



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

*The origin and functioning
of circumstantial clause linkers:
a cross-linguistic study*

Anna Martowicz

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

to
Linguistic and English Language
School of Philosophy, Psychology and Language Sciences
The University of Edinburgh

2011

© Copyright 2011

by

Anna Martowicz

Declaration

I hereby declare that this thesis is my own composition and that it contains no materials previously submitted for the award of any other degree as well as that the work reported here has been executed by myself.

ABSTRACT

The thesis analyses a variety of clause linkers used in the world's languages as dedicated markers of four circumstantial relations between states of affairs: anteriority ('after'), causality ('because'), purpose ('in order to') and conditionality ('if'). Focusing primarily on the function of the linkers rather than on their formal properties the study scrutinizes the origin and functioning of free-word adverbial subordinators, converbal endings as well as other subordinating affixes and polymorphemic structures specialized for clause-linking functions. The database consist of nearly 700 items which come from a geographically and genetically balanced sample of 84 languages.

In the first part of the thesis I discuss in detail the scope of the study, theoretical foundations (functional approach to language analysis) and methodology. In the second part I focus on the issues of grammaticalization and semantic scope of the linkers. The analysis of patterns of polysemy of the linkers, as well as the make-up of the polymorphemic items among them provide material for the reconstruction of the most common sources and pathways of grammaticalization. Looking closer at the patterns of semantic polyfunctionality (i.e. other circumstantial meanings that the clause linkers convey in addition to being markers of the four analysed relations) I reveal the architecture of the network of their cognitive affinities. The result of this investigation is a set of findings, which add to our understanding of the origin and functioning of the markers – a topic which has received little attention so far.

The third part of the dissertation is dedicated to the analysis of the cross-linguistic variation in the degree of grammaticalization, lexicalization and explicitness of the linkers and discussion on the motivations behind their development. The main body of the analyses is focused on the investigation of the potential influence of a variety of socio-cultural factors on the variations observed. The factors include those most often mentioned in the context of the forces shaping language structure: presence and vitality of written form as well as other media of displaced communication, presence of the language in schooling, population size of the speech community, and type of society. The results obtained show that the degrees of grammaticalization, lexicalization and explicitness are much stronger correlated with socio-cultural factors for anteriority and conditionality than for causality and purpose. In order to account for the differences I evoke the concepts of cognitive salience and communicative pressure arguing that the influence of socio-cultural factors on language structure should always be considered in a broader context which includes also pragmatics and cognitive psychology.

Acknowledgements

Both I and this dissertation owe a lot to many wonderful people.

First of all, I would like to thank Ronnie Cann for being the best supervisor a PhD student can dream of. Thank you for your wisdom, knowledge, kindness, patience, understanding and support. Thank you for allowing me to wander around the labyrinth of linguistics watching my steps at the same time and summoning me back to the right tracks on all those occasions when I was just about to enter the Dark Forests. It has been a great pleasure and a great honour to work under your wings.

My sincere thanks go also to Jim Hurford and Simon Kirby – my secondary supervisors – and to the members of the Language Evolution and Computation Research Unit as well as the entire Linguistics and English Language Department at the University of Edinburgh for creating a perfect research environment in which sharing of the linguistic experience and expertise is truly part of everyday life.

My PhD studies and completion of the thesis would not be possible without the generous support of the College of Arts and Humanities at the University of Edinburgh who awarded me with a scholarship and studentship making the most fantastic journey of my life possible. The same concerns all the small grants that the School of Philosophy, Psychology and Language Sciences awarded me throughout the years helping me to finance conference attendances and research visits abroad. Thank you for having confidence in my abilities.

I wish to express also my utmost gratitude towards all those linguists with whom I exchanged within these four years over a thousand e-mails in total, and whose contribution to this thesis cannot be underestimated. In no particular order I am expressing my thanks to: Charles Morfo, Sidney da Silva Facundes, Luise Hercus, Janet Watson, Xabier Artiagoitia, Swintha Danielsen, Ross McCallum Jones, Bertil Tikkanen, Steven Jacobson, Michael Dunn, Nancy Morse, Nicky de Jong, Volker Heeschen, Ronnie Cann, Ron Looker, Mati Erelt, Mark Post, Regine Koroma, Ger Reesink, Malami Buba, Philip Jaggard, Paul Newman, Amar Singh, Akira Yamamoto, Paul Frank, Carl Rubino, Niclas Burenhult, Yuri Shinya, Kaori Miura-Dowman, Robert Pensalfini, John Hutchison, David Solnit, Andrey Nefedov, Christa Kilian-Hatz, Ongaye Oda, Alain Kihm, Angela Terrill, Heleen Plaisier, Aone van Engelenhoven, Jan van Eijk, Wenshan Li, Gilles Gravelle, George Huttar, Noel Rude, Marie Lucie Tarpent, Ekaterina Gruzdeva, Chao-Lin Li, Chunming, Maciej Grochowski, David

Weber, Colette Grinevald, Clay Strom, Elizabeth Zeitoun, Christina Thornell, Madhusudan Mishra, Paul Sidwell, Arthur Holmer, Stephen Marlett, Zouhir Gabsi, Thomas Willett, Assibi Amidu, Bhuvana Narasimhan, Vipas Pothipath, Shoichi Iwasaki, René van den Berg, Rachel Nordlinger, Victoria Rau, John Bradley, William Foley, Jane Simpson

All of you have taught me more than I could have imagined – not only about the less and more exotic languages but also about the art of linguistics and fieldwork. Thank you for your time, kindness, patience and for your interest in my study. I hope that in the future I will have many occasions to share with others what I have learned from you – at both professional and personal level.

My thanks go also to the Linguistic Department at the Max Planck Institute for Evolutionary Anthropology in Leipzig for making my three research stays in the Institute: in November 2007, September-October 2008 and March-April 2009 such a wonderful experience. I am grateful to Bernard Comrie, Martin Haspelmath, David Gil and Juliette Blevins for finding time to share with me their thoughts on my research and linguistics in general and for all the useful comments I received from them. I will not forget about the staff of the Institute's library who were always ready to help me during the hundreds of hours that I spent among their collections and about Julia Cissewski for her help in organizing the research stays.

Ron Looker – thank you for proofreading of this thesis and for all the discussions we had about English and Polish in the last two years.

Finally, I wish to thank my husband and my daughter, my mum, my sisters, and parents-in-law for their constant support and encouragement throughout these years. You've been wonderful.

CONTENTS

Glossing conventions and abbreviations	XI
INTRODUCTION	1
<i>PART I Scope of the study, theoretical premises and methodology</i>	7
CHAPTER 1 Circumstantial relations between clauses	9
1.1. The notion of state of affairs and clause	9
1.2. Relations between states of affairs and relations between clauses	10
1.3. Classifications of circumstantial clauses	12
1.4. The focus of the study	14
1.4.1. Anteriority	15
1.4.2. Causality	16
1.4.3. Purpose	16
1.4.4. Conditionality	17
1.5. Strategies of encoding of circumstantial relations	18
1.5.1. Problems in defining subordination	19
1.5.2. Overview of the strategies	21
1.5.3. Thompson and Longacre's overview of strategies	28
1.5.4. Towards a unified, function-based classification	30
1.6. Summary	34
CHAPTER 2 Theoretical foundations and methodology	35
2.1. Theoretical foundations and approach to the analysed domain	35
2.1.1. Function-based definition of the studied subject	35
2.1.2. Instruments of analysis and explanatory apparatus	37
2.2. Language sample	40
2.3. Data collection	44
2.3.1. The database	44
2.3.2. Source 1: published materials	46
2.3.3. Source 2: the questionnaire	47
2.4. Summary	48
<i>PART II Sources and semantic scope of c-glossemes</i>	51
CHAPTER 3 Background and parameters of analysis	54
3.1. Forms and morphological complexity of c-glossemes	54
3.2. Insight into the origin of c-glossemes	57
3.2.1. Grammaticalization and its mechanisms	58

3.2.2. Polysemy	60
3.2.3. Grammaticalization pathways and the unidirectionality hypothesis	61
3.2.4. Multiple pathways and polygrammaticalization	63
3.2.5. Fixing, freezing and idiomatization	66
3.2.6. Types of evidence	67
3.2.7. Clause linkers and language contact	68
3.2.8. Previous studies on the origin of circumstantial clause linkers	70
3.2.9. Methods and assumptions – a summary	76
3.3. Insight into the semantic organization of the network of c-glossems	79
3.3.1. Degree of semantic polyfunctionality	80
3.3.2. Previous studies on the polyfunctionality of clause linkers	82
3.3.3. Goals, methods and assumptions	84
3.4. Summary	84
CHAPTER 4 Anteriority	87
4.1. Morphological complexity and forms of c-glossems	87
4.2. Insight into origins	89
4.2.1. Syntactic polyfunctionality and patterns of polysemy	89
4.2.2. Polymorphemic markers and their internal structure	100
4.3. Semantic polyfunctionality and cognitive affinity	105
4.4. Summary	115
CHAPTER 5 Causality	117
5.1. Morphological complexity and forms of c-glossems	117
5.2. Insight into origins	118
5.2.1. Syntactic polyfunctionality and patterns of polysemy	118
5.2.2. Polymorphemic markers and their internal structure	130
5.3. Semantic polyfunctionality and cognitive affinity	136
5.4. Summary	148
CHAPTER 6 Purpose	149
6.1. Morphological complexity and forms of c-glossems	149
6.2. Insight into origins	150
6.2.1. Syntactic polyfunctionality and patterns of polysemy	150
6.2.2. Polymorphemic markers and their internal structure	165
6.3. Semantic polyfunctionality and cognitive affinity	172
6.4. Summary	180
CHAPTER 7 Conditionality	183
7.1. Morphological complexity and forms of c-glossems	183
7.2. Insight into origins	184
7.2.1. Syntactic polyfunctionality and patterns of polysemy	184
7.2.2. Polymorphemic markers and their internal structure	197

7.3. Semantic polyfunctionality and cognitive affinity	202
7.4. Summary	212
CHAPTER 8 Comparisons and conclusions	215
8.1. Forms of c-glossemes	215
8.2. Degrees of morphological complexity	224
8.3. Degrees of syntactic polyfunctionality	225
8.4. Degrees of semantic polyfunctionality	228
8.5. Sources of c-glossemes	231
8.5.1. Evidence from the observation of patterns of polysemy	231
8.5.2. Evidence from the analysis of incorporated material	235
8.5.3. Other c-glossemes as sources	237
8.5.4. The emerging picture	238
8.6. Semantic affinities between the relations	239
 <i>PART III Cross-linguistic variation and its socio-cultural correlates</i>	 241
 CHAPTER 9 Degrees of grammaticalization, lexicalization and explicitness	 244
9.1. Degree of grammaticalization	244
9.1.1. Data coding and analysis	245
9.1.2. Geographic distribution	248
9.1.3. Summary	256
9.2. Degree of lexicalization	256
9.2.1. The notion of lexicalization	256
9.2.2. Data coding and analysis	257
9.2.3. Geographic distribution	260
9.2.4. Summary	265
9.3. Degree of explicitness	266
9.3.1. The notion of explicitness	266
9.3.2. Data coding and analysis	267
9.3.3. Geographic distribution	276
9.3.4. Explicitness and borrowing	279
9.3.5. Summary	282
9.4. Conclusions	283
 CHAPTER 10 Influence of socio-cultural factors	 285
10.1. Hypotheses and previous studies	286
10.1.1. Society structure	287
10.1.2. Written form and other modes of displaced communication	288
10.1.3. Contact	291
10.2. Parameters and design of the analysis	294

10.3. Results of the analysis	300
10.3.1. Correlations with degree of grammaticalization	300
10.3.2. Correlations with degree of lexicalization	305
10.3.3. Correlations with degree of explicitness	308
10.4. Summary	310
CHAPTER 11 Conclusions	313
11.1. Cognitive salience of the four relations	313
11.2. Pragmatic factors	316
11.3. Towards an explanation of variation	318
FINAL WORDS	323
REFERENCES	329
APPENDICES	361
Appendix I Genetic affiliation of the sample languages	363
Appendix II List of languages with references and names of consultants	367
Appendix III Questionnaire/list of recurrent questions	371
Appendix IV Degrees of grammaticalization (1)	374
Appendix V Degrees of grammaticalization (2)	376
Appendix VI Degrees of grammaticalization (3)	378
Appendix VII Degrees of lexicalization (1)	380
Appendix VIII Degrees of lexicalization (2)	383
Appendix IX Degrees of explicitness with borrowed c-glossemes included	386
Appendix X Socio-cultural profiles of the sample languages	390
Appendix XI Correlations between socio-cultural factors and degree of grammaticalization	393
Appendix XII Correlations between socio-cultural factors and degree of lexicalization	397
Appendix XIII Correlations between socio-cultural factors and degree of explicitness	400

GLOSSING CONVENTIONS AND ABBREVIATIONS

In glosses, labels for clause linkers are put in curly brackets (e.g. {COND}, {CAUSE}, {SIOVER}) according to the function they encode in a given example.

Polymorphemic markers are glossed morpheme-by-morpheme only when relevant. In such cases the label for circumstantial function is given after the morphological glossing e.g.

z **powod-u**
{from reason-GEN→CAUSE}

When a clause linker is discontinuous each of its parts is glossed separately with a number e.g.

U a kàrè lwɔhɔ tá-cya-ge e
she PRF go water {PURPOSE1}-seek-G.SG {PURPOSE2}

The clause linkers are additionally put in bold in the original sentences.

The abbreviations referring to particular circumstantial relations in the majority of cases follow those proposed by Kortmann (1997):

ANTE	anteriority ('after')
CAUSE	causality ('because')
COCOND	concessive conditionality ('even if')
COMMENT	comment/accord ('as')
COMPAR	comparison ('as if')
CONC	concession ('although')
COND	conditionality ('if')
CONTIN	contingency ('whenever')
CONTRA	contrast ('whereas')
IMMANTE	immediate anteriority ('as soon as')
MANNER	manner ('as', 'how')
MEANS	means ('by')
PLACE	place ('where')
POST	posteriority ('before')
PREFER	preference ('rather than')
PURPOSE	purpose ('in order to')
RESULT	result ('so that')
SICOEX	simultaneity co-extensiveness ('as long as')
SIDUR	simultaneity duration ('while')
SIMIL	similarity ('as', 'like')
SIOVER	simultaneity overlap ('when')
SUBSTI	substitution ('instead of')
TAQUEM	terminus ad quem ('until')
T AQUO	terminus a quo ('since')

The remaining abbreviations follow in principle the set proposed for the Leipzig Glossing Rules.

– morpheme boundary
= clitic boundary

1, 2, 3 first, second, third person

ABL	ablative
ABS	absolutive
ACC	accusative
ACT	action particle
ACTIV	active
ADEL	adelative (case)
ADES	adessive (case)
ADHORT	adhortative
ADJ	adjective
ADP	adposistion
ADV	adverb
ADV LZ	adverbializer
ALL	allative (case)
ANA	anaphoric
ANIM	animate
ANTICAUS	anticausative
AOR	aorist
AOPTCP	aorist particple
APPL	applicative marker
ART	article
AST	assertive particle
ATTR	attributive
AUX	auxiliary
BEN	benefictive
CAUS	causative
CLASS	class marker
CLF	classifier
CLsNMLZ	clause nominalizalizer
CNTRFCT	counterfactual
COMP	complementizer
CONJ	conjunction
CONT	continuative
CONV	converb
COP	copula
D	destination-orientation (verbal prefix)
DAT	dative
DEF	definite
DEM	demonstrative
DEP	dependency marker
DET	determiner
DIM	diminutive
DIR	directionality marker
DOBJ	direct object
DU	dual
DUR	durative marker
DYN	dynamic
E	epenthetic schwa

EMP	emphatic
EQUAT	equative marker
ERG	ergative (case)
EXT	extended (verb suffix)
F	feminine
FIN	finite
FOC	focus
FR	frequency marker
FUT	future
G	gender class marker
GEN	genitive
H	human
HAB	habitual
IMP	imperative
IMPRF	imperfect
IMPRFCONV	imperfective converb
INANIM	inanimate
INCL	inclusive
IND	indicative
INDEF	indefinite
INF	infinitive
INFCONV	infinitive converb
INS	instrumental
INTEN	intentional
INTER	interrogative
INTERJ	interjection
IRR	irrealis
LOC	locative
M	masculine
MID	middle
MOD	modality
N	neuter
NEG	negation, negative
NMLZ	nominalized/nominalization
NOM	nominative
NONFUT	non-future
NONFIN	non-finite
NOUNSUF	noun suffix
OBJ	object
OBL	oblique
OPT	optative
PART	particle
PASS	passive
PL	plural
POSS	possessive
POT	potential
PRF	perfect
PROG	progressive
PRON	pronoun
PROSP	prospective
PRS	present
PRTT	partitive marker
PST	past
PTCP	participle
Q	question particle
RDP	reduplication

REAL	realis
REL	relativizer (relative clause marker)
SEQ	sequentiality
SBJ	subject
SBJV	subjunctive
SBST	substantiviser
SG	singular
SPRX	speaker-proximate
SRDIR	superdirective (case)
SREL	superlative (case)
STAT	stative
STPTCP	static participle
TERM	non-subject marker
TNS	tense marker
TOP	topic
TR	transitivity marker
VERBSUF	verb suffix
VOC	vocative
VOLIT	volitional

Mojej Ukochanej Mamie, Cudownemu Mężowi i Najdroższej Córeczce

Of all mankind's manifold creations, language must take pride of place. Other inventions – the wheel, agriculture, sliced bread – may have transformed our material existence, but the advent of language is what made us human. Compared to language, all other inventions pale in significance, since everything we have ever achieved depends on language and originates from it. Without language, we could never have embarked on our ascent to unparalleled power over all other animals, and even over nature itself. (...) Language is mankind's greatest invention – except, of course, that it was never invented.

Guy Deutscher

INTRODUCTION

The development of grammar is commonly viewed as the final major step in the broadly understood evolution of human language (Givón 1979, Bickerton 1981; Liberman 1984). Within grammar, in turn, the emergence of means for complex sentence formation in the form of clause linkers which are the equivalents of English subordinators *if, as, although, who, that* etc., is, beyond any doubt, one of the most important achievements. The linking devices are missing from the so called “protolanguage” (Bickerton 1990) reconstructed on the basis of “fossils of language”: primate communication, language of children under the age of two, adults who have been deprived of language in the early years of their life and often also from the speech of patients with language disorders. For Jackendoff (1999), the evolution of “symbols that explicitly encode abstract semantic relationships”, which includes a variety of clause linkers, is the last step in the evolution of modern language. The evidence from grammaticalization studies also strongly suggest that of all the grammatical categories, clause linkers develop as one of the last groups (Heine and Kuteva 2007:111). The importance of the markers consists first and foremost in the novel communicative power with which they provide speakers i.e. the ability to express explicitly complex propositional thoughts. Although some go even as far as for arguing that – as a part of “utility vocabulary” – the linkers facilitate thinking:

Relational vocabulary plays an important role in thought. It has been argued that language enhances thoughts by making them available as perceptual objects (namely sentences), so that they can be attended to, focused on, modified, and remembered. Upon the invention of this ‘utility vocabulary’, it would all of a sudden be possible consciously to wonder if *p* and suppose that *p*, and to give reasons and purposes for actions, with a tremendous effect on the power of individual and communal reasons and planning. (What *should* I say to so-and-so? *If* he says this, *then maybe* I’ll do that; *but if...* Try to perform this reasoning without the italicized words). (Jackendoff 1999:277)

Four main types of clause linkers may be distinguished in the world’s languages: those that mark symmetrical relations between states of affairs (English *and, but, or*), those that introduce complement clauses (as in *I think that it was Lucy*), those introducing relative clauses (as in *This is the man who followed her*) and, finally, those linking clauses that stay in some sort of circumstantial relation (eg. *Kate went home because she was tired., Although I have not planned it, I will come to see you.*). The last group, which I call *circumstantial clause linkers*, is the most diverse of the four. It is a subset of markers belonging to this group that is the subject of interest in this study.

Despite the communicative (and allegedly also cognitive) importance of clause linkers, some languages have at their disposal much poorer sets of these markers than others. Interestingly, in many languages some of the types of clause linkers (eg. concessive linkers or purpose linkers) are not present at all and so the languages have to resort to other, less explicit strategies of clause-linking. On the other hand, the circumstantial clause linkers, when looked at from a cross-linguistic perspective reveal an astonishing diversity in morphosyntactic forms and morphological complexity. Often they display high levels of polysemy too – overlapping in form with members of other categories (such as adpositions, case markers etc.) and acting as exponents of more than one circumstantial meaning (eg. anteriority and causality). All of these issues are, however, still under-researched and have been used to facilitate our understanding of the origin and functioning of this interesting group of items only in a very limited scope.

The only systematic work that has looked in some detail into these problems is Kortmann's study from 1997. However, since it focused exclusively on adverbial subordinators and was limited to analysis of the languages of Europe, it covers only a fragment of the domain.

The present research is devoted to a cross-linguistic investigation of the origin and functioning of four, broadly understood, groups of circumstantial clause linkers: linkers of anteriority, causality, purpose and conditionality. It contributes to the rich body of grammaticalization literature by facilitating our understanding of various aspects of grammaticalization processes in the domain of clause-combining. The study is focused first and foremost on the reconstructions of grammaticalization pathways of a variety of clause-linkers, discussion on the motivation(s) for their development, the clues they give us on the organization of the network of circumstantial concepts in our mind and the picture of their cross-linguistic variation. In addition, clause-linking devices other than clause-linkers are also considered. This includes coordination and juxtaposition of clauses as well as strategies incorporating subjunctives, infinitives and other types of verb forms which are not allowed in independent clauses but which, by convention, became to be understood as encoding circumstantial relations between states of affairs with lesser or greater degree of ambiguity. In other words, the study offers a rich overview of more and less grammaticalized strategies which the world's languages employ in conveying circumstantial concepts.

The geographically and genetically balanced sample used for the purpose of this study consists of 84 languages and the number of analysed clause linkers approaches 700. Importantly, no structural constraints are put on the definition of clause linkers in this thesis and so the study covers a variety of markers which gives us the fullest possible picture of the analysed domain. The results of the analyses are interpreted in the spirit of functionalism – a research paradigm which assumes that grammars, i.e.

linguistic forms and structures, are ultimately explainable in terms of their function – human communication.

The thesis consists of three parts. In part I, which includes two chapters, the scope of the study, theoretical foundations and methodology are discussed. In chapter 1 the general idea of circumstantial relations between states of affairs and clauses is presented (section 1.1.) along with the distinction between symmetrical and asymmetrical relations made in this domain (section 1.2.). This is followed by the overview of the cross-linguistic classifications of circumstantial clauses proposed in the literature (section 1.3.) and a detailed presentation of the four circumstantial relations that are the focus of this study (section 1.4.). The variety of strategies that languages employ in encoding circumstantial relations between states of affairs is discussed in section 1.5. It begins with a brief overview of the problems in applying the term *subordination* in a cross-linguistic study. In section 1.5.2. terms such as *adverbial subordinator*, *converb*, *clause chaining* as well as *special verb forms*, *coordination* and *juxtaposition* of clauses are discussed as those most often mentioned in linguistic literature in the context of clause-combining strategies. The following section (1.5.3.) discussed the quasi-classification of the strategies proposed by Thompson and Longacre and indicates its weak points. In section 1.5.4. I propose my own function-based classification of the strategies which I apply later in the analyses. The most important point in the classification is the introduction of the category of *circumstantial-glosseme (c-glosseme)* which covers a variety of clause linkers serving the function of expressing circumstantial relations between clauses regardless of their form and complexity or any additional structural criteria which are parts of the definitions of adverbial subordinators, converbs etc. The chapter ends with a short summary.

The theoretical foundations and the methodology applied are the subject of chapter 2. It is emphasized in the introduction to the chapter that although the present study is not typological in nature, it draws extensively on the experience of linguistic typology in its basic assumptions, tools of analysis and explanatory apparatus as well as methodology. Section 2.1. presents the theoretical basis for choosing a functional approach to defining the studied subject as well as concepts such as: iconicity, economy and integrative functionalisms which are vital for approaching the topic of both origin and functioning of clause linkers. In section 2.2. the sampling method and the language sample used for the purpose of this study are discussed in detail. The issue of data collection is scrutinized in section 2.3. starting from the brief overview of the type of information needed for conducting the study and moving to the two types of sources used in this study: published materials and a questionnaire.

Part II of the thesis, grounded first and foremost in the grammaticalization theory, is devoted entirely to the investigation of the sources and semantic scope of the four groups of clause linkers. In this part the term *origin* is understood as historical origin and the term *function* as the role that a particular linker serves – a particular meaning(s) it encodes.

In the opening chapter of this part – chapter 3 – background information and parameters for the analysis to follow are discussed in detail. In the first section the classification of the linkers according to their forms and morphological complexity is presented. Section 3.2. discusses details of the possibilities of looking into the grammaticalization processes involving of c-glossems. Terms such *polysemy*, *unidirectionality*, *polygrammaticalization* and *idiomatization* are introduced in this section and the types of evidence used in grammaticalization studies are discussed along with issues of influence of language contact on the emergence and development of clause linkers. The section concludes with an overview of previous studies on the origin of circumstantial clause linkers with a special emphasis on the four semantic types which are the subject of this thesis. The penultimate section of chapter 3 presents in details the methods and assumptions that this study follows in the quest for the origin of these clause-linking markers. It is emphasized that the source of evidence used is exclusively of synchronic nature and is based on analysis of fossilized evidence of pathways of grammaticalization: cross-linguistic patterns of polysemy and morphological make-up of those clause-linkers which have been classified as polymorphemic. The last section of the chapter is, in turn, devoted to the issue of organization of the network of clause linkers and the insight it offers into the cognitive organization of our minds. It is explained that semantic polyfunctionality (i.e. the phenomenon of one clause linker serving as an exponent of more than one circumstantial relations) can be interpreted in terms of iconicity as an evidence for a close semantic proximity between the particular meanings, especially when the polyfunctionalities pattern cross-linguistically. Previous studies focused on this topic, as well as goals, methods and assumptions used in the investigation of the network of circumstantial meanings are also presented within this section.

Chapters 4-7, all of which have the same internal structure, are devoted to the presentation of the results of analysis of the morphological complexity and forms of the linkers; their origin; semantic polyfunctionality and the network of cognitive affinities. Chapter 4 discusses the findings concerning anteriority linkers, chapter 5 – causality linkers, chapter 6 – purpose linkers and chapter 7 – conditionality linkers. Each of the chapters concludes with a brief summary of the presented results and the entire part of the thesis is summarized in chapter 8 in which detailed comparisons of the four relations and conclusions are presented.

In part III, entitled “Cross-linguistic variation and its socio-cultural correlates”, the issue of origin and functioning of the linkers is looked at from the angle of motivations for the introduction of the linkers into a language system. The part begins with chapter 9 in which the cross-linguistic variation of degrees of grammaticalization, lexicalization and explicitness of the four groups of the linkers is discussed in sections 9.1, 9.2. and 9.3. respectively. For each of the sections information on data coding and analysis is explained separately and in each of them the quantitative analyses are accompanied by discussion on geographic distribution of the variation. The notions of *lexicalization* and *explicitness* are discussed in sections 9.2.1. and 9.3.1. respectively and in the section dealing with variation in the degrees of explicitness linguistic borrowing of clause linkers as well as alternative strategies of encoding are considered. The chapter concludes with a summary and comparison of the obtained results.

In chapter 10 the influence of socio-cultural factors on the encoding of circumstantial relations between clauses and emergence of clause linkers is considered. The chapter opens with an overview of hypotheses and previous studies focused on the extra-linguistic factors shaping language structure in section 10.1. The factors include in particular: society structure, written form and other modes of displaced communication and language contact. It is emphasized that of all the extra-linguistic factors it is the presence of written form that has been most commonly linked with the presence of clause-linking devices although the topic has never been investigated from a cross-linguistic perspective before. In section 10.2. the design of the analysis is discussed in detail and so are the main parameters of the analysis: number of speakers, level of written form development, presence and characteristics of the language in school teaching, radio and TV broadcasting and additional (general) parameters. Section 10.3. presents the results of the correlation analyses – separately for degree of grammaticalization, lexicalization and explicitness.

Finally, chapter 11 gathers the results presented in part III and aims to answer the question of the motivation behind the emergence of clause linkers and reasons for the cross-linguistic variation in their distribution. It is in this chapter that the cognitive, pragmatic and socio-cultural factors are considered jointly. The thesis concludes with a chapter entitled Final Words.

Taking into account the methods of analysis, the present study may be viewed as **empirical** in that it is based on material collected from real languages, **quantitative** in that it aims to reveal similarities and differences between languages by looking at frequencies of occurrences of certain phenomena, **qualitative** in that it explores clusters in distribution and tendencies, and **theoretical** in that it proposes hypotheses explaining the cross-linguistic variation.

PART ONE

SCOPE OF THE STUDY, THEORETICAL PREMISES AND METHODOLOGY

Those categories of thought connection which we express by means of 'if', 'because', 'although'... etc. do of course belong to the most indispensable tools of any reasoning mind; no people can do without them, each linguistic community operates in its thinking with conditions, causes, restrictions, alternatives.

Georg von der Gabelenz

CHAPTER 1

Circumstantial relations between clauses

The incredible cognitive apparatus that humans are equipped with is capable not only of perceiving and recording things that happen but also of abstracting about things that may happen and might have happened. Furthermore, it enables us to relate these “things that happen/happened/may/might have happened” to each other in a variety of scenarios. It is this very ability, which is universal among all humans and which is the basis of higher level propositional thinking and reasoning.

This chapter begins with the introduction of the complex notion of *state of affairs* and *clause* in section 1.1. In section 1.2. the basic distinction between symmetrical and asymmetrical relations between states of affairs and clauses is discussed. Classifications of the asymmetrical relations proposed in typological literature so far are presented in 1.3. Section 1.4 is devoted to more detailed presentation of the four relations which are the subject of this study: anteriority, causality, purpose and conditionality and in section 1.5. a detailed discussion on the strategies of encoding the circumstantial relations is presented. The chapter concludes with a summary in section 1.6.

1.1. The notion of state of affairs and clause

Of all terms such as *facts*, *situations*, *states of affairs*, *events* etc. used to describe the broad category of “things that happen” the notion of *event*, has been used most commonly in philosophy and logic as well as in linguistics. The term dates back to Plato’s dialogues but after two millennia of disputes, researchers still have not reached an agreement as to what “event” and the other terms such as “facts”, “activities”, “accomplishments”, “achievements”, and “states” exactly mean and how they are related to each other.¹ It is not my aim to enter here the discussion on events

¹ For a very detailed overview of the problems with distinguishing between events, facts, activities, accomplishments, achievements, and states from both theoretical and linguistic (semantic) angle and

which occupies a prominent role in modern semantic thought (cf. for instance the works by Davidson 1967, 1980, Parsons 1989, 1990, 1991, Bennett 1988, 1996 and the rich literature on event theory listed in the bibliography on event semantics by Condoravdi and Filip 2007). Instead, I simply apply the distinctions and disambiguate the concepts following Functional Grammar (FG) where the term *state of affairs* (henceforth SoA) is used unambiguously as a hyponym of different classes of predicates such as situations, actions, events and processes (see Dik 1997a:105). A SoA, as Siewierska puts it, “designates the conception of something that may be the case in some world” (1991:43), the term event being just one instance of SoA, namely a dynamic SoA.

In language, a given state of affairs is realized by predication, which – together with its propositional content – forms a clause. For instance, in the clause below the (nuclear) predication consists of the predicate *read* and the terms *John* and *book*.

(1.1.) *John read a book.*

The state of affairs the clause encodes (‘John reading a book’) becomes a proposition in an act of utterance. The working definition of clause I follow in this thesis has been adapted from Lehmann:

We will assume a broad concept of the clause which comprises any syntagm containing one predication. Syntactically, this means that – apart from nominal clauses – the uppermost controller of dependency in the syntagm is a verbal form. Since a verbal form may be finite or non-finite, this includes nominalized clauses. (1988:182)

1.2. Relations between states of affairs and relations between clauses

States of affairs may be related to each other in two ways: symmetrically or asymmetrically. In symmetrical relations both SoAs are viewed as equal in their importance. The asymmetrical relations involve, by contrast, certain representational dependences where one of the SoA is conceptualized as, for instance, a reason, condition or temporal antecedent.

The distinction between symmetrical and asymmetrical types of relations maps also onto their linguistic representations – complex clauses. In linguistics (as well as in logic) there is a long tradition of distinguishing 3 groups within the symmetrical relations: conjunction, adversativity and disjunction which all fall under the term *coordination*. In languages across the world these relations are most commonly (but not exclusively) expressed by using so called *coordinating structures* where two clauses are linked by coordinating particles such as *and*, *or*, *but*. Relations between various

additional references see Casati and Varzi (1996) as well as an annotated bibliography 1947 to 1997 prepared by the two authors (1997). For a linguistic-centered discussion see Hopper (1995).

structures that reveal dependency (the asymmetrical ones) are, in turn, traditionally associated on the sentential level with the notion of *subordination* (see section 1.5.1. for more detailed treatment of the topic). Within the complex structures involving clausal subordination, again, three types of constructions are commonly distinguished: complement clauses, relative clauses and adverbial clauses.

According to this distinction complement clauses (cf. the underlined fragment of the sentence in 1.2.) serve the function of arguments of predicates; relative clauses serve the function of noun phrase modifiers but not arguments of a predicate (cf. the underlined fragment in 1.3.); adverbial clauses act as modifiers of any verb phrases or sentences, serving the same role as manner, locative, temporal and other adverbs would serve (cf. the underlined fragment in 1.4.).

(1.2.) *I am not convinced that you did the right thing.*

(1.3.) *The girl who just left was my sister.*

(1.4.) *I went there to watch the parade.*

Cristofaro (2005, chapters 5-7) gives a good summary of the types of relations between SoA in these three types of clauses: in complement clauses the link is such that the main SoA entails that the other one is referred to, in relative clauses the dependent SoA provides some kind of specification about a participant of the other and for adverbial clauses two SoAs are linked such that one corresponds to the circumstances under which the other one takes place.

With regard to the last group, which is of special importance for this thesis, and asymmetrical relations between SoAs in general, Talmy (1978 and 2001) has introduced a distinction between Figure and Ground which has been applied also by Croft (2001). In this framework Figure and Ground (terms taken from Gestalt Psychology) refer to concepts that need anchoring and those which do the anchoring respectively. In order to illustrate the position of Figure and Ground with respect to arrangement of clauses Talmy describes the case of a temporal relation:

The Figure is an event whose location in time is conceived as a variable the particular value of which is the relevant issue. The Ground is a reference event, one that has a stationary setting relative to a reference frame (generally, the one-dimensional timeline), which respect to which the Figure's temporal location is characterized. (2001:320)

The simplest example illustrating the temporal Figure/Ground distinction between SoA in English is a complex sentence with a subordinator such as:

(1.5.) *I came in **after** John left.*

where the SoA 'John's leaving' has a Ground meaning assigned to it, since it is the reference point for the Figure SoA – 'my coming'.

The Figure/Ground distinction is very closely related to the distinction between foregrounding and backgrounding discussed in Hopper's (1979) and Hopper and Thompson's (1980) works and to the division between supporting and focal clause introduced by Dixon (2009). Despite the attractiveness of such cognitive approaches to the analysis of asymmetrical relations between SoAs and clauses there are, however, a couple of problems with cross-linguistic application of these distinctions, as Cristofaro (2005:26) has aptly noticed. The problem that is of relevance to this study is that the two types of phenomena have to be identified often on the basis of extensive contextual analysis, which is difficult in wide cross-linguistic studies (for similar points see also Myhill and Hibiya 1988:362 and Dixon 2009:2-5). For this reason I limit myself here to a more functional distinction delimiting between clauses conceptualizing the circumstantial concepts (circumstantial clauses) and the clauses which they are attached to (main clauses). In Functional Grammar terms these two types refer to *clauses* and *circumstance satellites* respectively (see Dik 1997b) and in traditional grammars to *adverbial adjuncts/modifiers/adverbial clauses* and *main/matrix clauses*. I choose to use the term *circumstantial clause* rather than *adverbial clause* since the adjective 'circumstantial' may be applied indicating not only the type of clauses but also the conceptual characteristics.

1.3. Classifications of circumstantial clauses

The classifications of circumstantial clauses (most often under the heading "classification of adverbial clauses" or "classification of interclausal relations") proposed in traditional grammars usually contain the following set of semantically defined classes: clauses of time, place, manner, reason, purpose, result, condition, concession, manner and degree. The classifications differ, however, for more fine-grained distinctions. The main reason for that is that the conceptual and formal boundaries between various circumstantial relations are fuzzy and tend to overlap, which translates directly onto the classification of clause types (c.f. Harris 1989:341; König 1986:229; Kortmann 1997:79).² For the purpose of this study those classifications that have been proposed on the basis of cross-linguistic research are of special importance and it is these that we should focus on here. There are three such proposals. The oldest one is the typology proposed by Thompson and Longacre (1985) where the following 12 types of adverbial clauses are distinguished:

² The conceptual fuzziness is a very important phenomenon in its own rights and is discussed in detail in part II of the thesis under the label *semantic polyfunctionality*.

Time ('when, after, before'), location ('where'), manner ('as, as if'), purpose ('in order to, in order that'), reason ('because'), circumstantial ('by, without'), simultaneous ('while'), conditional ('if, even if, unless'), concessive ('although'), substitutive ('instead of'), additive ('besides, in addition to') and absolute (subordinated clauses without explicit signals of the relationship between the main and subordinate clause which interpretation is inferred from the pragmatic and linguistic context).

Hengeveld's (1993) proposal of classification, based on the analysis of European languages, differs in three points from that by Thompson and Longacre. Firstly, the author has proposed to divide adverbial clauses of time into clauses that express simultaneity, anteriority and posteriority. Secondly, he has separated from condition the relation which he calls *potential circumstance*. Finally, he has excluded cases of non-explicit subordination. His typology therefore consists of 15 types of clauses:

simultaneity ('when'), anteriority, posteriority, manner, cause, reason, explanation, condition (if), potential circumstance (in case), concession, concessive condition, result, purpose, means and addition.

The most comprehensive set of adverbial clauses has been proposed by Kortmann (1997) in his cross-linguistic study on adverbial subordination based on 49 European languages. As the author explains this set "is fairly complete and includes all of the major types of semantic relations which are explicitly marked by means of clause linking devices" (1997:81). It consists of 32 interclausal relations divided into 4 groups:

TIME

simultaneity overlap 'when'
simultaneity duration 'while'
simultaneity co-extensiveness 'as long as'
anteriority 'after'
immediate anteriority 'as soon as'
terminus a quo 'since'
posteriority 'before'
terminus ad quem 'until'
contingency 'whenever'

CCC (causal, conditional, concessive and related interclausal relations)

cause/reason 'because'
condition 'if'
negative condition 'unless'
concessive condition 'even if'
concession 'although'
contrast 'whereas'
result 'so that'
purpose 'in order that'
negative purpose 'lest'

MODAL

manner 'as, how'
similarity 'as, like'
comment/accord 'as'
comparison 'as if'
instrument/means 'by'
proportion 'the ... the'

OTHER

place 'where'
substitution 'instead of '
preference 'rather than'
concomitance 'wobei' (German)
negative concomitance 'without'
addition 'in addition to'

All of the three authors whose classifications have been referred to here, have used overt markers of adverbial relations between clauses – the so called *adverbial subordinators* (cf. section 1.5.2.) – as exemplifications of particular types of clauses. It comes as no surprise then that the subordinators themselves adopt the name of the relation they encode – hence we have for instance: causal subordinator (*because*), subordinator of purpose (*in order that*), subordinator of manner (*as*) etc. Understandably, the labels of the types of circumstantial clauses are extended over circumstantial relations themselves and so we talk about the relation of cause, purpose, manner etc.

1.4. The focus of the study

The circumstantial relations whose encoding is the subject of this study are: anteriority, causality, purpose and conditionality. The number of relations included has been dictated by time and length limitations of this study. They have been chosen as to represent various semantic types of circumstantial concepts taking into account also the likelihood of finding sufficient information on their encoding in the published materials. Anteriority represents the temporal domain. Conditionality belongs to a group of concepts which include also concessive conditionality ('even if') and negative conditionality ('unless'). Purpose and negative purpose ('lest') form another domain while causality is a one-member subtype of the group of circumstantial relations. The encoding of these four concepts, unlike the encoding of relations such as contingency ('whenever'), terminus ad quem ('until'), contrast ('whereas'), comparison ('as if') and many others presented in the classifications in the previous section, is discussed almost without exceptions in grammars and grammar sketches of the world's languages which include sections on complex sentence formation.

Each of them is discussed below along with corresponding types of circumstantial clauses in English and with determination of the actual scope in which they are considered in this thesis. Of these four groups of clauses considerable amount of attention has been devoted so far only to the investigation of conditional clauses (see in particular Traugott et al. 1986; Athanasiadou and Dirven 1997; Dancygier 1998; Podlesskaya 2001; Khrakovskij 2005 and references therein). A typology of purpose

clauses has been proposed very recently by Schmidtke-Bode (2009) and apart from the earlier generative-oriented work on purpose clauses in English by Jones (1991) this is the only detailed study on this group of circumstantial clauses. In all these works, the focus has been, however, more on formal features of the clauses (such as finiteness of verb forms and TAM marking) than on the clause linkers used to encode the particular relations. The other two types of clauses: causal and anteriority still await more detailed cross-linguistic treatment.

1.4.1. Anteriority

The *anteriority relation* is defined as a temporal relation between two SoAs where one of them (called the anterior one) is realized and completed before the other one takes place. An *anteriority clause* (in the literature called sometimes also *sequential clause*) is the factual clause which encodes the temporally earlier SoA.³ In English the relation is prototypically marked by the use of the subordinating conjunction *after* as in (1.6.) where the underlined clause is the anteriority clause:

(1.6.) After I came home I called my friend.

No further distinction is to be made in this work between anteriority clauses (and anteriority markers) expressing sequencing of SoAs in the past (as in 1.6.) and in the future (1.7.).

(1.7.) *After/when* I come home I will call my friend.

Anteriority is distinguished from immediate anteriority i.e. such sequencing of SoAs where they follow each other in time (almost) without any interval (cf. 1.8.). It is exclusively the first relation that is considered in this study.

(1.8.) *Immediately after/as soon as* I came home I called my friend.

Finally, it is the marking of the anteriority relation on the clause expressing the temporally earlier SoA, as in (1.6.) and (1.7.) that is the subject of investigation here. The marking of this relation on the clause coding the temporally later SoAs – using so called *connective adverbs* or *discourse markers*, which are equivalents of (*and then/later/afterwards* (1.9.) – is therefore excluded from the majority of analyses but is

³ The term *factuality* has been applied by Pérez Quintero in her functional study on adverbial subordination in English. By *factual clause* the author understands a clause “which describes a property or relation as applicable; a SoA as real, a propositional content as full and a speech act as assertive” (2002:53). The concept is closely related to that of *factivity* introduced by Kiparsky and Kiparsky (1970) and adopted by Hengeveld (1993) and Dik (1997a,b).

considered in the discussion on degrees of explicitness of encoding of the relation in part III of the thesis.

(1.9.) *I came home (and) then/later/afterwards I called my friend.*

1.4.2. Causality

Two SoAs of which one (called the causal one) represents the reason for the other one to occur are to be viewed as being in the *relation of causality*. A *causal clause* is defined then as a factual clause expressing the SoAs stating the cause/reason. In English this type of clause is usually introduced by *because*, *since* or *as*:

(1.10.) *John had to cancel the meeting because/since/as he got sick.*

A variety of further distinctions could be introduced to the discussion of causal clauses and causality linkers such as: a distinction between cause and reason (especially popular in philosophical considerations⁴); a distinction between eventive, epistemic and illocutionary cause⁵; or a distinction between causal relations in co-referencing and non-co-referencing clauses. The main reason why these distinctions are not applied here is that they are very rarely reported in the linguistic literature as being grammaticalized.⁶ It would be a task far beyond the scope of this study to survey the sampled languages for such distinctions. Hence, all the clauses and all the clausal markers expressing the broadly understood causal relations between SoAs are to be considered in this thesis.

1.4.3. Purpose

Purpose relations between SoAs are to be understood as such relations where one of the SoAs (the purpose one, often involving a concept of motion) is performed with the goal of realization of the other one. Following from this, a *purpose clause* is a

⁴ The two have been delimited, for instance, by Wittgenstein (1979) who viewed causes as perceptual concepts (hypotheses about how events are connected in the world) and reasons as intentional concepts (justifications that we give for certain actions or propositions).

⁵ On this point see Pérez Quintero (2002:67-68).

⁶ In fact, I have found only four such examples in the sample of the 84 languages I analysed. The first two appear to be more or less evident grammaticalization of the distinction between cause and reason and come from Lezgian, where the *luhuz/lahana* marker is used most commonly to express subjective motivation in causal/reason clauses (Haspelmath 1993:390) and from Eipo, where the use of the cause/reason marker *tennen* is restricted to cases which derive from the following sequence: human agent thinking or saying something and then acting on something or someone (Volker Heeschen, personal communication). The other two are cases of markers specialized to encode causal relations in either same-subject or different-subject clauses. They come from Krongo, where the prefixal marker *má-* is used exclusively in coreferencing clauses (Reh 1985:349) and Retuarā which uses suffixal markers *pakā?ā* and *wa?ri* in coreferencing and non-coreferencing clauses respectively (Strom 1992:173-175).

clause encoding the SoAs that conceptualizes purpose. The most common and explicit way in which English expresses such relation is by use of the purpose marker (*in order to*):

(1.11.) *Kate went to the bank **in order to** apply for a mortgage.*

Purpose clauses (in more traditional grammars sometimes called also *final clauses*) are by definition non-factual. Moreover, they can be viewed as “reason formulated in terms of [the] intended outcome” (Jackson 1995:57).⁷

In this thesis no distinctions within the domain of purpose clauses and markers of purpose clauses are made with regard to motion and non-motion, eventive purpose and epistemic purpose.⁸ One important and commonly grammaticalized distinction that is taken into account in part III of the thesis is the distinction between coreferential (same-subject) and non-coreferential (different-subject) purpose clauses and markers. A coreferential purpose clause shares an agent with the main clause and the agent can, obviously, control the realization of the dependent SoA (such as in 1.11.), while in a non-coreferential purpose clause there is no agent sharing and the agent of the main clause cannot control the realization of the SoA encoded by the purpose clause (see 1.12.) As Cristofaro remarks “By their very nature, however, purpose relations imply that the performer of the main SoA is in some way involved in the realization of the dependent one, at least in that there is an element of will on his or her part towards such realization” (2003:157).⁹

(1.12.) *Kate went to the bank **so that** John could stay at home.*

In this study only cases of positive purpose are considered. Negated purpose, i.e. such relation where a certain SoA is performed in order to prevent another one from occurring (which is in English prototypically marked by the use of the subordinator *in order not to* or *lest*), is excluded from the range of relations looked at here.

1.4.4. Conditionality

A **conditional relation** establishes a connection between two SoAs such that the occurrence of one of them (the conditional one) is the condition for the occurrence of the other (which, in turn, can be viewed as the consequence of the former). Thus, a

⁷ I go back to that issue in chapter 6.

⁸ Again, for the latter distinction see (Pérez Quintero 2002:63-64)

⁹ Cf. also Longacre, Thomson and Hwang 2007:253.

conditional clause is a non-factual clause expressing the condition. In the linguistic literature it is often called the *protasis* (or *antecedent*), while the main clause it is attached to – the *apodosis* (or *consequent*).

The domain of conditionality involves a variety of situations which differ regarding the likelihood of their occurrence. This thesis is focused exclusively on the marking of protasis in what I call here **real conditions** – i.e. such that are possible to fulfil. This includes the following three types of conditionals distinguished by Schachter (1971) and adopted by Thompson, Longacre and Hwang (2007:254-255)¹⁰: real present (1.13.), real habitual/generic (1.14.) and unreal predictive (1.15.).

(1.13.) *If it's raining out there, my car is getting wet*

(1.14.) *If you step on the brake, the car slows down*

(1.15.) *If he gets the job, we'll all celebrate*

The distinction between real conditionals and other types of conditionals (especially imaginative – i.e. hypothetical and counterfactual) is quite often encoded in the configuration of TAM markers in the protasis and/or in the clause linkers.¹¹ I focus here exclusively on the markers of real condition and leave the discussion on the other types for future research. The obvious advantage of such an approach is that the outcomes of analysis of various aspects of encoding of conditionality will allow for extensive comparisons and more detailed coverage of the broad topic. For a similar reason, cases of past conditionals, as well as so called negative conditionals (prototypically encoded in English by the subordinator *unless*) and concessive conditionals (*even if*) are excluded from the analysis here. Finally, no distinction is made between eventive, epistemic or illocutionary conditions and the like.¹²

1.5. Strategies of encoding circumstantial relations

In this section I look at the range of strategies that are used in the languages of the world to encode circumstantial relations between SoAs. It needs to be emphasized,

¹⁰ This is only one of the many classifications proposed for conditionals. Other authors distinguish, for instance, between: closed, open, tentative, counterfactual (Declerck, Reed 2001); counterfactual, unlikely, hypothetical, given (Haiman, Kuteva 2001); open and remote (Huddleston, Pullum et al. 2002); indicative (predictive and nonpredictive), subjunctive and counterfactual (Kaufmann 2005).

¹¹ In the languages I have looked at, the distinction between real, hypothetical and counterfactual conditionals is grammaticalized in clause linkers most commonly in African and Austric languages. An example of this phenomenon is to be found, for instance, in the Konso language of Ethiopia where the *oo/oon* subordinator is used in real conditional clauses, *kandé* in hypothetical and *kanden* in counterfactual ones (see Oda 2000:23-26).

¹² Cf. Sweetser (1990), Quirk et al. (1991) and Pérez Quintero (2002).

however, that what interests us here is first of all the marking of circumstantial relations on subordinate clauses.

I begin the discussion with a consideration of the problem of defining *subordination* – a term that in Western linguistic tradition is most strongly associated with the encoding of circumstantial relations. In 1.5.2. I present an overview of the various strategies and various terms occurring in the works devoted to encoding of circumstantial relations. Sub-section 1.5.3. discusses the set of strategies listed by Thompson and Longacre in what is the only available (quasi-)classification of the strategies in question and scrutinizes the flaws of the classification. Finally, in 1.5.4., bearing in mind the main subject of this thesis, I present my own, functionally-oriented, classification of the strategies and the term *circumstantial glosseme* covering a variety of structural types of items with clause-linking functions.

1.5.1. Problems in defining subordination

It has been emphasized a number of times (Ramat 1999, Croft 1991, Gil 1991 and 2001, Haspelmath 2007 *inter alia*) that many of the syntactic concepts that are being used in modern linguistics have originated in Latin and Greek grammars and apply best to the description of Indo-European languages (and not all of them either). Since languages in other language families vary considerably from the European ones, the cross-linguistic validity of these categories is often problematic. To escape the, as Gil (2001) puts it, Europocentrism, field linguists are still being forced to adopt the Boasian approach of positing language-specific categories for the languages that they work with. This remains a serious problem for theoretical linguistics, including cross-linguistic and comparative investigations of linguistic phenomena (see section 2.1.1. for more discussion).

Regarding the scope of this thesis the problematic character of the concept of *clausal subordination*, has to be emphasized. The strongest opinion in this respect was expressed probably by Thompson:

“Subordinate clause” is not a grammatical category at all. That is, there does not seem to be a single function or even a group of functions that we can think of this ‘category’ as having been designed, as it were, to serve. So the term ‘subordination’ seems to be at best a negative term which lumps together all deviations from some ‘main clause’ norm, which means that it treats as unified a set of facts which we think is not a single phenomenon. (1985:86)

This fuzzy character of syntactic categories such as subordination and coordination has been a subject of discussion for decades. It was probably the Dutch grammarian Kruisinga (1932:501) who first noted that “it is perhaps hardly necessary to observe that the distinction between coordination and subordination is a relative one, allowing of intermediate cases”. Cristofaro has given a good summary of the problems

arguing that neither the dependency nor the embedding which have been proposed as criteria of distinguishing subordination from other complex structures are suitable for cross-linguistic comparisons¹³:

Any parameter chosen to distinguish between subordination and non-subordination will combine with a number of other parameters, yielding a variety of possibly very different clause linkage (sub)types. In this respect, the distinction between subordination and non-subordination should not be regarded as a discrete one (as implied by the opposition between subordination and coordination), but rather as a syntactic continuum involving a number of different and quite freely combinable parameters. (2005:20)

Huddleston (1984), Lehmann (1988) and Givón (1990) also argued for such a continuum. In Quirk et al. (1985:927-28) the idea of gradation between subordination and coordination has been worked out in detail. The authors have posited six syntactic criteria to characterize coordinating conjunctions in English – the more criteria a particular item conforms to, the more it resembles a coordinating conjunction and the less a subordinating conjunction. Van Valin (1984:546) and Olson (1981) and then van Valin and LaPolla (1997) proposed an additional term – *cosubordination* – for those cases where one clause is dependent but not embedded into another. However, the distinction of dependency and embedding is, again, not a universally acceptable one.¹⁴

The most reasonable way out of this confusion seems to be the approach chosen by Cristofaro who uses the term subordination as referring to:

a particular way to construe the cognitive relation between two events, such that one of them (which will be called the dependent event) lacks an autonomous profile, and is construed in the perspective of the other event (which will be called the main event). (2003:2)

According to this functional definition, subordination is, thus, defined with respect to asymmetrical relations between SoAs (cf. section 1.2.) that we perceive and process and not with respect to structural properties and arrangement of syntactic elements in a complex sentence. Because of the suitability for cross-linguistic comparisons (see section 2.1.) I choose to follow Cristofaro in her understanding of the term and, hence, whenever the notion of ‘subordination’ is used in this thesis it should be understood in its functional and not structural sense.

¹³ It has been often the case in traditional grammars (as well as in Quirk et al 1985; Van Valin and La Polla 1997, for instance) that the term *subordination* was used interchangeably with the term *embedding* and *subordinated clause* interchangeably with *embedded clause*. On the other hand, there are authors such as Huddleston (1984) and Hooper and Traugott (2003) who see embedding as a case where clause Y is a constituent of clause X as opposite to the case of subordination where X clause and Y clause are each immediate constituents of the sentence in which case ‘subordination’ is a hyponym of ‘embedding’. Needless to say, syntactic dependency itself is commonly perceived as a gradable phenomenon (cf. Mithun 1984; Lehmann 1988; Haspelmath 1995; Kortmann 1997; Van Valin and La Polla 1997; Fabricius-Hansen and Ramm 2008 for discussion and more references).

¹⁴ See footnote 13.

1.5.2. Overview of strategies

So far only a relatively small number of cross-linguistic surveys on means of expressing of circumstantial relations have been conducted. There are just a couple of more systemic studies which, however, rarely go beyond the languages of Europe and Asia (Haspelmath and König 1995, Kortmann 1997, Bisang 1998). In the majority of cases the discussions are limited either to very brief cross-linguistic overviews or descriptions of particular languages. The aim of this section is to put these bits together and portray the variety of phenomena that have been considered under the label “strategies of encoding of circumstantial relations”, as well as the terminological confusion that has arisen in this field.

Adverbial subordinators

The term *adverbial subordinator* (commonly also *adverbial conjunction*, *subordinating conjunction*, *subordinator*, *adverbializer*) is usually understood as a “free morpheme which operates over a subordinate clause serving as an optional adverbial modifier of the main clause” (Kortmann 2001:842). This includes one-word free morphemes such as the Mantauran Rukai *lo* in (1.16.), as well as combinations of morphemes as in the English purpose subordinator *in order to* (1.11.)

(1.16.) Mantauran Rukai (Zeitoun 2007:152)

Lo *pa'amaolro-li* *amo-paavanao-lrao.*
{COND} have time-1SG.GEN IRR.PASS–DYN.in.hurt-1SG.NOM
'If I have time, I will bathe.'

Adpositions with scope over nominalized verb forms, such as the causal postposition *kilinga* in Lezgian (1.17.), are also often classified as subordinators.¹⁵

(1.17.) Lezgian (Haspelmath 1993:389)

Wiči-n *wezifa-jar* *haqisağwil.e-ldi* *tamamar-un.i-zz* *kilinga*
self-GEN duty-PL conscientiousness-SRDIR fulfil-NMLZ-DAT {CAUSE}
kawaxa.di-z *xür.ü-n* *žemät.di-n* *arada* *jeki* *hürmet* *awa-j.*
chairman-DAT village-GEN people-GEN among big respect be.in-PST
'Since he fulfilled his duties conscientiously, the chairman enjoyed great respect among the villagers'

¹⁵ One strong rationale for counting adpositions operating over nominalized clauses as instances of clause linkers is that over time they tend to develop into traditionally understood adverbial subordinators (see section 3.2.8.) These various stages involving adpositions are still clearly visible in English, where *after* can be used with nouns (*After breakfast John went back to work*), nominalized verbs (*After finishing breakfast John went back to work*) as well as with finite verbs (*After he finished breakfast, John went back to work*).

Surprisingly, this prominent group of clause-linking devices, has received little attention so far. In the typological literature there is only one study focused on adverbial subordinators – Kortmann (1997). Although limited in scope (cf. section 3.2.8.), this work is, without any doubt, very insightful and inspiring. It is referred to in numerous places in this thesis.

Converbs

The category of converbs, initially used in Altaic linguistics, has been introduced to wider cross-linguistic discussions by Nedjalkov and Nedjalkov (1987). The term *converb* is nowadays most commonly used in the meaning proposed by Haspelmath i.e. “a nonfinite verb form whose main function is to mark adverbial subordination” (1995:3). As the author adds, “another way of putting it is that converbs are verbal adverbs, just like participles are verbal adjectives”. Understood this way the term covers a range of verb forms traditionally called *gerunds*, *adverbial participles*, *conjunctive participles* etc.

On the one hand converbs are devoid of certain TAM or agreement categories usually required on verbs in independent (main) clauses and, on the other hand, they are often marked by an addition of an adverbializing morpheme (or morphemes) to the verbal stem. The morpheme may be bound, as in the Japanese example in (1.18.), or free as the particle *en* in the French *gérondif* e.g. *en chantant* – ‘singing’ (Haspelmath 1995:9).

(1.18.) Japanese (Onishi 1994:375)

Omae ga ike-ba ore wa ika-nai.
 you NOM go-**{COND}** I TOP go-NEG

‘If you go, I will not go.’

An example of a converb which does not take any adverbializing morphemes, is to be found, for instance, in Ge’ez where the verbal adverb is formed by vowel pattern *CaCiC*:

(1.19.) Ge’ez (Haspelmath 1995:6)

Nabir-eya tanāgar-ku mesl-ēhomu.
 sit.down.CONV-1SG.POSS talk-PRF.1SG with-3PL

‘Having sat down, I spoke with them.’

So far the broadest study devoted to converbs is the selection of papers published in the volume edited by Haspelmath and König (1995). Some papers in the volume edited by van der Auwera and Ó Baoill (1998) also offer a more detailed treatment of the topic.

Clause chaining

Clause chaining (or simply: *chaining*) is a strategy of linking a number of clauses, which we find in the languages of New Guinea, Africa and in the Tibeto-Burman languages as well as in Australia and Americas.¹⁶ As to date (2011), however, the only two authors who have (very briefly) discussed chaining from a cross-linguistic perspective are Longacre (1985 and 2007) and Givón (1990). It is the first of the two authors whose definition I wish to quote here:

In a chaining structure (...) it is simply not possible to join two (...) verbs of the same rank in the same sentence. A sentence either ends in a dominating verb of fuller structure than that of the preceding verbs, or, alternatively, begins with a dominating verb of fuller structure than that of the following verbs. In the former case, the preceding verbs of restricted structure are often referred to as medial verbs (or as participles, gerunds, or even coverbs) while the dominating verb at the end is referred to as the final verb. In the latter case, the following verbs of restricted structure are referred to as consecutive (or sequential) verbs while the dominating verb at the beginning is referred to as the initial verb. In the former case we speak of medial-final chaining; in the latter case we speak of initial-consecutive chaining. (Longacre 2007:375)

The distinction between the final verb in the final clause, and the verbs of restricted structure (which are often labelled *medial verbs* irrespective of their place in the sentence) is, in principle, the same as between the traditionally distinguished finite and nonfinite verbs – the dominating verb is fully inflected while the verb of restricted structure is morphologically deficient. In the Yimas chaining structure in (1.20.), for instance, the verb stems *awŋkwi-* and *yampara-* take no actor or tense specification but are marked with the sequential suffix *-mp-* and additionally by the *-i* suffix that marks the verb as dependent.

(1.20.) Yimas (Foley 1986:178)

Mar-#mp-#n awŋkwi-mp-i ant-#nan yampara-mp-i
river-OBL down.in.water-{ANTE}-DEP ground-OBL stand-{ANTE}-DEP
ama-t#paŋ-#
1SG.SBJ-bath-PRF

‘I went down into the river, stood on the ground and washed.’

The nonfiniteness of medial verbs and the adverbial character of the clauses they occur in are often reasons for serious terminological confusion. Haspelmath (1995) tries to resolve the problem of mistaking converbs and medial verbs pointing at the fact that the former ones are examples of subordination and the latter of cosubordination (for similar point see also Longacre 2007:376). However, as it has been already said in section 1.5.1., the application of the notion of subordination as well as cosubordination is by no means unproblematic. Moreover, both in the definition of converb and the definition of

¹⁶ For some references see Thompson, Longacre and Hwang (2007).

chaining, as presented above, participles and gerunds are listed as embraced by the notions.¹⁷ All of this results in a terminological Gordian Knot.

Special verb forms

While discussing adverbial subordination countless authors have referred to *special verb forms* occurring in circumstantial clauses. What these forms are special in is that they either lack (completely or partially) the TAM or agreement categories normally required on a verb in an independent clause in a given language or use some special marking of these categories that is not allowed in independent clauses. The other frequently used adjective for describing these special verb forms is, simply, *nonfinite*.¹⁸ The three types of forms which are most often considered under this label are: (adverbial) participles (or simply: participles, in English confused often with gerunds due to their surface identity), infinitives and subjunctives.

Adverbial participles on their own, by definition, are capable of signalling circumstantial relations between clauses. Nonetheless, the signal is more often than not ambiguous and so, depending on context, the participles may receive a variety of different interpretations: temporal ('when', 'while', 'after'), causal ('because', 'since'), concessive ('although') etc. (see Dik 1997b:155). An example of such a construction with temporal/causal reading in English is:

(1.21.) *Finding the thief, the sheriff felt relieved.*

Verbs in *infinitive form* are also commonly encountered in the world's languages as specialized forms in circumstantial clauses. This seems to be especially common for same-subject purpose clauses which may be marked solely by infinitives without any additional marker of clause linkage. An example from Polish clearly illustrates such a situation¹⁹:

(1.22.) Polish

¹⁷ Givón (1990:865) himself uses participles in English as examples of restricted verbs in chaining structures. He argues that in the following sentence: *Coming out, stopping to check the mailbox, taking a look at the driveway, and pausing to adjust his hat, he turned and marched off* the first four participles are medial verbs.

¹⁸ The traditional distinction between finiteness and nonfiniteness has, however, received in recent years lots of critique due to the problems with defining the cross-linguistic criteria for identifying the phenomenon (cf. discussion and references in Koptjevskaja-Tamm 1994; Haspelmath 1995; Bisang 1998; and Nikolaeva 2007). Cristofaro (2003), addressing this problem has proposed to use the terms *deranked* and *balanced verb forms* instead (the distinction was originally proposed by Stassen 1985). She has used the term *deranking* as referring to "verbal categorial distinctions and/or use of special markers altering the status of the verb (nominal marking such as case endings, adjectival marking such as gender markers) or coding of verbal categorial distinctions not in the same way as in independent clauses" (2003:58).

¹⁹ The same could be said about English if we were to treat the *to* in the translation in (1.23.) as an infinitive marker rather than subordinator as some linguists prefer.

Id-ę do kin-a obejrz-eć film..
go-1SG to cinema-GEN watch-INF movie.SG.ACC

‘I am going to the cinema to watch a movie.’

The SoA encoded in the subordinate clause (‘watching the movie’) is understood unambiguously as a purpose for the SoA encoded in the main clause (‘going to the cinema’). An overt adverbial subordinator is often added (not only in Polish) to the subordinate clause, in which case the purpose relation is signalled in two ways – structurally and lexically:

(1.23.) Polish

Id-ę do kin-a żeby obejrz-eć film.
go-1SG to cinema-GEN {PURPOSE} watch-INF movie.SG.ACC

‘I am going to the cinema in order to watch a movie.’

The third category of special verb forms often encountered in circumstantial clauses are *subjunctives* – i.e. verbs in special mood which mark a clause as expressing something other than the a statement of what is certain (e.g.: a wish, possibility or an action that has not yet occurred). In circumstantial clauses subjunctives are often accompanied by an adverbial subordinator or an adverbializing suffix but this is not a universal rule. In English, for instance subjunctive mood may be used (especially in literary form) together with marked word order to express the meaning of conditionality:

(1.24.) *Were I Mary, I would have done things differently.*

None of these three special verb forms has been discussed in detail in the context of formation of complex sentences involving circumstantial clauses. Moreover, as we might expect, the range of the strategies of encoding interclausal relations which fall under the label ‘special verb form’ and ‘nonfinite form’ is broad enough to cause, yet again, terminological confusion. What is understood by ‘special verb form’, after all, overlaps with the scope of terms such as converbs and medial-verbs.

Coordination of clauses

“*And*” *coordination* (also called *conjunctive coordination* or *syndetic coordination*), understood as a strategy of linking clauses by the use of a marker/particle that renders the meaning of English *and* (Haspelmath 2004a), is another means of conveying the information about asymmetrical relations between SoAs. However, by contrast to the other strategies listed above, “and coordination”, gives a hearer only a very vague idea about the type of relation the speaker has in mind. It seems that, cross-linguistically, the meaning it most commonly implies is that of anteriority or causality, as in the English examples in (1.25a-b) and (1.26a-b) respectively.

(1.25a) *I cleaned the house and I went to cinema.*

(1.25b) *After I cleaned the house I went to cinema.*

(1.26a) *I woke up late and I missed my train.*

(1.26b) *Because I woke up late I missed my train.*

The anteriority and causality readings arise by implicatures, due to the diagrammatic iconicity displayed by language (Haiman 1980) where the structure of language reflects directly aspects of the structure of reality. In (1.25a) this concerns miming the temporal order of events (principle *post hoc, ergo post hoc* – after that, therefore after that) and in (1.26a) miming the fact that events following each other in time are often related causally (principle *post hoc, ergo propter hoc* – ‘after that, therefore, because of that’).

Although iconicity of sequence is clearly the most widespread type of iconicity, the number of circumstantial relations that one can infer from coordinated clauses is, of course, not limited to anteriority and causality.²⁰ In English, for instance, conditionality, relation of purpose and posteriority (and possibly also other relations) may be inferred about from clauses linked by *and*.²¹ This list is, however, not limitless. The limitations result from the likelihood of two (or more) SoAs being related in a particular way. It is unlikely, for instance that the clauses in (1.25a) or (1.26a) could be understood as concessively (or causally or conditionally) related.

Despite the fact that coordination itself has, in recent years, been a subject of a number of studies (see Haspelmath 2004b and Mauri 2008 for list of references), the issue of encoding circumstantial relations using coordinated structures has not been a subject of cross-linguistic enquiry so far. The main reason for that is that coordination is very often merely an additional strategy of indicating interclausal relations – accompanying the more specialized, and less ambiguous ones. Nonetheless, its prominent place especially in spoken communication (cf. sections 10.1.2. and discussion in chapter 11) is a strong argument for counting it among the most important strategies of encoding of circumstantial relations.

Juxtaposition of clauses

Juxtaposition of clauses (also known as *unmarked/zero/asyndetic coordination*) is the most implicit of all the strategies that may be resorted to in order to signal circumstantial relations between SoAs. All other things being equal (including the pragmatic interpretation directed by principles of iconicity) it differs from coordination

²⁰ Some interesting examples of languages for which such iconic motivation is foreign have been reported in South-East Asia. In Burmese, for instance, the unmarked interpretation would be not that of anteriority/posteriority or causality but of simultaneity (see Haiman 1980:533 and references there).

²¹ Cf. for instance *You are late and I am not going anywhere* read as *If you are late I am not going anywhere* or *She went to London and met Jim* read as *She went to London in order to meet Jim*.

in that there is virtually no structural linking element between the juxtaposed clauses – compare (1.25a) with (1.27.):

(1.27.) *I cleaned the house. I went to cinema.*

The only element that may be viewed as indicating some mutual relevance of the SoAs in the two clauses is intonation. As Hopper and Traugott have remarked:

Two juxtaposed clauses (...) with independent intonation contours and without any overt signal of linking do not constitute a single complex (...). However, when juxtaposed clauses are linked in some way, it is reason to think of the two clauses as united grammatically into one sentence by parataxis. (2003:179-180)

The problem of intonation as a marker of subordination is, indeed, very interesting but difficult to investigate, especially when it comes to cross-linguistic comparisons involving lesser studied languages for which the analyses would have to be conducted in the field. This is the precise reason for which such discussions are usually absent from reference grammars.²² The topic has not been well researched for English either. The only work dealing with the subject is the paper by Bolinger (1984). The author has tried to prove not only that intonation is an autonomous mean of expressing adverbial subordination in that language but also that it can indicate degrees of subordination. He has found that the intonation patterns of clauses marked with *if* are exactly the same as in juxtaposed clauses conveying conditional-imperative meaning such as in *If he had all that loot, I would arrest him* and *With all that loot, I would arrest him*. Bolinger has not elaborated, however, on the use of intonation in encoding other types of circumstantial relations in juxtaposed clauses nor has he gone beyond examples from English.

Of the 6 strategies of expressing circumstantial relations described here, converbs, medial verbs in chaining constructions, and the use of special verb forms involve verbal asymmetries and so may be viewed as structurally asymmetrical. Coordination and juxtaposition, on the other hand do not involve any phenomena of this kind, while clauses with adverbial subordinators – depending on the type of verb form they contain – can be either symmetrical or asymmetrical.

²² In November 2008 the subject of the role of intonation and juxtaposition (considered in the context of expression of possession and adjectival modification) was brought to the attention of a wider audience on the mailing list of the Association of Linguistic Typology. There was a general agreement that intonation is a form of “overt marking” and that typologists working on syntax should make more and more use of phonetic information, but, at the same time, it was remarked that the kind of information that would be needed is hardly extractable from published materials.

1.5.3. Thompson and Longacre's overview of strategies

The work that without any doubt occupies the most prominent place in the linguistic literature on adverbial subordination is the paper on adverbial clauses by Thompson and Longacre (1985, co-authored with Hwang in the 2007 edition), which is quoted in virtually every study that deals with this topic. Hence, I find it important to look at what the authors have to say about the strategies of signaling circumstantial relationships between clauses and how they relate to the list of strategies I collected and reported above.

The authors discuss first three major devices for marking adverbial clauses: subordinating morphemes, special verb forms and word order. Within the group of subordinating morphemes they distinguish grammatical morphemes with no lexical meaning (an instance of which is the English *to* as in *to buy beer*) and grammatical morphemes with lexical content (such as *before*, *when* and *if*).

The special verb forms are defined by them simply as verbs which are not used in independent clauses. The examples they give are Latin gerund (adverbial participle) and a defective verb from Wapoo – a Californian Indian language where a glottal stop that normally occurs at the end of a verb (1.28a) in an independent clause, is dropped in subordinate clause of all types (1.28b).

(1.28a) Wapoo (Thompson, Longacre and Hwang 2007:239)

Cephi *šawo paʔ-taʔ*
3SG bread eat-PST
'He ate bread.'

(1.28b) (ibidem)

Te *šawo paʔ-ta-wen,* *ah* *nale ʔš-khiʔ*
3SG.ACC bread eat-PST-*{SIOVER/CAUSE}* 1SG angry-NONFUT
'When/because he ate the bread, I got angry.'

In their description of the special word order which some languages employ in subordinate clauses, they refer to German which puts the finite verb at the end of the subordinate clause while in an independent clause the verb occupies second position in the clause (1.29.). The authors mention also a slightly different example of a word order from Swedish and remark that the position of adverbial clauses is the characteristic feature of some languages.

(1.29.) German (Thompson, Longacre and Hwang 2007:239)

Wir *wohn-ten auf dem* *Lande, wie* *ich* *dir* *schon*
1PL.NOM live-PST on ART.DAT land {COMMENT} 1SG 2SG.DAT already
gesagt *habe*
tell.PST.PTCP AUX.1SG

‘We lived in the country, as I have already told you.’

Following these three cases, the authors enumerate also other means of expressing the same relations as signalled by prototypical adverbial clauses: coordination, juxtaposition, chaining and serial verb constructions which they briefly discuss.

I have a number of reservations about the authors’ presentation of the topic as well as the content of the overview. First of all, by saying “There are three devices which are typically found among languages of the world for marking subordinate clauses” they create the impression that what they are going to discuss are three separate strategies. And this impression is sustained by the descriptions of the devices. It is only by the analysis of examples the authors give that we discover that the strategies are combined. This is most visible probably in the case of word order, which does not seem to be a strategy in its own right – nowhere in the linguistic literature have I found an example of marked word order as a sole indicator of subordination. It is either accompanied by one of the special verb forms (as in 1.24.) or an explicit subordinating morpheme as in (1.29). None of the other examples the authors give presents word order as a linking device in its own right either. Secondly, the authors fail to mention that the subordinating morphemes may be of various forms – bound as well as free and clitic and that cases of discontinuous clause linkers are by no means rare in the world’s languages. All the examples of subordinating morphemes they support the discussion with are examples of English adverbial subordinators. Thirdly, the authors do not mention the category of converbs which, by the time the second edition of their paper appeared, was already a well established term.²³ Finally, the authors list serial verb constructions (SVCs) among other strategies found in languages of the world for expressing circumstantial relations. They write: “A similar example [of an alternative way of signalling relationship between propositions – A.M.] can be found by comparing a language in which a purpose clause is expressed by a subordinate clause with one in which a serial verb construction is used for this function” (2007:242). It is a broadly accepted view now that serial verb constructions are monopredicates and as such are not viewed as forming complex sentences (cf. for instance the papers in the volume edited by Aikhenvald and Dixon 2006, and especially Aikhenvald 2006 for an overview of the subject literature). This remains in direct opposition to Thompson, Longacre and Hwang’s statement that SVCs occur in subordinate clauses and that they may “express a purpose clause”. Serial verb constructions can clearly encode meanings which can be viewed as purposive but this, most certainly, does not concern cases of purpose relations involving two different

²³ The Wappo verb form in (1.28b) can be clearly viewed as converbal.

agents in two different SoAs (i.e. cases which are expressed commonly by non-coreferential purpose clauses).²⁴

On the margin of the discussion here we may also refer to the set of strategies of linking clauses that has been listed recently in the article by Blühdorn (2008:60): connectives (subordinators, subordinating conjunctions, or adverbial subordinators), complementizers, relative pronouns and relative particles, specialized converbs, infinitive verb forms (e.g. infinitives), gerunds and participles, as well as inflectional case forms (e.g. locatives, instrumentals or ablatives). Although the author has considered not only adverbial clauses but also complement and relative clauses, we can clearly see that on the one hand he fails to mention clause chaining, and on the other he lists the use of inflectional case forms which are not strategies of linking clauses but merely one of the sources from which subordinating morphemes arise (see part II of the thesis for numerous examples).

1.5.4. Towards a unified, function-based classification

As I have shown in section 1.5.2., linguists have identified a variety of clause linking devices which, on the one hand, very often overlap with each other in scope (cf. converbs, chaining and special verb forms) and, on the other, co-exist within one clause (cf. for instance adverbial subordinators occurring with nonfinite verb forms). Moreover, as argued in the previous section, the only existing summary of the strategies – the overview presented by Thompson and Longacre in 1985 (which has made it to the 2007 edition of the paper almost unaltered) – fails to present the cross-linguistic diversity of strategies in a correct form.

The classification I propose here has been designed with the purpose of this study in mind – i.e. investigation into the function and origin of items that are direct exponents of circumstantial relations and an analysis of alternative strategies which languages resort to when they lack these designated items. Nonetheless, I believe that it can be successfully used in other studies on expressing circumstantial relations too. There are at least three reasons for this. Firstly, my classification takes into account the variety of phenomena that have been discussed under the heading “adverbial subordination” so far but instead of trying to undo the numerous terminological Gordian Knots, it looks at the issue from a functional perspective. Secondly, each of the strategies I distinguish can be used as a means of encoding of asymmetrical relations

²⁴ The fact of purposive readings of SVCs (as well as the aforementioned phenomena of infinitives as exclusive markers of purpose relations – see section 1.5.2.) can be easily explained by referring Van Valin and LaPolla’s (1997) interclausal relation hierarchy (altered slightly in Van Valin 2005). The authors argue that the closer the semantic relation between two propositions the stronger the semantic link between them. Purpose is placed on the hierarchy as syntactically marked by core cosubordination while majority of other circumstantial relations – including cause (reason) and conditionality – by clausal and sentential subordination and coordination.

between clauses in its own right (although it does not exclude the possibility of combinations of clause linkers with special verb forms). Finally, it takes into account also the least explicit of the strategies that languages employ to convey information about circumstantial relations.

The four groups of strategies of expressing circumstantial relations that I distinguish are:

- a) use of clause linkers;
- b) conventionalized structures;
- c) coordination;
- d) juxtaposition.

Use of clause linkers

The term *clause linker* will be used here as a synonym of *circumstantial-glosseme (c-glosseme)*, which I define as any morpheme or combination of morphemes (be they free or bound) which can be glossed or, in other words, which are dedicated exponents of circumstantial relations between clauses. The notion of clause linker, being very broad, seems, nonetheless, to have been used so far most commonly as a synonym of “adverbial subordinator”. For this reason I have decided to introduce the new term which would refer to the function of a unit rather than to its structural characteristics.²⁵

The term covers a variety of circumstantial clause-linking devices (both native to the language and borrowed as well as combinations of the two) irrespective of their historical origin, morphosyntactic form, internal complexity and type of the verb form they occur with. Moreover, the definition does not restrict the number of circumstantial meanings that a particular linking item serves (the topic is discussed in detail in section 3.3.1.). It is, therefore, suitable for a broad cross-linguistic analysis which have to deal with significant amount of variation in structures and forms.²⁶ Throughout the thesis c-glossemes will be marked in glosses of the examples quoted in curly brackets and using a unified set of abbreviations.

In relation to the terms distinguished in typological works and the notion of c-glosseme embraces:

²⁵ It needs to be noted that it was Bloomfield who first used the term glosseme to describe “the smallest meaningful unit of linguistic signalling” (1933:264). Within glossemes Bloomfield distinguished between lexical units (morphemes) and grammatical units (tagmemes).

²⁶ Note also that such understanding makes the group of c-glossemes very similar to the category of *relators* in Functional Grammar, which mark a relation of dependency linking a dependent constituent to a head and “comprise (i) adpositions, (ii) case markers, (iii) subordinating elements (= either independent subordinating particles or subordinating affixes)” (Dik 1997:398). In the FG framework, however, not much attention has been devoted to these subordinating elements so far.

- the traditional category of free-word adverbial subordinators (1.16.,1.17.,1.23.,1.25b, 1.26b)
- adverbializing affixes and adverbializing particles constituting converbs (1.18.)
- adverbializing particles on medial verbs in chaining structures (1.20.)
- adverbializing morphemes forming adverbial participles (1.21. and the *en* in French *gérondif*)

Three other types of markers, not discussed in section 1.5.2, also fall into the category of c-glossemes:

- affixes encoding circumstantial functions which get attached to finite verb forms, such as the Apache Jicarilla suffix *-go*:

(1.30.) Apache Jicarilla (Jung 2002:176)

koghqyaame'ïïayïï mi-ye' go-ghq-go
Teepee 3OBJ-inside INDF.OBJ-live-{COND}

géh its'a'-ye 'de-'-daa't-je
just middle-in fire-INDF.OBJ-PL-build

'If you live in a Teepee, then you build a fire in the middle'

- distributed markers, clitics and combinations of words and affixes which serve the function of expressing circumstantial relations (see section 3.1. for examples);
- affixes categorized as “dependent/conjunctive moods”.

The term *dependent/conjunctive mood* comes from linguistic literature on Eskimo-Aleut languages (cf., for instance, Harper 1974, Fortescue 1984, Kristoffersen 1992, Reed et al. 1997). It owes its name to the fact that its exponents occur in slots in which in main clauses indicative mood affixes are inserted.²⁷ These dependent moods are used (unlike subjunctives) exclusively in subordinate clauses and encode circumstantial relations such as conditionality, causality etc. In (1.31.) an example of such Central Alaskan Yup'ik dependent mood is given. The *-ng(a)-* affix is a primary strategy of expressing interclausal relation of causality in that Eskimo language.

(1.31.) Central Alaskan Yup'ik (Mather, Meade and Miyaoka 2002:97)

Cikir-nga-mki quya-ut.
give-{CAUSE}-1SG.SBJ.3PL.OBJ glad-IND.3PL.SBJ

'Because I gave (something) to them, the (others) are glad.'

²⁷ Hence, if we wanted to qualify the verb forms with dependent mood as either finite/nonfinite or balanced/deranked we would have to treat them as defective and, thus, nonfinite/deranked.

I found the functions served by these mood markers and the fact that they are restricted to subordinate clauses only good enough justifications to count these circumstantial dependent mood affixes among c-glossems.

Conventionalized structures

Under the heading *conventionalized structures* I include all the structures which are not accompanied by any c-glossems but which, due to displaying certain structural asymmetries on verb forms, are by convention understood as encoding circumstantial relations with lesser or greater degree of ambiguity.

This includes in particular:

- use of converbs, medial verbs and participles which are not accompanied by any adverbializing morphemes (1.19.);
- use of infinitives and subjunctive moods (as in 1.22. and 1.24.) not accompanied by any adverbializing morphemes.

I do not treat marked word order as a conventionalized strategy since, as said in the previous section, I do not know of any language that would use word order as a sole indicator of the circumstantial character of relations between SoAs.

Coordination and juxtaposition

These two strategies do not require any further introduction since they have been discussed in detail in section 1.5.2. and there is nothing we would need to add here. I wish to remark only that from a perspective of a language rich in subordinators, adpositions or converbal endings “and coordination” and juxtaposition may seem to be a non-elaborate way of expressing circumstantial relations – one that speakers rely on mainly when the actual relations between SoAs are not too relevant for a given communicative event. However, as many linguists have recorded in their grammars, these two strategies (and juxtaposition to much greater degree) are in various languages the favoured strategy of expressing cognitive asymmetries between SoAs. This is the case even in those languages which have at their disposal c-glossems (often in the form of adverbial subordinators) and possibility of forming asymmetrical verb forms.²⁸ Speakers of the world’s major languages, including English, often resort to juxtaposition and coordination too, as mentioned before, especially in day-to-day spoken communication. I go back to this issue in chapter 11.

²⁸ Some examples of the languages in my sample that have been said to favour juxtaposition over explicit linking are: Wambaya (Nordlinger 1998), Jingulu (Pensalfini 2003), Jahai (Burenhult 2005), Baure (Danielsen 2007), Didinga (Rosato and Santandrea 1980), Sapuan (Jacq and Sidwell 1999) and I’saka (Donohue and San Roque 2004).

1.6. Summary

We started the chapter in section 1.1. with the introduction of the notion of SoA and clause which are intrinsic elements of the discussion throughout this thesis. In section 1.2. the relations in which the SoAs and clauses may stay in have been divided into symmetrical and asymmetrical and the basic distinction between coordination, subordination, as well as between three types of subordinate clauses: complement, relative and adverbial clauses has been introduced. It was also in that chapter where the rationale behind favouring the notion of circumstantial clause over adverbial clause has been explained. Classifications of the circumstantial/adverbial/interclausal relations proposed in the typological literature have been briefly discussed in section 1.3. and in section 1.4. the 4 relations that this study is focused on have been presented in more detail along with examples.

Section 1.5., which has been devoted entirely to the discussion of the strategies of encoding of the relations has begun with an overview of the problem of cross-linguistic validity of the term *subordination* in 1.5.1. I have emphasized that in the light of the problems with application of the term in broad cross-linguistic studies, in my opinion it is most useful if defined, as Cristofaro suggested, in cognitive-semantic terms. In section 1.5.2. a variety of strategies of linking clauses identified in the world's languages so far has been presented. This has included discussion on adverbial subordinators, converbs, clause chaining, special verb forms, coordination and juxtaposition. On the examples of the last two of these strategies it has been demonstrated that the asymmetrical relations between SoAs can be inferred from clauses which are symmetrically linked. In the following section – 1.5.3. – an overview of the strategies presented in Thompson and Longacre (1985) and Thompson, Longacre and Hwang (2007) has been scrutinized with a brief remark on the list of strategies given by Blühdorn (2008). Finally, in section 1.5.4., taking into account a variety of formal criteria and recent findings and considerations of relevant cross-linguistic studies, I have presented my own classification of the strategies. The major innovation I have proposed is the introduction of a function-based category that covers a variety of items which are dedicated exponents of circumstantial relations between clauses. This category has been called *circumstantial-glossemes*.

The scope of the study, as discussed in this chapter, covers the c-glossemes which are the exponents of the relations of anteriority, causality, purpose and conditionality. The group is looked at in detail and from a variety of angles in part II and part III of the thesis. The other strategies of encoding are recalled, in context of their explicitness, in chapters 9, 10 and 11. The formal features of subordinate clauses, (including configuration of TAM and argument marking on verbs, switch-reference phenomena and alike) are, however, not to be considered any further.

CHAPTER 2

Theoretical foundations and methodology

The main aims of this thesis are, as already presented in the introduction to this work: reconstruction of the most common sources of clause linkers (c-glossems) of anteriority, causality, purpose and conditionality; reconstruction of the universal semantic affinities of these four relations; analyses of cross-linguistic variation in encoding of the relations and an attempt at explanation of the variation. The research this work reports on has encountered many problems similar to those other typological studies have to deal with and it draws heavily on the experience of typological investigations in solving them. This concerns the design of the study, data collection process and analysis.

In the two previous chapters all the necessary definitions have been introduced. Here I focus on the presentation of general theoretical and methodological issues accompanying the analyses. Further details concerning theory (and methods) relevant for each of the analysed aspects are discussed in parts II and III of the thesis.

The present chapter consists of 4 sections. In section 2.1. I discuss the general theoretical foundations of the analysed domain. In 2.2. I focus on the issue of language sampling and present the sampling method applied as well as the full list of languages included. Section 2.3. looks at the topic of data collection and is divided into three parts. In 2.3.1. I present the type of linguistic and extra-linguistic information that has been required for conducting the research and in 2.3.2. and 2.3.2. the two main sources of the data: published materials and questionnaires sent to specialists in particular languages. Finally, section 2.4. offers a brief summary of the information presented in the chapter.

2.1. Theoretical foundations and approach to the analysed domain

2.1.1. Function-based definition of the studied subject

The subject of this thesis involves dealing with data from a variety of languages. The branch of linguistics for which cross-linguistic investigation is daily

bread is linguistic typology which looks out for cross-linguistic patterns, and analyses complete sub-systems of languages as, for instance, agreement, past tense expressions, definiteness marking. The present study does not aim to provide a comprehensive description of the means by which languages encode circumstantial-relations and does not seek for universals or patterns. What it has in common with typological investigation, however, is the interest in cross-linguistic variation, distribution, comparison and the broad question why languages vary as to certain aspects of their systems. Thus, it applies many of the solutions used commonly in typology as discussed below.

The first common problem that every cross-linguistic (including typological) study encounters is the definition of the domain of enquiry which has to consider significant degree of structural variation displayed by the world's languages. As Croft claims:

The fundamental prerequisite for cross-linguistic comparison is crosslinguistic comparability, that is the ability to identify the same grammatical phenomena across languages (...) This is in fact a fundamental issue in all linguistic theory. (2003:13)

A natural inclination would be, of course, to apply formal criteria for the identification of the analysed phenomena. This would, however, exclude a number of languages from the analysis and restrict the validity and the power of the generalizations made, since it is known that purely formal criteria are not universally applicable (cf. Croft 2001, Dryer 1997, Haspelmath 2007). To overcome the problem, typological research has developed a standard research strategy based on **functional criteria**. The strategy may be formulated, again, in the words of Croft, in the following way:

- (i) Determine the particular semantic(-pragmatic) structure or situation that one is interested in studying.
- (ii) Examine the morphosyntactic construction(s) or **strategies** used to **encode** that situation type.
- (iii) Search for dependencies between the construction(s) used for that situation and other linguistic factors: other structural features, other external functions expressed by the construction in question, or both. (2003:14)

This strategy, which has been successfully applied since the 1970's, became to be known as a functional-typological approach, since what it implies is that the link between form and function should be closely considered and, ultimately, that the phenomena under investigation should be defined in functional or functional-structural terms but not exclusively in structural ones.

Since in my research I am interested in the very issue of how a given function is expressed cross-linguistically, it is an additional argument for defining the phenomena under investigation in functional rather than structural terms. Thus, recall that there

have been no structural overtones in the definition of the four relations between SoAs: anteriority, causality, purpose and conditionality presented in sections 1.4.1.-1.4.4. The definitions of the strategies used to encode these four relations I proposed in my classification (section 1.5.4) are also function-related. Even though the definitions of conventionalized structure, juxtaposition and coordination are not free from certain structural add-ons, they are defined with the semantic-pragmatic categories in mind. The definition of c-glosseme – the *tertium comparationis* term presented in section 1.5.4., is, in turn, purely notional (i.e. structure independent). If I were to limit my quest for the origin of clause linking devices only to adverbial subordinators, for instance, the universe from which the data could be elicited would shrink significantly and would be dominated by the languages in which morphology is scant. Since scant morphology is usually a genetic and geographic feature the results obtained in such structure-based research would be additionally biased. The application of the functional approach to the definition of the subject of this study allows effectively for including a variety of forms into the database ensuring the best possible coverage of the investigated domain.

2.1.2. Instruments of analysis and explanatory apparatus

In this study, the inspirations drawn from a functional approach go much deeper – into the analytical tools and explanations of cross-linguistic similarities and cross-linguistic variation.

As Newmeyer (1998:17) has noticed, it is the belief that the analysis of function may reveal the forces shaping the language and the fact that any assumptions about these forces may be investigated and verified only when a large number of diverse languages is investigated that led to functionalists taking the lead in typological research. The functional approach, indeed, assumes that language structure may be explained in terms of language function – this includes cognitive and pragmatic explanations as well as influence of language-external factors.

The two main principles that shape language, and which are responsible for cross-linguistic similarities are, according to functionalists, **iconicity** and **economy** (Croft 2003:201-226). The first of them assumes that the structure (form) of linguistic expressions is motivated (at least partially) by their function. This includes, as Haiman (1980) expounded in his influential paper, **isomorphism** and **motivation**.¹ The principle of economy, on the other hand, assumes the tendency to reduce as much as possible the phonetic substance and the information encoded in linguistic expressions and, as such, is obviously related to the pragmatic theories by Grice (1975) and Sperber

¹ The term *isomorphism* is understood as a one-to-one correspondence between *signans* and the *signatum* i.e. one linguistic form assigned to one meaning. *Motivation*, in turn, refers to the order of elements in language miming the order of elements in the external world.

and Wilson (1987). These two principles: iconicity and economy can be seen as cognition-related and communication-related respectively.

In typological practice these two have been sometimes enriched by considering **external social factors**, resulting in what has been labelled *integrative functionalism* and what has been captured by one of the most prominent functionalists in the following words:

Linguistic phenomena [are considered] systematic, and may be (partly) arbitrary, but they would involve such a close interaction of cognitive and external social factors that one could not reasonably describe the internal cognitive system as self contained. (Croft 1995:516)²

What integrative functionalism effectively denies then is, as Newmeyer (1998:16) has noticed, the Saussurian separation of *langue* from *parole* and synchrony from diachrony. It is here, more than anywhere else, where functionalism shares a number of traits with cognitive linguistics.

All of the phenomena listed here: isomorphism and iconic motivation, economy principle and external social factors are, in one way or another, relevant for approaching the problems analysed in this study. What follows is a brief discussion of their application.

In part II of the thesis my attention is focused on the reconstruction of the origin of c-glossems and the analysis of the semantic affinities within the domain of circumstantial relations. The theoretical background, methodology and assumptions relevant for that part of research are discussed in detail in chapter 3 and so here I wish to recall only what has already been said in the introduction – all the analyses in part II are based on the identification of synchronic patterns of homonymy/polysemy. This concerns both the polysemous syntactic categories that c-glossems overlap with (such as adpositions, case markers etc.) together with their function/meaning and the variety of circumstantial meanings that a given c-glosseme is capable of encoding. It is a well researched and a well known fact that the underlying mechanism behind the emergence of polysemy (and grammaticalization processes in general) is semantically motivated. In fact, it is iconic in nature – similar morphological shapes or syntactic behaviour of categories are icons of their underlying semantic homogeneity (Haiman 1980:517). This becomes especially obvious once cross-linguistic patterns of homonymy/polysemy (also called polyfunctionality or macrofunctionality) are reconstructed since the presence of these patterns – the recurrent use of the same marker for different functions – is viewed as an indicator of their conceptual proximity. Hence, although polysemy may seem a counterexample for the isomorphic “one form-one meaning” principle it is

² We find elements of integrative functionalism in the works of Bybee (1985), Du Bois (1985 and 1987), Givón (1990) and Hopper (1987) to name just some of the authors.

explainable by the iconic mechanism of our cognition which translates into grammaticalization processes. It is precisely in this point where the issue of economy enters the discussion. As Hopper and Traugott have remarked:

The optimal language would be one in which every meaning was distinct, just like every numeral is distinct (...). However, such “optimality” would clearly in actual fact be dysfunctional since there are far too many meanings for the brain to remember individual expressions for them. “One form – one meaning” is an ideal on the dimension of choice of form and the motivation to maximize information. It is balanced and offset by another optimality, that of associating like forms with like meanings, in other words, of developing polysemies. (2003:78)

Hence, isomorphism and economy are the two principles which not only allow us to gain insight into grammaticalization processes from a strictly synchronic perspective by analysis of the effects of their exploitation, but which allow us also to understand the cognitive mechanisms behind the presence of cross-linguistic patterns of polysemy. Moreover, the surface effects of these two processes in the form of c-glossems capable of expressing more than one circumstantial meaning allow us additionally to gain an insight into the cognitive organization of our minds, as discussed in more detail in section 3.3.

In part III of the work, in turn, I look at the motivations behind the cross-linguistic variation in the degree of grammaticalization, lexicalization and explicitness of particular types of c-glossems. In answering questions such as “why do some languages have a dedicated, fully grammaticalized conditionality c-glossems and others don’t?”, “why in some languages these are lexicalized?” and “why do some languages seem to be more explicit than others?”, the cognitive as well as language-external (socio-cultural) factors are discussed in the spirit of functionalism as an explanatory apparatus. The issue of economy – pragmatic motivation – is, understandably, considered in that part of the analysis too and it is linked to the discussion on iconic motivation behind the phenomena of marking certain circumstantial relations by coordination and juxtaposition – i.e. following the order of SoAs in the real-word (cf. section 1.5.2.).

The broadly understood functionalism – the mixture of cognitive, pragmatic and extra-linguistic elements and the focus on usage-based account of language structure – are then the main instruments with which this thesis approaches the problems it has aimed to address.

2.2. Language sample

The problem of choosing languages for one's cross-linguistic study is by no means trivial since a bad sampling technique may significantly affect the results of the planned investigation. For this reason it is important to choose such a sampling technique that will ensure that various biases (including the genetic and areal one in the first place) are minimized. Every decent research project whose questions require collecting data from a variety of languages acknowledges this problem. It comes as no surprise, therefore, that the issue has received a lot of attention and a number of language sampling techniques have been proposed so far (see for instance Bell 1978; Dryer 1989 and 1992; Dahl 1985; Perkins 1989 and 1992; Ramat 1998; Rijkhoff et al. 1993; Rijkhoff, Bakker 1998; Bybee, Perkins and Pagliuca 1994). The techniques have been recognized, in their own right, as important instruments bringing us closer to the goal of understanding rules and principles that are responsible for the differences and similarities among the world's languages (Rijkhoff and Bakker 1998:305).

In the practice of sampling, two main types of samples can be distinguished: *probability samples* and *variety samples* (see Rijkhoff et al. 1993:171; Croft 2003:23; Bakker, forthcoming). The first type of sample is used when a study aims to determine significant correlations between grammatical traits and so it is important for the languages to represent independent cases i.e. to be unrelated in terms of geographic distribution, genetic affiliation etc. A variety sample, by contrast, is designed to capture diversity (full range of linguistic variation) and so the languages are selected from different genetic families (i.e. such that, in principle, evolved independently).³ At the same time the design of a variety sample should ensure maximal geographic dispersion of languages to control for areal diffusion of typological traits.

In this study a variety sample designed using the recognized sampling method proposed by Rijkhoff et al. (1993) has been used. The method is the only fully formalized general sampling technique proposed in the typological literature to date and is based on an algorithm called the *Diversity Value calculation*. The universe from which the sample is taken consists of all known – living and extinct – languages; all language families (phyla) are represented by at least one language and all language isolates are included. This minimal sample may be then extended on the basis of calculations of the diversity of languages in particular family trees. Each node of a given tree is assigned a diversity value which expresses the complexity of the tree below the node taking into account the number of sub-nodes and their arrangement. The value is then taken as a determinant of the number of languages to be drawn from a

³ In the case of larger samples, more than one language may be selected from a single genetic family. However, in such situations attention should be paid to make sure that the languages belonging to one family come from as distantly related branches of that family as possible.

given node and used in the final calculation which considers the total number of languages one wishes to use in a given study. The method was originally demonstrated on Ruhlen's (1987) classification of languages⁴ but, as the authors argued (and proved in Rijkhoff and Bakker 1998), it is suitable for applying to other classifications (such as Ethnologue's or Voegelin and Voegelin's 1977) too.

For the purpose of the research reported on in this thesis, a geographically and genetically balanced sample of 100 languages applied to Ruhlen's classification has been chosen as a starting point and altered slightly. The principal change I have introduced was the exclusion of the extinct language families including extinct language isolates. The main rationale behind this decision was theoretical – since the socio-cultural setting of the extinct languages was very different to the one languages find themselves in nowadays, the former ones could not be considered in the discussion on the reasons behind cross-linguistic variation in the form and explicitness of c-glossemes (see part III of the thesis). Additionally, on many occasions a practical issue of availability of quality reference grammars and availability of consultants (see next section) has also determined the decision of exclusion of the extinct languages. For the same practical reason several other language families had to be excluded from the initial sample. Considering that two thirds of the known languages still have not been described at any level of linguistic sophistication (Bakker, forthcoming), the availability of data very often determines the final list of languages included in cross-linguistic studies. As long as in choosing languages for one's sample (especially when a world choice of languages is planned) the issue of convenience – i.e. use of materials which are readily available – does not overwrite the effort to ensure the genetic (and geographical) representativeness and balance of the sample, the scientific generalizations drawn from the analyses of the sample are valid. In this study the principles of representativeness and balance have been meticulously adhered to.

Apart from the exclusion of the extinct languages and language families for which neither good quality grammars nor consultants have been found, one Niger-Kordofanian and one Afro-Asiatic language have also been removed from the initial sample since in the final stage of data collection contact has been lost with the consultants and I was not satisfied with the quality of the data on those languages which I elicited from published materials.

The final number of the languages in the sample, after introduction of the abovementioned changes, has been brought down to 84. All the languages are listed in (Fig. 2.1.) below and their geographic distribution is depicted in (Fig.2.2.).

⁴ Ruhlen's classification has been criticized by several authors because of its "mass lexical comparison" approach and the disputable status of some of the phyla it distinguished – especially the Amerind and Indo-Pacific phylum (see Rijkhoff et al. 1993:1999). However, as Blake (1988) noticed – any worldwide genetic classification will be controversial.

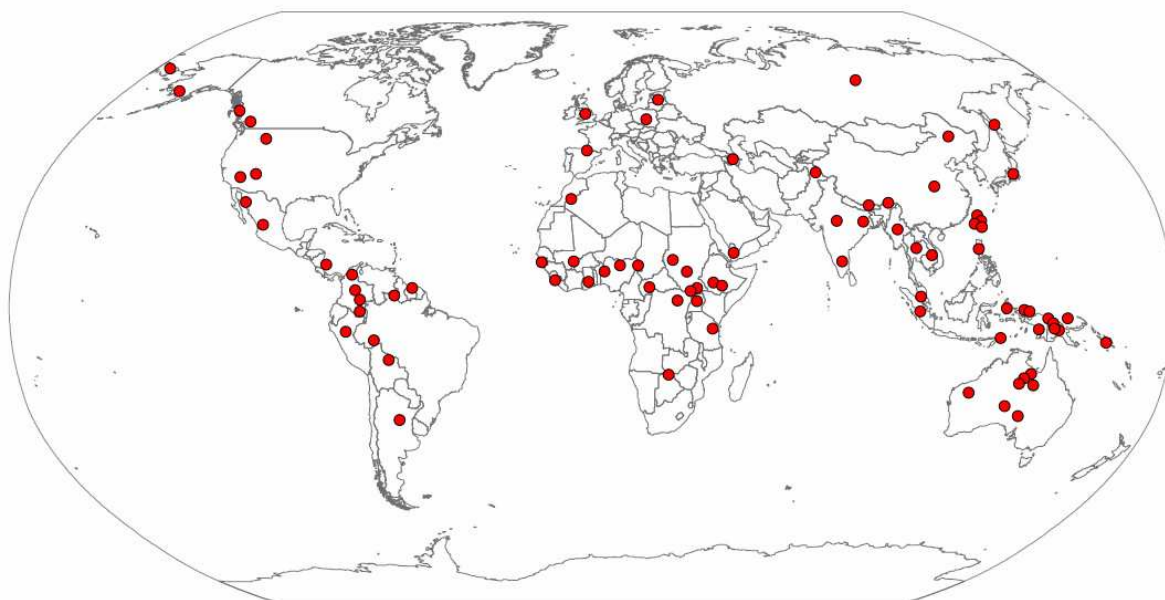
Langue phyla**Languages' names**

AFRO-ASIATIC (5/6)	Arabic San'ani, Hausa, Konso, Maale, Shelha (Douiret dialect)
ALTAIC (2/2)	Japanese, Dagur
AMERIND (16/18)	Achagua, Apurina, Baure, Cubeo, Hualapai, Ika (Arhuaco), Letuma (Retuarã), Lillooet (St'át'imcets), Macushi, Mocoví, Nisga'a, Nez Perce, Quechua Huallaga, Rama, Seri, Southeastern Tepehuan
AUSTRALIAN (7/7)	Arabana, Jingulu, Pitjantjatjara, Wambaya, Warlpiri, Yanyuwa, Yindjibarndi
AUSTRIC (13/14)	Batak Karo, Ilokano, Jahai, Leti, Rukai (Mantauran dialect), Paiwan, Santali, Sapuan, Seediq, Taba (East Makian), Thai, Vitu, Yami
CAUCASIAN (1/1)	Lezgian
CHUKCHI-KAMCHATKAN (1/1)	Chukchi (Telqep dialect)
ELAMO-DRAVIDIAN (1/1)	Tamil (standard form)
ESKIMO-ALEUT (1/1)	Central Alaskan Yup'ik
INDO-HITTITE (3/4)	English, Polish, Hindi
INDO-PACIFIC (8/13)	Ama, Au, Eipo, Hatam, I'saka (Krisa), Lavukaleve, Meyah, Yimas
KHOISAN (1/1)	Khwe
NA-DENE (1/1)	Apache Jicarilla
NIGER-KORDOFANIAN (8/9)	Akan, Boko, Gola, Krongo, Mayogo, Swahili, Sango, Suppyire
NILO-SAHARAN (5/5)	Didinga, Fur, Kanuri, Lango, Ma'di
PIDGINS AND CREOLES (2/2)	Kryiol, Ndyuka (Aukan)
SINO-TIBETAN (4/4)	Galo, Kayah Li, Lepcha, Mandarin
URALIC-YUKAGHIR (1/1)	Estonian
LANGUAGE ISOLATES (4/9)	Basque, Nivkh (Gilyak), Ket, Burushaski

(Fig.2.1.) The sample⁵

The numbers in the brackets indicate for each of the phylum how many of the languages required in the initial 100 sample have been included in the final sample.

⁵ The list with complete details on the genetic affiliation of the languages is given in Appendix I.



(Fig.2.2.) Distribution of the sample languages

Since the number of languages in each of the language families is not big enough to allow for drawing reliable conclusions concerning certain patterns within each of the families such conclusions are rather avoided in this thesis. Occasional observations concerning language families are discussed in part III of the thesis as well as in chapter 9 where the results reported on in part II are compared. Observations concerning geographical patterns are discussed more widely in part III of the thesis. This concerns especially the correlation between the cross-linguistic variation in encoding of circumstantial relations and the socio-cultural specifics of certain regions of the world.

In relation to the analysis of the influence of socio-cultural factors on language structure (see part III) it needs to be born in mind that the sampling method is designed for linguistic and not sociological or anthropological purposes. Hence, it is not expected to be balanced when it comes to the socio-cultural profiles of the languages. The extra-linguistic aspects and the linguistic aspect of sampling will never be possible to combine in such a way that would ensure balanced representation of both of them simultaneously. To explain the reasons behind this claim we may look at the distribution of the number of speakers across languages since population size is, more often than not, correlated with the development of other cultural traits, including advances in technology, education, medicine etc. According to the Ethnologue (2009), less than 1.2% of the world's population speaks what amounts to 80% of the languages. This means that the socio-cultural profiles assigned to particular speech communities are very strongly biased towards the ones which do not reveal the aforementioned

advances.⁶ The only way in which we may take the socio-cultural variation into account in choosing languages for a sample is to try to pick languages of speech communities displaying as various socio-cultural profiles as possible. However, this is not easy for two reasons. The first one is that societies speaking languages belonging to one language family very often display very similar extra-linguistic characteristics. The other one is that languages of bigger communities are usually better described and as a linguist has to, in the first place, make sure that he uses good quality linguistic data, he would naturally go rather for the more detailed description. These are the problems I faced while designing the sample for my research too. Only in several cases I managed to find good quality published sources for languages of various socio-cultural profiles within one language family. I have partially overcome the problems by choosing nonparametric statistical tests (i.e. designed for samples with non-normal distribution of values) for analysis of correlations between socio-cultural factors and linguistic phenomena (see chapter 10).

2.3. Data collection

The extensive data required to conduct analyses necessary to answer the research questions stated in the introduction to the thesis and recalled at the beginning of this chapter have been elicited from two types of sources: published materials and analytical questionnaires sent to linguists who are specialists in the particular languages included in the sample. Both of these sources are described in sections 2.3.2. and 2.3.3. below. However, before we move onto discussing them I believe that the reader would find it useful to learn more about the type of information that I have been seeking and that this knowledge will allow him to better understand the choices I made. Thus, in section 2.3.1. I present a concise overview of the types of data that I have been collecting.

2.3.1. The database

For each of the almost 700 c-glossems included in the analysis in this study information on the following has been sought for:

- form (free word, affix, clitic, combination of words and affixes, discontinuous linker);

⁶ Of course, the advances are not the only element of socio-cultural reality that one may want to look at. Others may include, for instance, type of agriculture, kinship relations within a community etc. However, for these and similar traits it is not possible to easily interweave them into a language sample either. In any case, I wish to emphasize that differences in cultural complexity (regardless of how we define cultural complexity) do not, in any way, imply or determine mental inequality between the populations considered.

- morphological complexity i.e. mono- vs polymorphemicity and – in the case of polymorphemic markers – also information on the internal make-up;
- origin (i.e. is the linker original to a given language or borrowed);
- homonyms/polysemes of the c-glosseme in other syntactic categories;
- other circumstantial meanings encoded by the marker;
- additional information concerning restrictions of usage (e.g. occurring only in same-subject or different-subject clauses in the case of purpose markers).

For each of the 84 languages the following details concerning the socio-cultural profile of the speech communities have been needed:

- number of speakers who speak the language as their first one;
- type of society the language is spoken in;
- level of written form development;
- presence of the language in school teaching;
- presence of the language in radio and TV broadcasting;
- information about the languages with which a given language stays (or stayed) in close contact.⁷

The data for each of the languages have been entered into a Microsoft Excel file with separate fields devoted to the socio-cultural profiles and each of the four relations analysed. The fields have been divided into subfields which have covered all of the specific groups of information described above. The in-built options of Excel provided the basic tools necessary for managing the database including highlighting, sorting, and searching the entries. Due to the non-numerical character of the entries, however, almost all the calculations had to be performed manually. Using a dedicated data base designed specially for the purpose of this study would clearly have made the process of data analysis easier and more efficient but the costs of preparation of such a program (including the time investment) would have been too high and therefore the more readily available solution has been applied.

Since not for all of the c-glossemes and not for all of the languages all of the desired information has been obtained, the database does not claim the right to be called complete. Occasional errors in interpretation, coding, and entry of the information are inevitable. As the author of this thesis, I take full responsibility for all of these and other imperfections as well as for the effect they might have on the results of this research.

⁷ All of these elements are discussed in detail in section 10.2.

2.3.2. Source 1: published materials

As in the majority of cross-linguistic studies, also here the main and primary sources of information on the investigated topic have been published reference grammars, grammatical sketches and doctoral dissertations providing descriptions of the particular languages. Despite the fact that an effort has been made to choose the languages for which good quality materials are available, there have been some recurrent problems in the elicitation of the data required for the purpose of the study. The major issue has been the identification of homonyms/polysemes of particular c-glossems. This kind of information is only rarely given directly in the parts of grammars discussing formation of adverbial clauses and so eliciting it required in the vast majority of cases tedious searches through the volumes in order to identify the homonyms/polysemes.⁸

The second major problem has been the determination of the morphological complexity of particular c-glossems and the morphemes incorporated in the structure of those of them which are polymorphemic. More often than not I had to simply deduce the information and then ask the consultants for the verification.

On many occasions, in addition to the reference grammars and grammar sketches, available dictionaries of the languages – sources by definition especially useful for identification of homonyms/polysemes – have also been consulted. The quality of these sources, however, was not always satisfactory. For instance, dictionaries more often than not list only lexical items omitting affixes (or combinations of words and affixes) which, as already said in chapter 1, are by no means rare among clause linkers. Moreover, in many cases they do not provide the full list of homonyms/polysemes and in extreme cases, they give only English equivalents of particular words without part of speech classification so that one is unsure whether the word for “after” functions as an adposition, an adverb or a clause linker. Even after consulting collections of texts from the languages in the sample (published separately or as appendices to reference grammars), journal articles and papers in volumes devoted to clause combining and clause linkers many questions still remain unanswered.⁹

The information on socio-cultural profiles of particular speech communities has been gathered from a variety of publications, the most important of which have been the introductions to the reference grammars, Ethnologue (2009), and the “Dictionary of

⁸ The only grammar which made the process very straightforward was the grammar of Lezgian by Haspelmath (1993) which, apart from giving information about homonyms and polysemes of many of the clause linkers in the chapter on adverbial clauses includes also a list of English-Lezgian vocabulary and a very useful index of affixes with annotations.

⁹ The complete list of linguistic publications consulted for each of the languages, together with the names of consultants (see next section) is to be found in Appendix II.

Languages” by Dalby (2006).¹⁰ In addition to these, a variety of online sources dedicated to particular speech communities and languages have been used as a source of information too. Even with their help, however, additional questions very often had to be sent to the consultants – this concerned especially questions about the presence of radio and TV broadcasting and history of written tradition which are really discussed in the available materials in the amount of detail required for this study.

2.3.3. Source 2: the questionnaire

The second source of data, whose role would be difficult to overestimate, has been a questionnaire sent to the specialists in the respective languages. In the pilot data collection on 15 languages the questionnaire was designed as a combination of elicitation and analytical questionnaire. The consultants were asked to translate a number of structures from English and to answer a list of specific questions concerning clause linkers. However, as it often happens in this type of research, the questionnaire has been modified as the feedback has been received. In its final version it has been devoid of the elicitation part completely¹¹, focusing on language-specific questions instead. The main aims of the latter version of the questionnaire can be summarized in the following way:

- to verify the information I collected analysing published materials;
- to supplement the linguistic information;
- to obtain missing information on socio-cultural profiles of the analysed languages.

The method of data collection I arrived at consisted of three stages: a) reading of the available materials, b) preparation of summaries of the elicited information, c) sending the summaries (with the request for verification) along with the lists of remaining questions to the consultants.

The type of information I have been seeking determined also the profile of the consultant – since a significant level of purely linguistic knowledge was required to answer the questions, non-linguists had to be automatically excluded. Although only a small number of my consultants are native speakers of the respective languages, I was very lucky that many of the highly respected and experienced researchers, including many authors of highly valued reference grammars, agreed to help me in my data collection. In total the questionnaires were filled in by 70 consultants for a total of 64 languages.¹² The verification of validity of the information collected from printed

¹⁰ For the other ones see section 10.2.

¹¹ The main reason for excluding elicitation was that the consultants found it too time-consuming and many of them pointed that many of similar sentences can be found scattered around the chapters on complex sentence formation and samples of texts in the relevant reference grammars.

¹² Although it needs to be emphasized that the questionnaire has never constituted the only source of data for a language.

materials was especially important since much of the data needed, as discussed in the previous section, had to be simply deduced. The consultants not only verified the information (sometimes using evidence from historical studies that I would not be able to look into given the time and scope restrictions of this study) but on several occasions also elicited additional information from their native informants. On the whole only for two of the 84 languages: Lezgian and Lango the reference grammars of truly excellent quality provided all the information needed without the necessity of verifying and supplementing it with the help of consultants.

As the reader may expect from what has been said here, the questionnaire had to be tailored for each of the languages separately which may give an impression that the method used was, in fact, not that of a questionnaire (which by definition, contains a standard list of question), but of an research interview. Nonetheless, since there has been a clear set of recurrent questions the consultants were asked, I believe it is still sound to call the method a questionnaire method.¹³

2.4. Summary

In this chapter, the general theoretical background of the study built on the assumptions that language structure can and should be investigated with relation to language function has been presented in section 2.1. It has been said that the study draws on the instruments of analysis and explanatory apparatus of functional (and functional-typological) approach which includes: lack of assumptions about universal formal categories; interest in identifying cross-linguistic patterns and explaining them (grammaticalization), usage-based explanation of language structure with special emphasis on cognitive (iconicity), pragmatic (economy) as well as socio-cultural forces shaping language.

In section 2.2. the importance of good sampling in cross-linguistic studies has been discussed along with a brief description of the sampling technique used in this work – Rijkhoff et al. (1993) Diversity Value calculation. The variety sample consisting of 84 languages has been presented and the reasons for introducing changes to the initial sample of 100 languages have been explained. Finally, the problems of balancing socio-cultural profiles of speech communities within genetically and geographically balanced samples of languages have been addressed.

Data collection issues – including information on the database and the collection methods have been presented in section 2.3. The difficulties of searching for

¹³ The recurrent list of questions is listed in Appendix III.

clues on diachronic processes in materials discussing synchronic state of languages have been presented and the need for verifying and supplementing the information elicited from those sources has been emphasized. Finally, the analytical, language-specific questionnaires filled in by specialists in the analysed languages have been discussed.

PART TWO

SOURCES AND SEMANTIC SCOPE OF C-GLOSSEMES

The perspective of variationist typology to connectives positions them in a variational space in which polyfunctionality and fuzziness are inherent features.

Ursula Lenker and Anneli Meurman-Solin

The following part of the thesis has two main aims. The first one is to reconstruct, on the basis of available synchronic evidence, lexical and/or grammatical sources from which c-glossems of anteriority, causality, purpose and conditionality most commonly develop in the world's languages. In the reconstructions of grammaticalization pathways two methods are used: analysis of cross-linguistic patterns of polysemy and analysis of the material incorporated in the structure of polymorphemic markers. The discussion is, on numerous occasions, supported by evidence from other cross-linguistic and language-specific studies devoted to the topic of the emergence of clause linkers. The second aim of this part of the thesis, inherently related to the first one, is to investigate the semantic space of c-glossems by revealing semantic affinities between the circumstantial relations that the particular markers encode. Also here the cross-linguistic patterns emerging from the observations are of special interest.

In chapter 3. the reader finds background information on theoretical and practical issues concerning the task undertaken here, as well as a description of particular parameters of analysis, terminology and methodological solutions that are applied in chapters 4 (Anteriority), 5 (Causality), 6 (Purpose) and 7 (Conditionality). Each of the four chapters is divided into four sections: discussion of morphological complexity and forms of a particular group of c-glossems; presentation of the results of the investigation into the origin of the linkers; more detailed discussion of semantic extensions of meanings in the domain of circumstantial relations and presentation of the findings concerning affinities between the circumstantial concepts encoded by the linkers; and a short summary. In chapter 8 the results of analyses are compared and general conclusions are drawn.

The focus of this part is exclusively on the c-glossems themselves. Discussion on the functioning of the linkers in particular languages as well as on the motivations for the emergence of c-glossems are presented in part III.

CHAPTER 3

Background and parameters of analysis

3.1. Forms and morphological complexity of c-glossemes

The first two elements that I wish to look at before I approach the more detailed issues of origin and functioning of c-glossemes, is their form and internal complexity as revealed in the analysed sample. The issue has been briefly mentioned already in chapter 1. Since it is relevant for the problems considered in the later chapters, here the subject is addressed in a more systematic way.

Regarding the form, the items are classified into the following types: **free words**, **affixes**, **linear combinations of words and affixes**, **distributed** (i.e. non-continuous) **markers** and **clitics**. For each of the four circumstantial relations a detailed summary of distribution of these forms in the data set is presented in sections 4.1., 5.1., 6.1. and 7.1. respectively. In this place I restrict myself to giving some examples of these forms:

- a) free words (see also, for instance, examples 1.16.,1.17.,1.23.)

(3.1.) Ndyuka (Huttar and Huttar 1994:119)

Efu mi tyai en te doo a boto, mi o siki
{COND} 1SG carry 3SG.OBL until arrive ART.SG boat 1SG FUT sick

‘If I carry it all the way to the boat, I’ll be sick.’

- b) affixes (including prefixes as in 3.2., suffixes as in 1.18.,1.28b,1.30, circumfixes as in 3.3. and non-final bound morphemes as in 1.20,1.31.)

(3.2.) Krongo (Reh 1985:349)

n-áfàlàŋ àʔàŋ níinò m-áamàamà
1/2-IMPRF.open 1SG mouth {PURPOSE}-INF.yawn

‘I am opening the mouth in order to yawn.’

(3.3.) Burushaski (Tikkanen 1995:494)

Iné garoóni nu-mú-çu-n dayóan
that.H bride {ANTE}-3SG.F.DOBJ-take-{ANTE} flour
du-mó-sku-n, móo-dil-um-an.

D-3SG.F.DOBJ-lower-{-ANTE} 3SG.F.BEN-throw-STPTCP-H.PL.SBJ

‘Taking the bride along, they threw flour on her [to welcome her], having helped her down [from the horse].’

c) linear combinations of words and affixes¹

(3.4.) Basque (Hualde, Ortiz de Urbina 2003:744)

Behin lana amaitu-z gero, gusta-tzen zaio lagunekin ardo
once job finish-{-ANTE} like-IMPRF AUX friends.with wine
pare bat har-tze-a
pair one take-NOM-DET

‘After having finished work, she/he likes having a couple of glasses of wine with her/his friends.’

d) distributed markers

(3.5.) Khwe (Kilian-Hatz 2008:327)

Nò teá kú lám-lōā à wòó nò
{COND/SIOVER} 2SG.M be little time-DIM OBJ have {COND/SIOVER}
nllòmàá khòè-teè ti à llxáo è!
please.2SG.M.VOC 1SG OBJ help IMP

‘If/when you have a little time, please help me.’

(3.6.) Yanyuwa (Kirton and Charlie 1996:134)

Karna-wuluma nyala wariba-ntha-lu wurnda-a
1SG-run {PURP1} climb-PTCP-{-PURP2} tree-ABL

‘I ran to climb a tree.’

e) clitics

(3.7.) Yindjibarndi (Wordick 1982:185)

Nyinta ngarri-ngu pampa yaala=yhu warrung-ka=yhu mirta-wa pampa ngarr-ii
You sleep-IMPRF now={COND} night-LOC=DET not-EMP sleep-POT

‘If you sleep now, then you won’t sleep tonight.’

I wish to emphasize here that I second Dixon and Aikhenvald (2002) in their argumentation that it is impossible to introduce a cross-linguistically universal set of criteria for distinguishing between words, clitics and affixes. Consequently, when it comes to these three notions, in this study, I repeat the labels attributed to particular clause linkers by the authors of grammars of particular languages and the consultants who helped me in the data collection.

¹ Note: only c-glossems made up of free forms and suffixes that are not morphologically bound qualify as combinations. Hence cases of inflected nouns such as *ostean* in Burushaski consisting of the noun *ostea* (‘back’) and the locative case ending *-an* are not considered combinations but polymorphic, lexical markers.

The second initial element which is relevant for the analysis throughout chapters of part II is **synchronic internal complexity** of the c-glosseemes. The linkers are classified into two main groups: the simplest c-glosseemes, consisting of just one morpheme, are called **monomorphemic** (cf. examples 3.1.,3.2.,3.7.) and those consisting of more than two morphemes – **polymorphemic** (cf. examples 3.3.,3.4.,3.5.,3.6.). In those few instances where historical evidence on the formation of polymorphemic markers was readily available, it had, understandably, priority over the conclusions drawn from synchronic observations. Such is the case with the English *because* which is seemingly (in its orthographic form in UK English and its orthographic and phonological form in American English) made up of the copula *be* and the noun *cause*. We know, however, that it developed from the combination of preposition *by* and noun *cause* (cf. the entry for *because* in OED, for instance). Since this make-up is not evident from its recent form, *because* is treated here as synchronically monomorphemic.

The group of polymorphemic markers is classified, where possible, according to the number of morphemes: two (bimorphemic), three (trimorphemic), four (quadrimorphemic) etc. The exact make-up of the polymorphemic markers is the subject of analysis in sections 4.2.2, 5.2.2, 6.2.2 and 7.2.2. Importantly, on numerous occasions in the case of polymorphemic markers we are dealing with so called **double-marking** i.e. two clause linkers on their own are capable of expressing a circumstantial relation of a given kind occurring in one clause (they may be of the same form, as in the Khwe example in (3.5.) or of different, as in (3.8).

(3.8.) Khwe (Kilian-Hatz 2008:328)

NgýĚĚ *ti winà-à-gòè nò,* *to †ũú-à-gòè ngú à*
 {COND} 1SG be rich-1-FUT {COND/SIOVER} 1SG buy-1-FUT house OBJ

‘When/if I am rich, I will buy a house.’

In some cases, one of the two linkers in such a double-marking configuration is a borrowed word. This are clearly marked in the discussion on material incorporated in polymorphemic linkers.² Furthermore, in cases where a linker which is capable of encoding a given circumstantial relation on its own may also form a part of a more complex linker, these are counted as two separate c-glosseemes. Finally, any borrowed clause linker is always qualified as monomorphemic since it should be viewed as such from the perspective of the borrowing language.

² The issue of borrowing is discussed separately in the context of explicitness of encoding of the four relations in section 9.3.4.

In the Chibchan language Rama we find a suffixal purposive marker *-bang* (3.12a) which is identical to the prospective aspect marker (3.12b):

(3.12a) Rama (Craig 1991:457)

Tiiskama ni-sung-bang taak-i
baby 1SG-see-{PURPOSE} go-TNS

‘I am going in order to see/look at the baby.’

(3.12b) Rama (Craig, 1990, chapter 9:18)

Bal-n-aating-bang
PREF-1-talk-PROSP

‘I am going to talk.’

And in Lepcha we notice that the purposive clause linker *-ká* is identical in form with the locative case marker (3.13.)

(3.13.) Lepcha (Plaisier 2006:126)

Hu cho rok-shang-ká yánthó-ká nóng ma
3SG book read-INF-{PURPOSE} school-LOC go AST

‘He went to school in order to study.’

Finally, it will certainly strike the observer that some particular combinations of these various polyfunctionalities occur in a variety of languages, creating what is called here *patterns*.

Both the fact that we encounter numerous examples of polyfunctionality (a topic dealt with in detail in section 3.2.3.) and patterns of these polyfunctionalities are only rarely accidental. In the vast majority of cases the identical items can be viewed as polysemes and their independent emergence in a variety of languages can be explained by processes of *grammaticalization*.

3.2.1. Grammaticalization and its mechanisms

In the last three decades, grammaticalization has been a hot topic in linguistics. Among the most prominent works in this domain, which have looked at the phenomenon of subordination more closely, we should enumerate the pioneering publications by Givón (1979), studies by Lehmann (1982), Saxena (1988), Traugott and Heine (1991b), Hopper and Traugott (2003 [1993]), Bybee, Perkins and Pagliuca (1994), Harris and Campbell (1995), Bybee and Noonan (2001), Diessel (2005), and finally works by Heine and Kuteva (2002, 2005, 2007). It would be far beyond the scope of this thesis to give even a brief overview of all that has been said about

grammaticalization of clause-linking markers in these publications. Hence, I limit myself to the presentation of the ideas that are most relevant for this study.

Grammaticalization is most commonly understood as “the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions” (Hopper and Traugott 2003:232).³ There is, however, far less agreement between linguists as to what sub-processes and mechanisms constitute grammaticalization. Consequently, there is no theory of grammaticalization that all would be ready to adhere to.⁴ The version of the theory I am most convinced by, and which I choose to follow here, is the one proposed by Heine and Kuteva (2002 and especially 2007). Starting from the most general issues, I follow the authors in their depiction of grammaticalization as a process that typically involves four mechanisms:

1. **desemantization** (semantic bleaching, loss of meaning);
2. **extension** (context generalization, use in new context);
3. **decategorialization** (loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms);
4. **erosion** (phonetic reduction, loss of phonetic substance).

Heine and Kuteva’s list of mechanisms has an important additional advantage here – it covers effectively most of the other processes mentioned often in the context on grammaticalization including metaphorical transfer, metonymic transfer and reanalysis. And so metaphor and metonymy qualify as cases of desemantization while *reanalysis*, (understood widely as the complex process where the syntactic, morphological and semantic properties of forms are modified without modifications to their phonological shape) is embraced by Heine and Kuteva’s mechanisms 1-3.⁵

The term grammaticalization refers then to a macro-change which is an interaction of pragmatic, semantic, morphosyntactic and phonetic factors. Moreover, these sub-processes are often viewed as following each other in time in a particular order: “Grammaticalization tends to begin with extension, which triggers desemantization, and subsequently decategorialization and erosion” (Heine and Kuteva 2007:35). Not all of the four mechanisms are *conditio sine qua non* for grammaticalization to occur and among them it is erosion that quite often does not take place which results in the emergence of polysemy (see next section). The pragmatic

³ For a list of various definitions of grammaticalization proposed in linguistic literature see Campbell and Janda (2001).

⁴ Cf. the versions presented in Lehmann (1982); Heine, Claudi, and Hünemeyer (1991a); Bybee, Perkins, and Pagliuca (1994) and Hopper and Traugott (2003).

⁵ For discussions on the components of reanalysis see, for instance: Langacker (1977:58), Traugott and König (1991) and Hopper and Traugott (2003:39).

element (context) and the semantic element, on the other hand, are of special importance since they license grammaticalization (Traugott and Heine 1991a:8).

Putting together what has been said here, grammaticalization is to be viewed as a process which is triggered by a context-induced reinterpretation of a linguistic unit and it is the semantic content of that unit – the motivation referred to in section 2.1.2 – that makes it suitable for this reinterpretation. As a result, readily available units are used for novel purposes. The frequency of use of these units in their new function allows for the change to spread in the population of speakers (see Heine and Traugott 1991a:9, Bybee and Hopper 2001).

3.2.2. Polysemy

Whenever a grammaticalization process occurs without phonetic reduction, it results in the presence of phonologically identical items sharing their etymology but encoding different meanings and, possibly, belonging to different syntactic categories. This phenomenon is widely known as *polysemy*.

A simple example of an emergence of polysemy in the case of clause linkers is the history of the aforementioned English *after*. In the oldest historical records *after* was used exclusively as an adverb indicating temporal and spatial setting (3.11b). It developed then into a preposition (3.11c) indicating both temporal and spatial setting and finally into a subordinator indicating temporal relation between clauses (3.11a). Throughout these stages there has been no phonetic reduction although decategorization and semantic extension (limited as it was) certainly took place. This and numerous similar cases provide us with readily available material for grammaticalization studies.

However, the apparently simple hypothesis of the reasons behind the existence of identical forms in a given language is not without its problems. To prove that a lexical or grammatical item became reanalysed from some other item, we ideally need attested historical evidence showing the changes in the meaning and, where applicable, also in the category class of that item. This is a challenge that we can rarely overcome (see section 3.2.6.). Hence, on numerous occasions, uncertainty may arise as to whether we are dealing with a case of polysemy or homonymy (a phenomenon of phonological identicalness that is not a result of grammaticalization but, for instance, of borrowing or phonological change).⁶ There are two simple criteria that may help us to distinguish between these two. The first one is purely semantic:

⁶ The definitions of homonymy and polysemy given are somewhat simplified but it is not possible to go into details here and do justice to the vast amount of literature on lexical ambiguity, vagueness and polysemy and homonymy themselves. Moreover, it needs to be noted that although the notions of polysemy and homonymy have been most commonly used in the context of lexical items and lexical semantics, in works on grammaticalization they are often extended over grammatical items too (be they bound or free forms). The same approach is followed here.

In general, from the perspective of grammaticalization it is methodologically essential to assume polysemy if there is a plausible semantic relationship, whether or not the forms belong to the same syntactic category. (Hopper and Traugott 2003:77-78)

The second one uses cross-linguistic evidence: “if many diverse languages independently have the same pattern of ‘homonymy’, then the meanings are closely related” (Croft 2003:106). In other words, if items of the same phonological form share their meanings/functions in variety of languages the meanings/functions are assumed to be polysemous.

Both these criteria are, needless to say, far from being ideal analytical tools. The reason for that is that we do not have either a readily available measurement of semantic distance between concepts to apply the first criterion or a parameter of frequency which would indicate in how many languages the same pattern (e.g. identicalness of a demonstrative and a clause linker) has to occur to apply the second one. Hence, following Kortmann (1997), a more general term that would cover both homonymy and polysemy is used here: **polyfunctionality**⁷ and a distinction between two types of polyfunctionality: **syntactic polyfunctionality** and **semantic polyfunctionality** is introduced.

By syntactic polyfunctionality I understand the existence of forms which are identical in their phonological shape but belong to different syntactic categories such as the abovementioned English *after* (3.11.). The phrase “semantic polyfunctionality” is, in turn, used to refer to the phenomenon where one form is used to express more than one circumstantial relation between clauses such as the English temporal and causal *since* (3.9.) Whenever two or more items share one form the phenomenon is called an **overlap** and a distinction between syntactic and semantic overlaps is applied following the rule for the distinction between semantic and syntactic polyfunctionality. Whenever only two categories/meanings share one phonological form I talk about **binary overlaps**. Where polyfunctionalities involve more than two items – I talk about **multiple overlaps**.

3.2.3. Grammaticalization pathways and the unidirectionality hypothesis

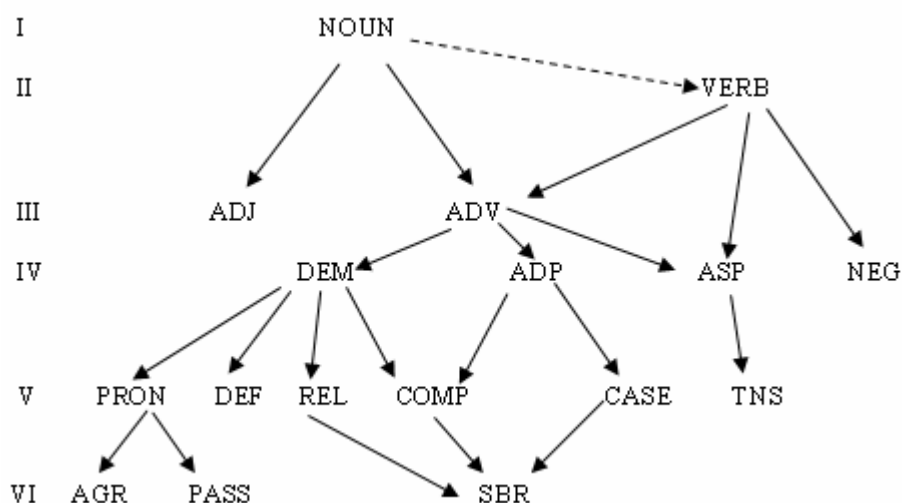
The development of items in the processes of grammaticalization is commonly depicted in the form of so called **pathways** (paths, clines) **of grammaticalization** – schemas that can be seen from the distance of time and from the linguist’s analytic

⁷ Polyfunctionality should not be mistaken with *layering* (see Kuryłowicz’s fourth law of analogy 1949 and Hopper 1991). The latter refers to the synchronic presence of both the newly grammaticalized and the older method of conveying a particular function.

perspective (Andersen 2001). The particular (hypothesized or attested) stages of development are arranged in a chronological order, for instance:

French *là* ‘there’, adverb > *-là* ‘that, distal demonstrative’
 Rama *bang* ‘go’, verb > *-bang*, clause linker of goal, purpose

Since, as has already been said, grammaticalization is first and foremost a semantic and pragmatic process and since many aspects of perception and cognition (such as the parallels between spatial and temporal organization of world and events) are universal irrespective of what language a person speaks, it becomes clear why we detect numerous cross-linguistic patterns of the grammaticalization pathways.⁸ These patterns have been reconstructed recently by Heine and Kuteva (2007). The authors, having gathered data from over 500 languages, come up with a diagram of the most salient grammaticalization pathways in the world’s languages (Fig.3.1.).



(Fig.3.1.) Layers of grammatical development according to Heine and Kuteva (2007:111)

Abbreviations: I, II, etc. = layers, ADP = adposition, ADV = adverb, AGR = agreement, ASP = (verbal) aspect, CASE = case marker, COMP = complementizer, DEF = marker of definiteness (definite article), DEM=demonstrative, NEG = negation marker, PASS = passive, PRN = pronoun, REL = relative clause marker, SBD = subordinating marker of adverbial clause, TNS = tense marker.

The diagram describes evolution in terms of a set of layers indicating the co-existing categories and the elements they develop into: the lower on the hierarchy the more grammaticalized the categories are with layer VI containing agreement markers, passive morphemes and subordinating markers of adverbial clauses.

⁸ I refrain here from bringing up the topic of the possible small differences and variations in cognition between speakers of various languages that are dealt with in literature on linguistic and cultural relativism.

The idea of pathways of grammaticalization is closely linked to the concept of **unidirectionality**, according to which grammaticalization operates in one direction. It concerns both the development of categories:

lexical form > grammatical form > more grammatical form

and the morphosyntactic development:

word > clitic > affix

The unidirectionality of the development of categories, which is the one I am focused on in this study, is also depicted in (Fig.3.1.) – none of the arrows is bipolar and so it is understood that, for instance, it is not usual for complementizers to develop into demonstratives or adpositions or for aspect markers to give rise to verbs.

In recent years numerous researchers have been presenting counterexamples to the unidirectionality hypothesis (see for instance Nevis 1984, 1985; Joseph and Janda 1988; Ramat 1992; Frajzyngier 1996; Newmeyer 1998, Campbell 2001; Fischer, Norde and Perridon 2004). Although these examples come from various languages, on the whole their number is very small in comparison to the number of examples following the unidirectionality pathways. Moreover, the adherents of unidirectionality state firmly that diachronic universals, like synchronic ones, are observed tendencies rather than theoretical absolutes and so also unidirectionality should be viewed as a hypothesis rather than an absolute principle (cf. Hopper and Traugott 2003:17, Harris and Campbell 1995:330, Heine and Kuteva 2007, Haspelmath 2004c). I agree with them.

3.2.4. Multiple pathways and polygrammaticalization

The paths of grammaticalization are not always straight in the sense of items developing in a single line one from another. Some of them show development along two or even more different clines. A more complex scenario such as this has been well described by Craig (1991) for the verb *bang* ('go') in Rama which, to put things briefly, developed independently into a variety of temporal, aspectual and modal markers on the one hand and into an adposition on the other. It was from the latter that it subsequently developed into a variety of other items including a clause linker (see also section 6.2.1.). The author called this phenomenon of development involving separate clines **polygrammaticalization**.⁹

It is not unreasonable to expect that examples of polygrammaticalization are far more frequent than it seems to us. When it comes to details, our reconstruction of grammaticalization is, without any doubt, more often than not influenced by the types

⁹ Similar scenario of multiple paths have been showed, for instance, by Givón (1991) for Biblical Hebrew relative clause morphology which developed into both adverbial clause and complementizer domains.

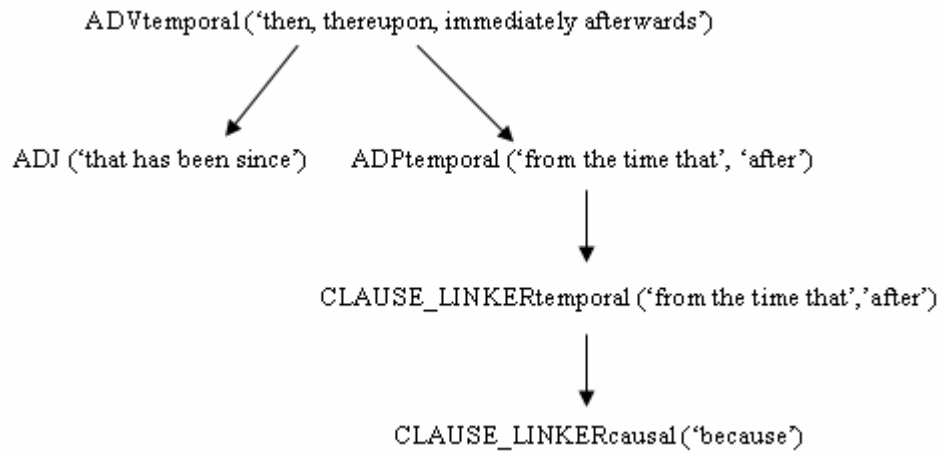
of conclusions we may draw from the poor historical records we have at our disposal, and the synchronic data we may use (see 3.2.6.). The unavoidable fate of historical reconstruction in linguistics is to be an idealized and regularized version of the past. Taking these facts into account it is advisable to keep an open mind for alternative scenarios of the ways of grammaticalization. This does not concern examples of polygrammaticalization only but also cases where competing scenarios of development of a particular item emerge from linguistic enquiry.

It is essential to say here a few additional words on explanations of affinities between particular items in grammaticalization clines involving numerous items. If we go back to the Rama example described above, it is clear that there is no direct link between the clause linker which developed from *bang* and the various TAM markers that the verb gave rise to. As a result any similarities in form between the clause linker and TAM markers are of **epiphenomenal character** – they are a side-effect of the (hypothesized) development of one item in two different directions. The existence of epiphenomenal similarities is especially important if, as in this study, our inquiry into grammaticalization is based on observations of polysemy. Let us look now at the already mentioned example of English *since*.

We know that in Old English its first attested function was that of an adverb with the meaning ‘then, thereupon, immediately afterwards’. It then developed separately into an adjective ‘that has been since’ and preposition ‘from the time that’, ‘after’¹⁰. From the latter it developed its function as a temporal clause linker ‘from the time that’, ‘after’. Already in OE in some contexts causal relations between states of affairs were inferred from the temporally linked clauses and in Middle English the causative implicatures were conventionalized and so the polysemy CLAUSE_LINKERtemporal-CLAUSE_LINKERcausal emerged in addition to the syntactic polyfunctionality of *since* as a temporal adverb, adjective adposition and temporal subordinator (for details on the evolution of *since* see Traugott 1989:34, Hopper and Traugott 2003:82-83, Molencki 2007 and OED). The complex grammaticalization of *since* that we reconstruct on the basis of historical evidence is depicted schematically in (Fig.3.2.).

Although over time the original adverb *siþþan* and the items it gave rise to changed their phonological form becoming finally the modern *since*, many of the semantic and syntactic polyfunctionalities have been retained and the item even nowadays acts as an adverb, preposition, and temporal and causal c-glosseme.

¹⁰ The adjective use was rare, as OED remarks. Its early example is to be found for instance in Josuah Sylvester’s *Du Bartas* from 1598: “That first travel had no sympathy. With our since-travel’s wretched cruelty”.



(Fig.3.2.) Grammaticalization pathways of English *since*

While discussing grammaticalization in such complex cases we encounter numerous examples of epiphenomenal affinities. The most obvious one in the case of *since* is the identicalness of forms of the preposition and clause linkers on the one hand with the adjective on the other which can be explained by polygrammaticalization that the source adverb has undergone. But as epiphenomenal can be viewed also the identity of the causal c-glosseme and adposition (or adverb) since these two, although aligned in a straight line, are not related directly – the polysemy arose through the temporal linker ‘from the time that’. As it is argued in many places in the following chapters, the fact that a given c-glosseme has polysemes in several syntactic categories is an important element of the grammaticalization puzzle and cannot be underestimated when it comes to the reconstruction of grammaticalization pathways. For this reason the notion of epiphenomenon is frequently referred to in this thesis.

The emergence of epiphenomenal identicalness of items in a grammaticalization chain involving c-glossemes is depicted schematically in (Fig.3.3). A source category undergoing desemantization, extension and decategorialization becomes reanalysed as a c-glosseme with meaning A. Then, the c-glosseme, due to context-induced changes, develops into marker B encoding other circumstantial relations and so becomes a polyfunctional linker. All this happens without changes in phonological form of the items in the chain. Since there is no direct link between the source item and item B, their identicalness is of epiphenomenal character.

difference between the idiomatic expressions and idiomatized effects of grammaticalization lies in the fact that in the latter the unit may consist entirely of bound morphemes.

The differences between the two types of grammaticalization – grammaticalization of monomorphemic units and of polymorphemic units posits certain problems for a synchrony based study – such as the one this thesis reports on. The details are explained in the following section.

3.2.6. Types of evidence

As Heine and Kuteva (2007:212-213) have put it, there are two kinds of evidence that a linguist can use in grammaticalization studies. The first one, called *diachronic*, relies on historical evidence from earlier periods of development of a given language which confirm that there was a stage at which element X existed but Y did not and hence the hypothesized pathway of development from X to Y is verified positively. The second kind of evidence, called *structural*, identifies similarities between items co-existing at a certain point in time and follows the same assumptions as the former method. For the sake of simplicity the second one is called here *synchronic*.

Both the diachronic and synchronic evidence are commonly used in grammaticalization studies. There is no doubt that the former one, based on written records, is the more reliable of the two. The truth is, however, that only a small fraction of the 6000 or so languages spoken in the world today have at their disposal such historical evidence. Moreover, even when rich evidence is available, in many cases it cannot be trusted completely. As Traugott remarks:

All claims about the order of development that are based (...) on written records and evidence from grammars and dictionaries, must be regarded with caution. As is well known, attestation is often a matter of accident. Furthermore, it does not necessarily reflect changes in the spoken language. What is significant is cumulative evidence from different but related semantic domains and, wherever possible, from other languages, of the same order of attestation among exemplars, whatever the time lag. (1989:34)

Harris and Campbell also comment on this and other problems in diachronic studies:

The problems of obtaining evidence for diachronic studies in syntax have frequently been addressed (...) Often cited are the lack of a native speaker's intuitions, accidental gaps in the corpus, the need for philological skills and thorough knowledge of the languages under investigation, and the small number and variety of languages attested over a long period of time. (1995:10)

Due to obvious limitations dictated by the scope and duration of my research, in the quest for explanation of the origin of clause linking devices, I choose to restrict myself to using synchronic evidence only. The main tool that I use is the observation of cross-linguistic patterns of polysemy (see section 3.2.2) and the results are presented in

sections 4.2.1., 5.2.1., 6.2.1. and 7.2.1. In the case of units categorized as synchronically polymorphemic, additional analyses of their internal make-up are performed (again, from a synchronic perspective) and the results are reported in sections 4.2.2., 5.2.2., 6.2.2. and 7.2.2. The analyses follow principles and assumptions which are discussed in section 3.2.9. An important point to make here is that, where possible, the hypothesized directions of grammaticalization are informed and supported by findings from other studies – this follows from Traugott’s valuable observation that:

The evidence is substantial that the process of semantic change outlined for the semantics of grammaticalization belongs to a larger set of crosslinguistic processes of semantic change that are in general quite regular. Indeed, they are so regular that it is possible to develop predictive hypotheses that can be tested against historical data. They are sufficiently predictive that one can take synchronic polysemies from any period in any language and project change back into the past. (1989:31)

The method I apply here can be viewed then as a simplified structural method that does not involve reconstructions of items that have undergone phonological erosion in the course of their historical development. An identical approach was applied by Kortmann (1997) in his study on adverbial subordinators in the languages of Europe (see section 3.2.8. for details).

Being aware of the limitations that such simplified approach puts on linguistic enquiry, I nonetheless strongly defend its value. Taking into account the aforementioned paucity of diachronic cross-linguistic data and the fact that the vast majority of languages are very poorly described, we have to admit that when it comes to broad cross-linguistic studies the synchronic evidence is often all that we have. As for the value and originality of the findings obtained by using this approach, I will let the results speak for themselves.

3.2.7. Clause linkers and language contact

The discussion on the origin of c-glossems would not be complete if we didn’t mention the significant impact that language contact may have on their introduction into a language system. There are three mechanisms by which language contact may influence the domain of clause linkers:

- a) borrowing;
- b) externally-motivated grammaticalization;
- c) polysemy copying.

Borrowing, understood as an introduction of a foreign unit into a language system (be it a free word, a bound morpheme or a construction), has always been a very interesting topic for linguists. Recent studies show clearly that not only lexical but also grammatical borrowing (borrowing of structural patterns, inflectional paradigms etc.) even from typologically divergent languages is not uncommon (see for instance

Aikhenvald and Dixon 2001a,b, Matras and Sakel 2007, Stolz, Bakker and Palomo 2008). As for clause linkers, a recent cross-linguistic study by Matras (2007) has shown that the markers are by far the most susceptible to borrowing among the grammatical categories looked at in the study. Moreover, it has been revealed that the linkers are the second most popular group of loanwords after nouns.

The second phrase – “externally-motivated grammaticalization” – refers to situations of language contact where one language copies the grammaticalization processes occurring in the other language using its own material (the process is sometimes called *grammatical replication*). As a result, grammaticalization processes may display geographic patterns, in which cases we talk about grammaticalization areas (see Kuteva 1998; Stolz & Stolz 2001:1549). This interaction of grammaticalization and language contact has received attention only very recently (Heine 1994; Bisang 1996; Kuteva 2000; Heine and Kuteva 2001) and it needs to be emphasized that the surface results of externally and internally-motivated grammaticalization are likely to be the same since the replication is based on the same principles of change – reanalysis triggered by context and salient semantic features of the source items. The examples of externally-motivated grammaticalization of clause linkers are for instance: the development of temporal linker *gaan* ‘when’ from noun *gaan* ‘time’ in Tigak of New Ireland (Papua New Guinea) as a replica of grammaticalization pathway in Tok Pisin, or the development of the relational noun – *pal* ‘possession’ into a purpose clause linker *pal* in Pipil (El Salvador) on the model of Spanish *para* ‘for, in order to’ (Heine and Kuteva 2005:91,246)

The third mechanism – polysemy copying – is a process in which one of the languages in contact copies entire patterns of polysemy from another language using its own lexical material (this is also called *calquing* or *loan translation*). Again, the effects of this mechanisms from a synchronic perspective are exactly the same as of grammaticalization and it is not a surprise that only recently linguists have considered polysemy copying as an alternative explanation for existence of patterns of polysemy (for some examples see Heine and Kuteva 2005:100-103).

In this thesis I do not attempt to go into the difficult territory of determining whether a particular pattern exists due to internally or externally-motivated grammaticalization, polysemy copying or accidental similarities between languages. This is a task that would go far beyond the main focus of this study. The issue of borrowing, on the other hand is taken into account when classifying markers as mono or polymorphemic and in the discussion of the motivations behind borrowing in section 9.3.4.

3.2.8. Previous studies on the origin of circumstantial clause linkers

The subject of the origin of clause markers of circumstantial relations has been present in linguistic literature for quite some time now and various points of view on it have been presented. The strongest opinion has been expressed by Meillet (1914) who wrote:

Les origines des conjonctions sont d'une diversité infinie, on le sait. Il n'y a pas d'especè de mot qui ne puisse livrer des conjonctions.

When it comes to the sources of conjunctions, there is an infinite diversity. There is no part of speech that could not give rise to them. [translation by A.M]

The views expressed in more recent studies on grammaticalization are less definite. Hopper and Traugott, for instance, have stated that:

Clause linking markers have their sources in nouns, verbs, adverbs pronouns, case morphemes (including prepositions and postpositions), derivational prefixes and in phrasal combination of these. (2003:177)

Also Heine and Kuteva (2007) are far from making any universal judgements and limit their summary to the cross-linguistically “most common sources”, listing, as we may read from the diagram presented in (Fig.3.1.) nouns, verbs, adverbs, adpositions, demonstratives, relativizers, complementizers and case markers. Notably, as they have explained, in the case of subordinators we can talk about direct and indirect pathways of grammaticalization:

Subordinators can arise directly from the Noun > subordinator or Verb > subordinator pathway, but they can also be the result of chain of pathways, for example Noun > adverb > adposition > case marker > subordinator. (2007:114)

Apart from the more general observations there are quite a lot of minor studies which have mentioned the role of certain specific categories in the formation of clause linkers. One of such commonly quoted works is that by Genetti (1991) who discussed the development of adverbial subordinators out of postpositions in Tibeto-Burman languages. The author indicates the following patterns of extensions (1991:3):

Locative > if/although, when/while/after

Ablative > when/while/after, because, non-final

Allative, Dative > purpose

Ergative/Instrumental > because, when, while

Numerous other studies have also independently confirmed that adpositions often give rise to clause linkers (cf. for instance Genetti 1986, Thurgood 1986, Craig 1991, Lichtenberk 1991, Heine and Kuteva 2002). It does not come as a surprise that case markers too have often been reported to serve as grammaticalization material for clause

linkers (see for instance Harris and Campbell 1995, Hopper and Traugott 2003, Heine and Kuteva 2002). An especially interesting work contributing to the discussion of grammaticalization of case markers into clause linkers is the paper by Aikhenvald (2008). On the basis of observations of the most common functions of case markers in almost 30 languages from various language families the author identified a number of patterns of semantic polyfunctionality. This includes the overlaps CASE_MARKER-CLAUSE_LINKER As the author summarizes:

The most common semantic correspondences between the same morpheme as marker of the function of a noun phrase and as clause-linking device are:

- I. Dative or purposive marking on a noun phrase tends to have a purposive meaning when used as a clause linker.
- II. Locational marking on a noun phrase tends to have temporal or more rarely, conditional or purposive meaning when used as a clause linker.
- III. Instrumental marking on a noun phrase tends to have a causal or temporal, or (more rarely) a manner meaning when used as a clause linker. (2008:594)¹²

There seem to be a general agreement between linguists regarding the types of syntactic changes that trigger grammaticalization of clause markers out of adpositions and case markers – as Harris and Campbell have remark “It is believed that in many cases the development [of clause linkers] from adpositions and case markers involved nominalized verb forms and only then developed to non-nominalized ones” (1995:293).

Other frequently mentioned sources of circumstantial clause linkers are complementizers. Saxena (1995) has even argued for a cross-linguistic implicational and unidirectional hierarchy:

word meaning ‘say’ or ‘thus’ → direct quote marker/complementizer →
reason/purpose marker → conditional marker → comparative marker

Many more authors (Lord 1976, Heine, Claudi and Hünemeyer 1991a:158, Ebert 1991:87, Frajzyngier 1996, Klamer 2000, Crass 2002, Heine and Kuteva 2007) have also indicated that the grammaticalization channel involves the following main stages:

speech act verb ‘say’ → ‘say’ as a quotative marker →
complementizer of object clauses → (complementizer of subject clauses →)
subordinator of purpose clauses → subordinator of cause clauses

We may add here also the observation that “prepositions and postpositions, being heads of noun phrases, commonly develop into markers of complement clauses, this is

¹² We can clearly see that despite differences in language samples, there are correspondences between the grammaticalization potential of the functions of adpositions reported by Genetti (1991) and the functions of case markers discussed in Aikhenvald’s study.

complementizers” (Heine and Kuteva 2007:92).¹³ Horie (2001:981) supplements the set of sources of complementizers by case markers and conjunctions.

The last of the categories most often mentioned in the context of grammaticalization of clause linkers are interrogatives (Traugott 1985, Hopper and Traugott 2003:186, Harris and Campbell 1995:293-308).¹⁴ Their role in the development of relative clause markers and subordinators seems to be, however, especially important in European languages and not widespread outside Indo-European languages (Heine and Kuteva 2007:112-113).

Although in the more general statements, such as those by Hopper and Traugott or Heine and Kuteva quoted at the beginning of this section, nouns and verbs are included among the categories commonly serving as sources of clause linkers, I am not aware of any works discussing such grammaticalization pathways in detail.

As can be seen from the above, there are a number of works that contribute findings concerning general issues in the development of clause linkers, but there are far fewer works that would discuss the grammaticalization of markers of particular types of circumstantial clauses from cross-linguistic perspective. One of the works in which we would wish to find such discussion is the already mentioned study by Kortmann (1997) which is the most detailed cross-linguistic study on adverbial subordinators conducted so far. Unfortunately, when it comes to our understanding of the origin of clause linkers it offers only a very general insight. The author presents merely overall summaries of the data collected without informing us about the grammaticalization pathways for specific types of adverbial subordinators.¹⁵ He devotes only four pages in total to the discussion of the sources of adverbial subordinators. All that we may elicit from it is that his analysis of syntactic polyfunctionality suggest that:

- a) the most common sources of adverbial subordinators in the languages of Europe are in the order of decreasing significance: adverbs, adpositions, interrogatives, complementizers and relativizers;

¹³ This is, as the authors indicate, how the Old English location and purpose preposition *for* became a complementizer by early Middle English.

¹⁴ Some other minor studies dealing with more specific cases, which are relevant for the analyses presented in this thesis, are considered in chapters 4-7.

¹⁵ Similarly to this study, Kortmann’s investigation is limited to observation of synchronic patterns of polyfunctionality and analysis of material incorporated in the structure of synchronically polymorphemic markers. For both these aspects of investigation the author presents separate tables with results. He also includes tables with adjusted values removing cases of syntactic polyfunctionality of particular items and taking into account “only the syntactic functions relevant for the formation of the adverbial subordinator”. I find this step quite controversial since the author does not discuss how exactly he decides which of the polysemous meanings is relevant for grammaticalization and which is not. As everyone who has worked on grammaticalization knows, there are many doubts and competing scenarios available in such cases.

- b) among the material incorporated in the polymorphemic subordinators it is complementizers, adverbs, adpositions relativizers and interrogatives that are most popular.

Kortmann's study is also limited in other aspects, the major limitation being the fact that it is focused only on one specific group of clause linkers – “ideal adverbial subordinators”. In practice it means that the author looked only at a subgroup of clause linkers which fulfil a number of pre-established criteria such as: being of the form of a free word (or string of words), operating over finite clauses, occurring on margins of the clause, used in non-marked register etc. (for the full list see Kortmann 1997:71-77). The study is also restricted in its geographical (and, consequently, genetic) scope – it is based exclusively on the languages of Europe. For all these reasons, its contribution to the discussion on grammaticalization of clause linkers can be viewed as very limited and it does not contribute at all to our understanding of the origin of markers of particular types of circumstantial relations.

One of the works that does look at one type of clause linkers in detail is the paper by Traugott (1985) devoted to the grammaticalization of conditionals. The author has listed five types of cross-linguistically common lexical sources for markers of the protasis: words for modality (especially epistemic and optative); copular constructions; interrogatives; words that mark something as known or given (including topic markers and demonstratives); words temporal in origin – “most especially words that in some way express the temporal notion “for a time” (that is, duration), or at least are, like *when*, ambiguous between durative meanings and non-durative (or “punctual”) ones” (Traugott 1985:292). Importantly, the author has emphasized also that it is the last group that is apparently the most common source of conditional clause linkers. I come back to the Traugott's list in section 7.2.

Schmidtke-Bode in her recently published typology of purpose clauses discusses what she labels “the developmental trajectories of purpose clauses” (2009:197-198) and what may be viewed as a list of sources of purpose clause linkers. The data are based on her own research as well as previously published works and the set includes: verbs of motion, allative and benefactive markers, complementizers and quotative markers, causality and result markers, coordinate ‘VP-and’ and serial verb constructions, as well as expressions of temporal concepts (such as ‘future’, ‘later’, ‘until’) and expressions of desire (‘order’, ‘tell to’, ‘ask to’, ‘want’). I refer to this list later on in section 6.2.

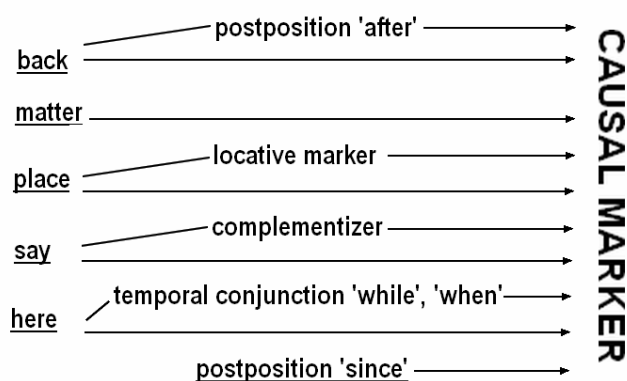
The third important and relevant for this study work in which we find more detailed information on the origin of specific classes of clause linkers is the *World Lexicon of Grammaticalization* (2002) by Heine and Kuteva (WLoG henceforth) which has become an indispensable reference book for all those who are interested in grammaticalization processes. Using material from several hundreds of languages the authors identified over 400 various grammaticalization processes, including those

involved in the development of some types of clause linkers. The data were provided by both synchronic and diachronic analyses (see 3.2.6). By extracting information scattered around the lexicon one may easily put together the development pathways that the authors identified for particular lexical and grammatical items.

Today, 8 years from its publication, WLoG remains the only source of cross-linguistic information on the development of various types of clause linkers. For this reason I find it both appropriate and beneficial to cite and acknowledge here the information it presents on grammaticalization pathways of the markers of circumstantial relations that are the subject of this thesis. I go back to these summaries in each of the analytical chapters (4-7).

CAUSALITY

Sources of causal relations mentioned in WLoG, as depicted in (Fig.3.4.) include nouns ('back' – body part, 'matter', 'place'), verb 'say', adverb of place ('here') as well as adpositions ('since', 'after').

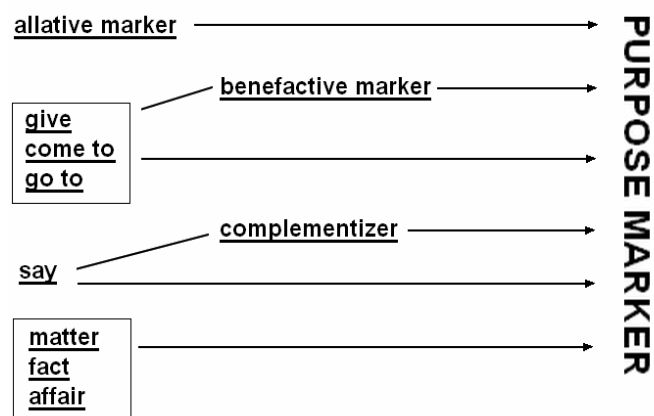


(Fig.3.4) Sources of causal clause markers according to WLoG¹⁶
(2002:48,171,200,210,239,261,275)

PURPOSE

Purpose markers, according to the information presented in WLoG, have their sources in verbs ('say', 'give', 'come to', 'go to'), nouns ('matter', 'fact', 'affair') benefactive and allative markers (be they case markers or adpositions) and complementizers (Fig.3.5.).

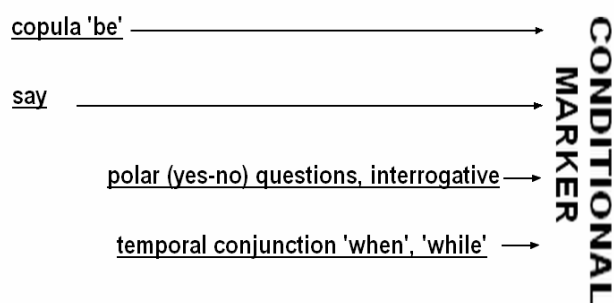
¹⁶ Some of these development pathways are direct, others involve intermediate steps. Underlining in the diagrams indicates elements that Heine and Kuteva have listed as "sources" for development of markers of a given circumstantial relations. In (Fig.3.4.), for instance, the postposition 'since' should be, therefore, interpreted as capable of being a direct source of causal markers in some cases and as acting as an intermediate point in others. Clearly, the sources of adpositions, adverbs etc. may also be traced back but I stick to those pathways that have been explicitly listed by the authors.



(Fig.3.5.) Sources of purpose clause markers according to WLoG (2002:39,56,91,79,155,163,212,265-7)

CONDITIONALITY

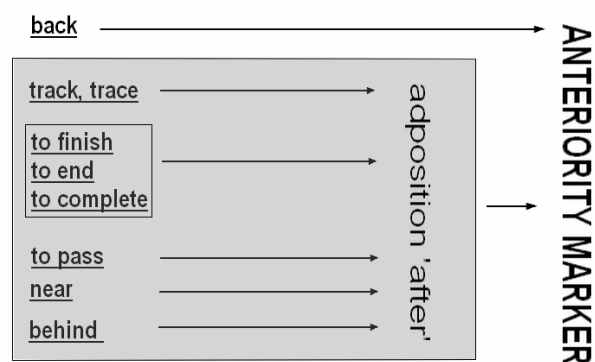
The number of reconstructed sources of conditional clause markers (Fig.3.6.) is smaller than in the two former cases and includes copula, verb ‘say’, interrogatives (polar questions) as well as temporal conjunctions like ‘while’ and ‘when’.



(Fig.3.6.) Sources of conditional clause markers according to WLoG (2002:94,249,265,293)

ANTERIORITY

Regarding the exponents of circumstantial relation of anteriority, the only explicitly given source in the lexicon is the noun ‘back’. I decided to make a small adjustment here and in (Fig.3.7.) I have included, in a grey box, information about the origin of the adposition ‘after’ elicited from WLoG which I then linked with the clausal anteriority marker. The decision to include the adposition was dictated by the well attested fact (mentioned on numerous occasions by Heine and Kuteva themselves) that temporal adpositions often develop into markers of temporal clauses.



(Fig.3.7.) Sources of anteriority clause markers according to WLoG (2002: 46,52,134,214,228,300)

The content of the lexicon, though valuable, does not answer at least three interesting questions: what is the statistical distribution of sources of particular items? (e.g. how many clause linkers of purpose in the sample originated from adpositions, case markers, verbs?); what linguistic material is most often incorporated in the polymorphemic clause linkers?; or are there any geographic/genetic patterns in the pathways of grammaticalization? Kortmann's work (1997), as already emphasized, answers the first two questions in a very general manner. My study aspires to find its place in the half way between WLoG and Kortmann's monograph by attempting to address the first and the second question with respect to particular semantic types of clause linkers. The third question, interesting as it is, would require a far larger language sample than the one used here and so, for now will have to remain unanswered.

3.2.9. *Methods and assumptions – a summary*

In the preceding sections, while discussing the general background for this part of the study, on several occasions remarks have been added regarding the methods and assumptions that are used in the analytical chapters of part II. The aim of this section is to put these remarks together and to add what needs to be added.

The general procedure of reconstruction I adopt is very similar to that described by Heine and Kuteva (2007:20):

- a) X and Y are phenomena that are related in some way
- b) Hypothesis 1: X existed prior to Y.
- c) Hypothesis 2: There was a change $X > Y$ (but X continues to exist parallel to Y).
- d) There is evidence in support of (c).
- e) There are specific factors that explain (c).

This procedure, as the authors have emphasized has been used in internal reconstruction in historical linguistics but in the case of grammaticalization, rather than

being restricted to just one language, it allows reconstructions across languages since the motivations underlying language change are cross-linguistically essentially the same. It needs to be noted, however, that it is not my aim to present a detail discussion of the specific factors explaining the changes (point e above) and so only general motivations are considered.

As already mentioned, the material used for this study is exclusively synchronic. Following Kortmann's approach (1997) I report on the syntactic and semantic overlaps observed from a synchronic perspective and draw conclusions from patterns of syntactic and semantic polyfunctionalities (polysemy). Being aware of the problem of cross-linguistic validity of syntactic terms such as for instance *adverb*, *mood* or even *noun* or *verb* (see section 2.1.1.) I have trusted the authors and consultants' judgements and in the majority of cases I follow directly the syntactic and semantic distinctions made by them. I adjust them only when enough evidence is available to combine certain categories together. Moreover, since the aim of this part of analysis is to depict the general cross-linguistic tendencies in grammaticalization, a certain level of generalization is, by definition, unavoidable. To support the hypotheses I put forward in this synchrony-based study and to prevent overgeneralizations, I use evidence from other studies on grammaticalization.

It is good practice in any type of research to approach the data with as few assumptions as possible. In a study on linguistic change, as in any study of evolutionary character, it is, however, especially difficult to avoid them due to the problem of gaps in data (cf. section 3.2.6.). The best we can do is to try to make the necessary assumptions as sound as possible. This rule motivated the choice of the 6 main assumptions I made for the purpose of this study. I call them henceforth **grammaticalization heuristics**:

1. If a given syntactic category (e.g. noun) overlaps in form with a specific type of c-glosseme (e.g. anteriority c-glosseme) and this overlap is frequent cross-linguistically, the category and the c-glosseme are closely related.
2. If it has been established that the meaning encoded by a c-glosseme and the syntactic category that the c-glosseme overlaps with are closely related, this affinity is an effect of grammaticalization.
3. Grammaticalization operates always or almost always in a unidirectional manner. The order of emergence of layers of new syntactic categories is as Heine and Kuteva describe it (cf. Fig.3.1.) – with adverbial subordinators (and other types of c-glossemes) at the very bottom.

4. The likelihood that a category is the direct source of a c-glosseme is highest for those categories that display the highest proportions of binary overlaps with the given type of c-glosseme (eg. anteriority c-glosseme).
5. The conclusions made following assumptions 1-4 are most reliable for the semantically monofunctional markers.
6. Where no other evidence is available, the general tendency of less abstract meanings developing into more abstract ones is used as a guiding principle.

Some additional remarks on the assumptions should be made here. The first concerns the issue of unidirectionality which, as has been described in section 3.2.3. has provoked much discussions among linguists. However, since there seems to be only one counterexample in the grammaticalization literature on clause linkers brought to the attention of linguists so far (the Japanese concessive marker *ga*, which, apparently developed into adversative conjunction ‘but’, see Hopper and Traugott 2003:210) I find no reasons to abandon the conviction of the unidirectionality tendency.

The second remark concerns the fourth assumption – when a syntactically polyfunctional c-glosseme displays also semantic polyfunctionality various scenarios are considered and the one which finds more support in other data is favoured. On those occasions where the patterns of polyfunctionality seem to suggest the possibility of multiple pathways of grammaticalization, the existence of polygrammaticalization and epiphenomenal character of affinities (cf. section 3.2.4.) are considered.

The grammaticalization pathway less abstract > more abstract, referred to in assumption 6, also needs to be elaborated on. Such a pathway is a commonly observed phenomenon. Heine and Kuteva, for instance, on many occasions mention the type of change when a form used for a visible object (e.g. the body part ‘back’) is used also to refer to a non-visible item (spatial notion ‘behind’) or a form used for an action (‘go to’) is used to refer to a grammatical notion (future tense). They describe this type of change in terms of metaphorical transfer that leads for example “from the domain of concrete objects to that of space, from space to time, from (“real word”) space to discourse space etc.” (2002:3). Similarly, Genetti (1991:231) explains the grammatical extension of postpositions into clause markers in Tibeto-Burman languages as “a process by which the basic case relations are extended to more abstract domains, as they proceed from coding relations between arguments to relations between propositions”. The same principle has been postulated also for the development of more specific semantic types of c-glossemes – as Hopper and Traugott have noticed “it appears that temporals can be the sources of conditionals (and causals), not vice versa; conditionals can in turn be the source of concessives (as can temporals such as *while* and focus particles such as *even* or universal quantifiers such as *any* as in *anyhow*)”

(2003:187).¹⁷ The works by Traugott (1987, 1989, 1999, 2003) on the development of deontic meanings into epistemic ones also provide support for the principle less abstract > more abstract. Using the term *subjectification* Traugott emphasizes three important tendencies:

- a) meanings based in external described situations develop into meanings based in the internal (evaluative/perceptive/cognitive) described situation;
- b) meanings based in the external or internal described situation develop into meanings based in the textual and metalinguistic situations;
- c) meanings tend to become increasingly based in the speaker's subjective belief-state/attitude towards the proposition.

On the whole, as Traugott has noticed “the meaning based in the sociophysical world precedes that based in the speaker's mental attitude” (1989:46, cf. also Langacker 1990).¹⁸

Finally, it is important to mention that in actual reconstruction (as well as in preparing the list of grammaticalization heuristics), I take into account the three general rules of inquiry into the history of language (Hock 1986:535-541)¹⁹:

- a) naturalness – given two otherwise equally acceptable competing analyses the one which postulates more common or more natural processes is preferred;
- b) explanation – given two alternative analyses the one which provides greater explanation or motivation for postulated changes and for the attested synchronic facts is preferred;
- c) Occam's Razor *entititia non sunt multiplicanda praeter necessitatem* – entities (in an argument) are not to be multiplied beyond necessity, and so the simplest possible scenario is to be preferred.

3.3. Insight into the semantic organization of the network of c-glossems

As it has been discussed in chapter 1, linguists agree that the semantic space of circumstantial relations is far from being discrete – cases of linkers used to express more than one relation (which are called here, recall, *semantically polyfunctional*) are by no means rare and we encounter them in every part of the world.

¹⁷ For similar remarks see also Claudi and Heine (1986), Hock (1986:290), Heine, Claudi and Hünnemeyer (1991a), Heine and Kuteva (2007:33).

¹⁸ See also Sweetser (1984).

¹⁹ The rules, or principles, are applicable first and foremost to sound changes but are guidelines also for semantic and morphosyntactic changes.

The analysis of the range of the circumstantial functions expressed by particular linkers offers us an invaluable insight both into processes of grammaticalization and cognitive organization of the human mind.²⁰ The observations allow us to create (and, where possible, also test) hypotheses on the semantic extensions made by speakers from one meaning into another. These extensions more often than not appear to be of metonymic character.²¹

3.3.1. Degree of semantic polyfunctionality

The degree of underspecification of meanings that semantic polyfunctionality introduces can vary. The simplest cases are those when one c-glosseme covers only two meanings (as in the case of causal and temporal readings of English *since*) but among the world's languages we encounter cases where one marker may convey a much wider variety of circumstantial meanings. Although these highly polyfunctional linkers have not been a subject of any systematic study yet, their presence has certainly been noticed by authors of grammars and other linguists. In Tepehuan, for instance, such a marker has been simply called *subordinate clause introducer* (Willett 1991), in Hualapai *general subordinating particle* (Watahomigie et al. 1982), in Quechua Huallaga *adverbializer* or *adverbial clause marker* (Weber 1989), in Estonian *non-marked subordinating conjunction* (Mati Ereht, personal communication). Heine and Kuteva labelled a similar case in !Xun *general subordinator of adverbial clauses* (2007:250) while Harris and Campbell have talked about *generic* or *all-purpose subordinators* (1995:148,313).

Between the semantically monofunctional and highly polyfunctional linkers there are c-glossemes encoding two, three, four and more functions. However, when it comes to determining the exact degree of polyfunctionality (exact number of circumstantial meanings that a particular linker encodes) there is a serious methodological problem – how to distinguish between actual meanings of the markers and the implicatures they bring about? Let us consider the following example:

(3.15.) *After* Jack told his sister the truth she didn't want to talk to him any more.

The meaning directly encoded by *after* is clearly that of temporal antecedence. However, from such a sentence we would also understand that there is a direct causal link between the two states of affairs:

(3.16.) *Because* Jack told his sister the truth she didn't want to talk to him any more.

²⁰ There are good reasons, as cognitive linguists and psycholinguists have already convinced us, to use linguistic analysis in the studies of cognition.

²¹ I define metonymy, following Kövecses and Radden (1998:38), as a cognitive process in which “one conceptual entity (...) provides access to another conceptual entity within the same domain”.

Does it mean that *after* is a polyfunctional marker capable of encoding both temporal and causal meaning? Native speakers would surely answer ‘no’ indicating that the causal reading of *after* is heavily context-dependent.²² Indeed, what we are actually here dealing with is an issue of distinction between semantics and pragmatics on the one hand and determining the degree of grammaticalization on the other.²³ Since it is, as has been said in section 3.2.1. linguistic inference – implicature – that often triggers the grammaticalization process it is often virtually impossible to distinguish between a meaning that already has been, through frequent use, grammaticalized and the one that is simply inferred (cf. also Dahl 1985:11). We have to join Kortmann in his opinion that

Even for native speaker linguists or specialists of the language it is often difficult to decide in whether the given reading or use of a lexical item falls into the realm of semantics or rather that one of pragmatics. Even greater are the problems in a typological project involving many languages which are not nearly as well documented as the major European languages, not to mention the nonexistence or inaccessibility of fine-grained semantic analyses, for which, in other words, no independent evidence is available which helps to verify data collected on the basis of informant interviews plus the odd grammar or dictionary. (1997: 93)

If we do not have reliable tools to distinguish between meanings and implicatures in the case of clause linkers, can we at least distinguish between more and less salient (or primary and secondary) meanings of the markers? The question, again, boils down to the criteria we have to our disposal. Kortmann (1997:94), who has attempted to apply the distinction between primary and secondary readings decided that the most important criterion for him would be “whether a given reading does or does not require a special context (...). Primary meanings are assumed to be those which a lexical item has in isolation whereas secondary meanings are nonautonomous (or: contextbound)”. There are, however, obvious problems with determining the contextual requirements for each particular item in a broad cross-linguistic study and, consequently for distinguishing between secondary and primary meanings as Kortmann himself admits (1997:94). For this reason in my study the distinction between these two types of meanings is ignored. Understandably, I do not attempt to approach the topic of distinction between meaning and implicature of clause linkers either. In determining the specific meanings of a linker and, consequently, its degree of polyfunctionality I rely exclusively on the list of meanings extracted from published materials and communication with consultants. Similarly as in the case of syntactic overlaps, where one marker encodes two circumstantial functions the overlap is labelled **binary overlap**. The overlaps involving three and more circumstantial meanings are called

²² Cf. *After I did the shopping, I went to see my aunt* or *After I come home I have to call Lisa* which clearly do not receive causal readings.

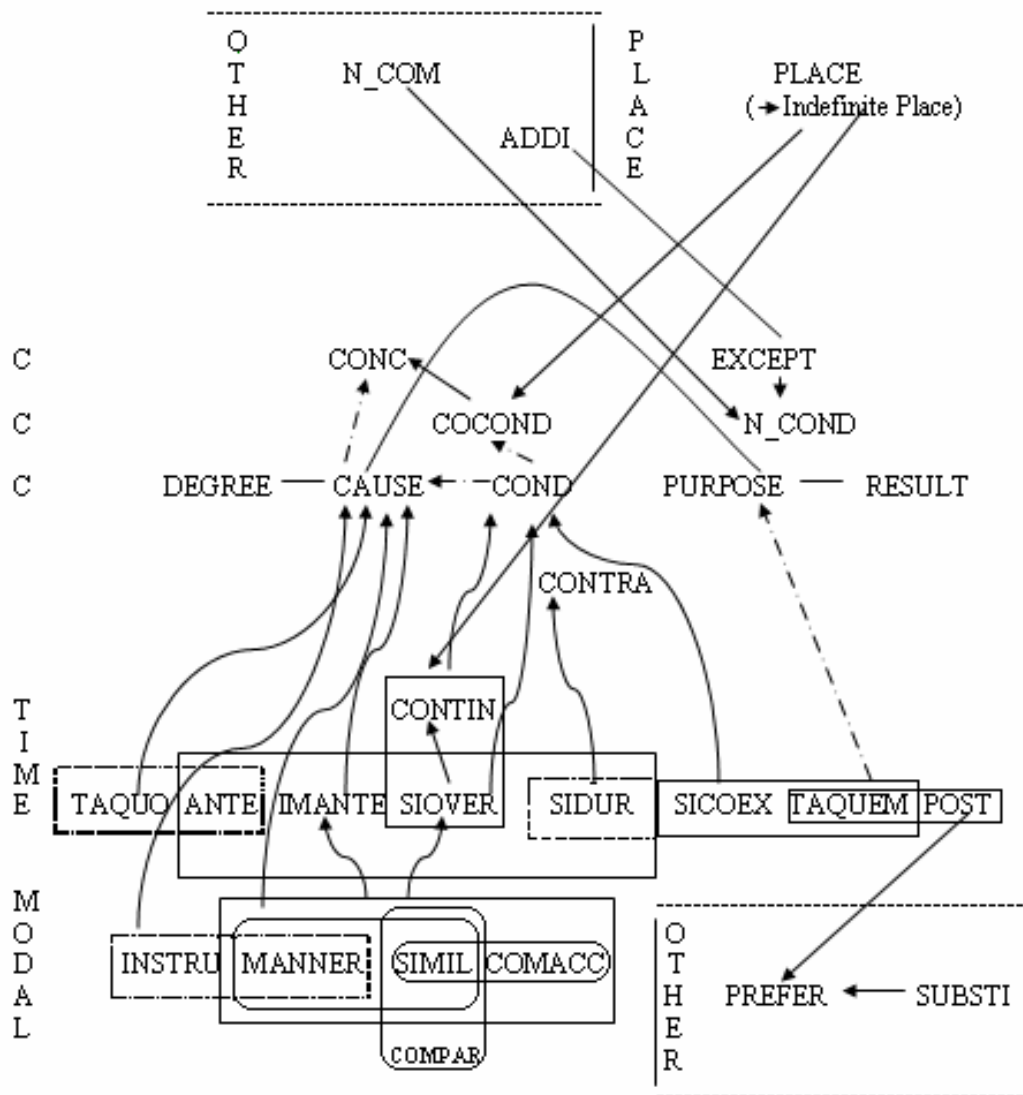
²³ Hopper and Traugott (2003:82) discuss the causal implicatures of clauses linked by *after* in terms of *pragmatic polysemy* distinguishing it from *semantic polysemy* such as the one displayed by *since*.

multiple overlaps and within this group, for those which are used to encode five and more meanings, the term **general c-glosseme** is used.

The various readings of semantically polyfunctional linkers may be, and often are, disambiguated by the presence and/or configuration of some other elements of the clause (e.g. special TAM marking on verbs or word order changes). In other cases, however, it is entirely up to the context of the utterance, and ultimately up to the reader to identify the type of relation (the meaning of the c-glosseme) meant by the speaker. The problem of disambiguation would require a separate, detailed treatment and is not entered into here.

3.3.2. *Previous studies on the polyfunctionality of clause linkers*

Although it would be a great injustice to say that the phenomenon of semantic polyfunctionality of various c-glossemes has escaped linguists' notice, it is by all means true that it has not received much typological attention. Apart from brief remarks made on the margins of discussions concerning more general topics in works such as Thompson and Longacre (1985), Heine and Kuteva (2002, 2007), Dixon (2009), or in papers focused on particular c-glossemes in a given language (cf. the volume devoted to English connectives edited by Lenker and Meurman-Solin 2007 or the two volumes edited by Traugott and Heine 1991b), which are referred to in the analytical chapters, the only study that investigates the area in more detail is the aforementioned volume on adverbial subordinators by Kortmann (1997). Applying the method of semantic map (discussed in detail in Haspelmath 2003), Kortmann revealed both a complex picture of cross-linguistic patterns of semantic polyfunctionalities of adverbial subordinators and the most common semantic affinities in the domain of circumstantial relations (Fig.3.8.).



(Fig.3.8.) A cognitive map of the most important affinities within the semantic space of interclausal relations proposed by Kortmann (1997:210)

The abbreviations stand for: N_COM – negative concomitance ‘without’; ADDI – addition; CCC – causal, conditional, concessive and related interclausal relations; CONC – concession, COCOND – concessive condition ‘even if’; EXCEPT – exception; N_COND – negative condition ‘unless’; COND – condition; CONTRA – contrast; TAQUO – terminus a quo ‘since’; ANTE – anteriority; IMANTE – immediate anteriority ‘as soon as’; CONTIN – contingency ‘whenever’; SIOVER – simultaneity overlap ‘when’; SIDUR – simultaneity duration ‘while’; SICOEX – simultaneity co-extensiveness ‘as long as’; TAQUEM – terminus ad quem ‘until’; POST – posteriority ‘before’; INSTRU – instrument/means ‘by’; SIMIL – similarity; COMACC – comment/accord ‘as’; PREFER – preference ‘rather than’; SUBSTI – substitution ‘instead of’; COMPAR – comparison ‘as if’.

3.3.3. Goals, methods and assumptions

When it comes to the analysis of the phenomenon of semantic polyfunctionality of c-glossems, the ambition of this study is to contribute to our understanding of grammaticalization processes within the domain of clause linkers and to pursue Kortmann's investigation of the semantic affinities between circumstantial relations on the other. Since these two tasks have much in common they both are discussed in detail within one section separately for each of the circumstantial relations (i.e. in sections 4.3., 5.3., 6.3. and 7.3).

In the aspect of grammaticalization the analysis follows the principles described in section 3.2.9 with a special emphasis on the less abstract > more abstract tendency of development. For the more salient grammaticalization pathways, I attempt to propose brief explanations for the route of semantic extensions by which the polyfunctionalities arose.

The insight into the cognitive affinities in the semantic space of circumstantial relations, in turn, follows two assumptions which henceforth are called **affinity heuristics**:

1. If a given circumstantial relation (e.g. anteriority) overlaps in form with another circumstantial relation (e.g. causality), and this overlap is frequent cross-linguistically, the two relations are closely related cognitively.
2. The more frequently two circumstantial relations overlap in form and the more frequently the overlap is binary (i.e. involves only those two relations) the more cognitively close the relations are.

The reconstructed patterns of semantic overlaps are presented in a form of semantic maps of affinities. The labels for particular circumstantial relations follow, in principle, those used by Kortmann as listed in (Fig.3.8), the only difference being the use of the label 'manner' which in this thesis denotes meanings of both manner and means (the reasons for that are explained in section 6.3.).

3.4. Summary

In this chapter background information and the parameters for analysis of c-glossems have been discussed. In section 3.1. a description of classification of c-glossems according to their form and internal complexity has been presented along with some examples. Section 3.2. dealt with relevant issues related to grammaticalization, its mechanisms and tendencies as well as the methods and assumptions used in grammaticalization theory. It was accompanied by an overview of

previous studies on the grammaticalization of clause linkers with a special emphasis on cross-linguistic studies. Other mechanisms of introduction of clause-linking devices into a language system, such as borrowing and polysemy copying, have also been discussed. Finally, the goals, methods and assumptions for the analysis in the relevant sections of chapters 4-7 have been listed.

In section 3.3. we have looked at the problem of semantic polyfunctionality of c-glossems discussing its usefulness for the grammaticalization studies and studies in the organization of the network of affinities between various concepts in the domain of circumstantial relations. Methodological problems concerning distinction between the number of meanings of a particular polyfunctional item have been emphasized and the only broad study in the subject has been briefly described. The section concluded with a description of the goals, methods and assumptions to be used in the following chapters when it comes to the analysis of semantic space of c-glossems of anteriority, causality, purpose and conditionality.

CHAPTER 4

Anteriority

4.1. MORPHOLOGICAL COMPLEXITY AND FORMS OF C-GLOSSEMES

Among the 84 languages in the sample, 20 do not have at their disposal any c-glossemes that would act as an exponent of anteriority (ANTE henceforth) on subordinate clauses (see section 9.2.2. for details). The 64 languages for which the presence of anteriority linkers has been recorded contribute in total 150 c-glossemes. For 5 of the linkers, it was not possible to verify information on their morphological complexity and among the remaining 145 linkers almost 75% are monomorphemic (Fig.4.1.). The polymorphemic ones are in majority bimorphemic. Those made up of 3 and more morphemes contribute only 7 items (less than 5%).

	count	%
monomorphemic	105	72.4%
2 morphemes	33	22.8%
3 morphemes	4	2.7%
3+ morphemes	3	2.1%
TOTAL	145	100%

(Fig.4.1.) Morphological complexity of anteriority c-glossemes

The data I collected have revealed that the proportion of free-word and affixal c-glossemes is almost equal in the encoding of anteriority (Fig.4.2.).

	count			%
free word	70			46.6%
affix	64			42.7%
	suffix	prefix	circumfix	
	59	4	1	
combination	6			4.0%
distributed	7			4.7%
clitic	3			2.0%
TOTAL	150			100%

(Fig.4.2.) Forms of anteriority c-glosseemes

Among the affixal c-glosseemes we find some prefixes as in Mantaoran Rukai (4.1.), one example of a circumfix in Burushaski (3.3.) and numerous suffixes (including a postbase in Central Alaskan Yup'ik – see 4.2.).

(4.1.) Mantaoran Rukai (Zeitoun 2007:439)

Ona'i alrepenge-nga-li kone,
 that {ANTE}.finish-already-1SG.GEN DYN.SBJ.eat
maava'i 'i Dhipolo
 DYN.FIN.com then Dhipolo

‘When I had finished eating, Dhipolo came.’

(4.2.) Central Alaskan Yup'ik (Reed, Myaoka, Jacobson et al. 1997:244)¹

Tuntuq tuquterraarluku amuullruarput.

‘After killing the caribou, we skinned it’.

Morphologically complex c-glosseemes that consist of both free words and affixes, which I call *combinations* (see section 3.1. and example 3.4.), contribute nearly 5% of all anteriority linkers. The next class, also contributing 5%, are distributed c-glosseemes, i.e. polymorphic linkers whose morphemes are not adjacent but are separated by some other elements of the supporting clause. Such anteriority markers have been identified in Kanuri (example 4.3. below), Konso, Akan and Burushaski and they are looked at closely in section 4.2.2.

(4.3.) Kanuri (Hutchison 1976:113)

Ngawo korkure darye-bela gudowum cize
 {ANTE} crow last-{ANTE} cock get.up

‘After the final crow (of the morning) the cock got up’

Finally, there are also isolated cases of clitics coming from Konso (-yyé) Lillooet (ǻ-) and Ama (-mo).

¹ The example is quoted as presented in the reference grammar – without exact glosses.

4.2. INSIGHT INTO ORIGINS

4.2.1. Syntactic polyfunctionality and patterns of polysemy

Out of the 150 anteriority c-glossems, information on their syntactic mono/polyfunctionality has been confidently encoded for 141 items. The proportion of syntactically monofunctional and polyfunctional anteriority linkers is similar – the first group outnumbers the second one only by 5 items as the summary in (Fig.4.3.) shows. Among the syntactically monofunctional markers, the number of monomorphemic and polymorphemic c-glossems is also very similar, whereas among syntactically polyfunctional linkers the monomorphemic ones are over ten times more frequent than the polymorphemic ones.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	39	27.66%	62	43.97%	101
polymorphemic	34	24.11%	6	4.26%	40
TOTAL	73	51.77%	68	48.23%	141

(Fig.4.3.) Distribution of c-glossems of anteriority according to their syntactic mono/polyfunctionality and morphological complexity

More detailed analysis of the polymorphemic anteriority c-glossems shows also (Fig.4.4.) that in this group 82.5% of the markers are bimorphemic. Among the 40 polymorphemic linkers, they contribute 67.5% of syntactically monofunctional markers and they are the only group contributing syntactically polyfunctional c-glossems. The remaining 7 more complex c-glossems are exclusively monofunctional.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
2 morphemes	27	67.50%	6	15.00%	33
3 morphemes	4	10.00%	-	-	4
3+ morphemes	3	7.50%	-	-	3
TOTAL	34	85.00%	6	15.00%	40

(Fig.4.4.) Distribution of syntactic mono- and polyfunctionality in polymorphic anteriority c-glossems

The findings depicted in (Fig.4.3.) and (Fig.4.4.) imply the following correlation: the more morphologically complex a marker is, the less likely it is to serve numerous syntactic functions (this tendency is called henceforth *complexity* → *monofunctionality tendency*). The resistance of the polymorphemic anteriority clause linkers to syntactic polyfunctionality becomes apparent also when we look exclusively at the *degree of syntactic polyfunctionality* (Fig.4.5) i.e. at the number of other syntactic functions that a

given c-glosseme serves. Comparing the data in (Fig.4.3.) and (Fig.4.5.) we see that in addition to being clause linkers, 4 of the 40 polymorphemic anteriority c-glossemes (10%) serve only one additional function and two (5%) are trifunctional. Among the monomorphemic markers 43.5% are bifunctional, 15% trifunctional and 4% serve four and more functions.

	1 additional function	2 additional functions	3 additional functions	TOTAL
monomorphemic	43	15	4	62
polymorphemic	4	2	-	6

(Fig.4.5.) Degree of syntactic polyfunctionality of anteriority c-glossemes

It is the 68 syntactically polyfunctional c-glossemes that I am focused on here, aiming to reconstruct the most common sources of anteriority linkers, following the methodology and assumptions explained in section 3.2.9.

The analysis has revealed that anteriority c-glossemes overlap in form with a number of other syntactic categories. All of the *syntactic overlaps* – both binary and multiple (see section 3.2.2.) are reported in (Fig.4.6.). Column I gives information on the total number of occurrences of a particular category among the syntactic overlaps of anteriority linkers. Columns II and III present data on the configuration of those overlaps.

The first thing that surely attracts attention in the table is the high number of adpositions. They constitute over 28% of all the overlaps that have been identified and almost 38% of the overlaps that occur more than two times (rows 1-10 in the table). Half as frequent than the overlaps with adpositions are overlaps with adverbs and case markers which are then followed by nouns and verbs. What brings all these categories together is also the fact that all of them are often encountered in binary overlaps.²

The three most probable explanations for the cross-linguistic rarity of the items in rows 14-19 are:

1. The categories overlaps in form with the anteriority c-glosseme accidentally (they are homonyms of ANTE linkers).
2. The categories are polysemous with c-glossemes but the grammaticalization pathway is cross-linguistically rare or even language-specific.

² Nouns are an exception here – only 2 out of 7 occur in binary overlaps. However, as discussed on the following pages the 5 multiple overlaps in which we find nouns are exclusively NOUN-ADPOSITION-LINKER overlaps. This configuration follows exactly the pathways indicated by Heine and Kuteva as depicted in (Fig.3.1.) Hence in all the NOUN-LINKER overlaps nouns (whether overlapping additionally with adpositions or not) are most likely the category that the grammaticalization processes leading to the emergence of c-glossemes of anteriority originated from.

3. The fact that the categories overlap with c-glossemes of anteriority in form is an epiphenomenon of semantic polyfunctionality of the c-glosseme (see the discussion in section 3.2.4.)

An initial application of the grammaticalization heuristics 1-4 (see section 4.2.9.) to the findings presented in (Fig.4.6.) allows us to conclude that the syntactic categories that are most common sources of anteriority c-glossemes are adpositions, case markers, adverbs, verbs and nouns. The conclusions would not be complete, however, if we didn't look at the particular meanings/functions of those source categories that triggered the processes of reanalysis and grammaticalization and led to the emergence of syntactic polyfunctionality. The specific meanings/functions together with the number of their occurrences are listed in (Fig.4.7.).

		I occurrences	II occurrences in binary overlaps	III occurrences in multiple overlaps
1.	ADPOSITION	26	11	15
2.	ADVERB	12	4	8
3.	CASE	11	10	1
4.	NOUN	7	2	5
5.	VERB	5	5	-
6.	CONJUNCTION	5	1	4
7.	RELATIVE CLAUSE MARKER	4	2	2
8.	COMPLEMENTIZER	3	-	3
9.	PARTICIPLE MARKER	3	-	3
10.	ASPECT MARKER	3	2	1
11.	NOMINALIZER	2	1	1
12.	ADVERBIALIZER	2	1	1
13.	INFINITIVE MARKER	2	1	1
14.	AORIST MARKER	1	1	-
15.	AUXILIARY	1	1	-
16.	PARTITIVE MARKER	1	1	-
17.	PREDICATIVE MARKER	1	1	-
18.	SETTING PARTICLE	1	1	-
19.	TOPIC MARKER	1	1	-
	TOTAL	91	46	45

(Fig.4.6.) Syntactic overlaps of anteriority c-glossemes

The first row of the table, for instance, should be read as follows: in the sample anteriority c-glossemes overlap with adpositions 26 times; this includes 11 cases of ANTE linker-ADPOSITION overlaps and 15 overlaps that involve also one or more additional categories.

Such closer analysis reveals that among all the categories it is the spatial, temporal and spatio-temporal adpositions with the meaning of ‘after’ that are the most common sources of anteriority c-glossemes. Adpositions with the meaning ‘behind’ are much rarer and in the sample they always occur in multiple overlaps being accompanied by an adverb or a noun. The overlap ADP-ADV is especially common for the adpositions with the meaning ‘after’ (there are 7 such cases in the table). According to Heine and Kuteva (2001) it is adpositions that develop from adverbs and not the other way round. If this really proves to be a rule the ‘after’ adpositions should be viewed as direct sources of the c-glossemes and the grammaticalization pathways would look as follows:

ADV → ADP → LINKER_{anteriority}

At the same time there are 4 adverbs occurring in binary overlaps with clause linkers which allows us to conclude that the mediation of adpositions is not always the case.

A very similar story can be told about another case of multiple overlaps involving adpositions, namely the NOUN-ADP-LINKER syntactic polyfunctionality. Nouns ‘back’ and ‘posteriority’ themselves (or accompanied by case markers, as shown in section 4.2.2) can act as sources of anteriority c-glossemes:

NOUN → LINKER_{anteriority}

or they can develop into adpositions which then may further develop into the c-glossemes.:

NOUN → ADP → LINKER_{anteriority}

However, neither adverbs nor nouns but case markers appear to be the second most common source of anteriority linkers. Ablative, with four occurrences, is the most frequent function overlapping with c-glossemes. Interestingly, as depicted in (Fig.4.7.) by the numbers in the brackets and in (Fig.4.9.) below, two of the linkers that overlap in form with ablative case endings are semantically polyfunctional conveying also the meaning of causality in Lepcha and causality, comparison and contrast in Quechua Hualлага. Applying the assumption that less abstract meanings develop into more abstract ones we can hypothesize that in these cases the pathway of development looked as follows:

CASE_{ablative} → LINKER_{anteriority} → LINKER_{causality} (→ LINKER_{other meanings})

The other case markers listed in (Fig.4.7.), i.e. locative, comitative, dative, accusative, instrumental/allative, occur less frequently and so we cannot draw any

strong conclusions about their cross-linguistic salience in grammaticalization of anteriority linkers on the basis of the material collected here.³

The final category that was already said to give rise to ANTE c-glossemes is the category of verbs. In the sample we encounter 5 binary VERB-LINKERanteriority overlaps. This includes 3 overlaps with verbs ‘to finish’/’to complete’, one with verb ‘to do’ and one with ‘be so/be true’.

³ It is worth remembering that, according to Aikhenvald’s study (2008), it is the locational marking on a noun phrase that tends to have temporal meaning when used as a clause linker (cf. section 3.2.8).

POLYSEMOUS/ HOMONYMOUS CATEGORY	CHARACTERISTICS OF POLYFUNCTIONALITY	
ADPOSITION Total:26	BINARY OVERLAPS:11 ADP'after'(spatial/temporal):9 ADP'after';locative:1 ADPlocative:1 (1)	MULTIPLE OVERLAPS:15 ADP + ADVERB:7 ADP'after' + ADV'after':3 ADP'after' + ADV'then':1 ADP'behind' + ADV'then':1 ADP'after' + ADV'in back' (spatial):1 ADP'after' + ADV:1 ADP + NOUN: 5 ADP'after' + NOUN'back':2 ADP'after' + NOUN'posteriority':1 ADP'behind' + NOUN'back':1 ADP'behind';'after'(temporal) + NOUN'back':1 (1) ADP + other: 3 ADP + ASPprogressive + CONJ'and':1 (1) ADPcomitative('with') + CONJ'and':1 ADP'against' + OBL:1 (1)
ADVERB Total:12	BINARY OVERLAPS:4 ADV'afterwards':1 ADVmanner:1 ADV'first':1 ADV'then':1 (1)	MULTIPLE OVERLAPS:8 ADV+ ADP:7 ADV'after' + ADP'after':3 ADV'then' + ADP'after':1 ADV'then' + ADP'behind': 1 ADV'in back' (spatial) + ADP'after': 1 ADV + ADP'after': 1 ADV + CONJ + REL:1 (1) ADV'after';'for a while';'first of all' + CONJ'but'+ REL

CASE Total:11	BINARY OVERLAPS:10 CASEablative:4 (2) CASElocative:2 CASEcomitative:1 CASEdative:1 (1) CASEaccusative:1 (1) CASEinstrumental;allative:1 (1)	MULTIPLE OVERLAPS:1 CASE + ADVLZ + INF:1 (1) CASEdative + ADVLZ + INF
NOUN Total:7	BINARY OVERLAPS:2 NOUN'posteriority':1 NOUN'time':1	MULTIPLE OVERLAPS: 5 NOUN + ADP:5 ADP'after' + NOUN'back':2 ADP'after' + NOUN'posteriority':1 ADP'behind' + NOUN'back':1 ADP'behind';'after'(temporal) + NOUN'back':1
VERB Total:5	BINARY OVERLAPS:5 VERB'finish';'complete':3 VERB'do':1 VERB'be so';'be true':1 (1)	MULTIPLE OVERLAPS:0
CONJUNCTION Total:5	BINARY OVERLAPS:1 CONJ'and':1 (1)	MULTIPLE OVERLAPS:4 CONJ + COMP: 2 CONJ'and'(NPs) + COMP:1 (1) CONJ'and' + ADPcomitative ('with'):1 (1) CONJ + ADP + ASP:1 CONJ'and' + ADP + ASPprogressive:1 (1) CONJ + ADV + REL:1 CONJ'but' + ADV'after';'for a while';'first of all' + REL:1 (1)

RELATIVE CLAUSE MARKER Total: 4	BINARY OVERLAPS: 2 (2)	MULTIPLE OVERLAPS: 2 <i>REL + COMP + PTCP: 1 (1)</i> <i>REL + CONJ + ADV: 1 (1)</i> <i>REL+ CONJ'but' + ADV</i>
COMPLEMENTIZER Total:3	BINARY OVERLAPS:0	MULTIPLE OVERLAPS:3 <i>COMP + REL + PTCP:1 (1)</i> <i>COMP + PTCP:1 (1)</i> <i>COMP + CONJ : 1</i> CONJ'and'(NPs) + COMP
PARTICIPLE MARKER Total:3	BINARY OVERLAPS:0	MULTIPLE OVERLAPS:3 <i>PTCP + COMP:1 (1)</i> <i>PTCP + COMP + REL:1 (1)</i> <i>PTCP + NMLZ:1 (1)</i>
ASPECT MARKER Total:3	BINARY OVERLAPS:2 ASPsubsequent:1 ASPcompletive;continuat.;inchoative1 (1)	MULTIPLE OVERLAPS:1 <i>ASP + CONJ + ADP:1 (1)</i> ASPprogressive + CONJ'and' + ADP
NOMINALIZER Total:2	BINARY OVERLAPS:1	MULTIPLE OVERLAPS:1 <i>NMLZ + PTCP:1</i>
ADVERBIALIZER Total:2	BINARY OVERLAPS:1	MULTIPLE OVERLAPS:1 <i>ADV LZ + INF + CASE:1 (1)</i> ADV LZ + INF + CASEdative:

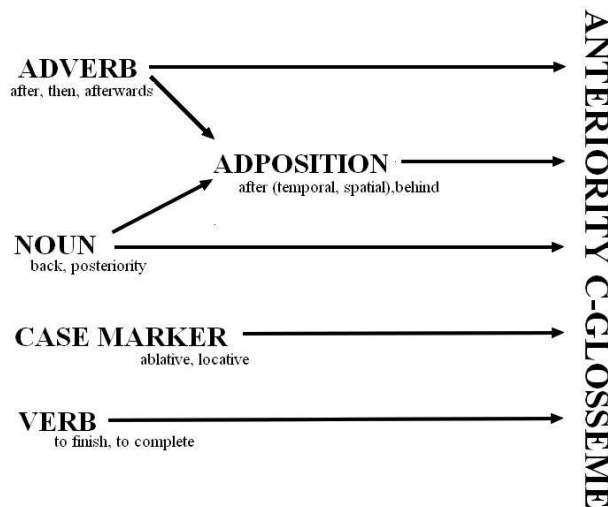
INFINITIVE MARKER Total:2	BINARY OVERLAPS:1	MULTIPLE OVERLAPS:1 <i>INF + CASE + ADVLZ:1</i> (1) INF + CASE_{dative} + ADVLZ
--	--------------------------	--

(Fig.4.7.) Details of syntactic overlaps of anteriority c-glossemes

All the categories that occurred as overlaps of anteriority c-glossemes more than once are listed. Whenever an item has two or more meanings/functions all of them are listed (hence cases such as ADP_{after};locative or CASE_{instrumental};allative). The meanings are given as they have been presented in the grammars or reported by the consultants. For each of the categories overlaps are listed separately, hence the repetitions of the configurations in the table. In the case of the multiple overlaps if a specific meaning of a category is not identified (due to missing information or simply because no further divisions are made within the category) only an abbreviation is used. Finally the numbers in brackets mark the number of items which, displaying a given syntactic overlaps, are also semantically polyfunctional (this problem is elaborated on later on in the chapter).

In the core set of sources of anteriority c-glossemes reconstructed on the basis of the analysis of patterns of syntactic overlaps, the spatial/temporal adpositions ‘after’ and various case markers occupy the central position. Temporal adverbs such as ‘then’, ‘afterwards’, nouns ‘back’ and ‘posteriority’ and verbs are less frequent although their position is stronger than that of the remaining categories. The latter ones are either cross-linguistically rare or occur as overlaps of semantically polyfunctional markers. The picture of the most common sources of c-glossemes together with hypothesized pathways of grammaticalization emerging from this part of analysis is depicted in (Fig.4.8.).

The major difference between this diagram and the one reconstructed on the basis of WLoG (cf. Fig.3.7.) is the absence of the category of case markers and adverbs from the latter one. The two categories appear to be very important sources of c-glossemes in the languages analysed in this study. Adverbs, as depicted in (Fig.4.8.) may be direct or indirect sources of ANTE linkers while case markers seems to be almost always the direct ones.⁴ On the other hand WLoG indicates nouns ‘trace, ’track’ and the verb ‘to pass’ as sources of the adposition ‘after’. They have not been found as overlaps of anteriority c-glossemes in my sample.



(Fig.4.8.) The most common sources of anteriority c-glossemes reconstructed on the basis of analysis of patterns of syntactic overlaps

The more general conclusion emerging from this part of the analysis is that the anteriority c-glossemes are in the majority of cases derived from items with spatial and space-related meanings/functions and temporal and time-related meanings/functions. Among the first group we find various locative adpositions (‘behind’, ‘after’ and others

⁴ The only example of a case marker occurring in a multiple overlap concerns a semantically polyfunctional c-glosseme (cf. Fig.4.9.).

marked as ‘locative’), locative and ablative case markers, adverb ‘in back’ and nouns ‘back’. The second group consist of temporal ‘after’ adpositions, nouns ‘posteriority’ and ‘time’, various adverbs of time (‘later’, ‘then’, ‘afterwards’) as well as verbs ‘to finish’, ‘to complete’.⁵

Before I move to the next subchapter, which also adds to the state of our knowledge on the origin of anteriority c-glossemes, there are two more observations I shall point out to the reader here. The first one concerns the other syntactic overlaps that have been omitted from the discussion so far. As was mentioned earlier on in this chapter, there are three most probable explanations for the cross-linguistic rarity of these overlaps: homonymy, language-specific polysemy or a complex scenario in which the overlaps are epiphenomena of the fact that one linker encodes synchronically a number of circumstantial meanings. The exclusively synchronic data collected for the purpose of this thesis do not allow for elaboration on the two first explanations. I therefore focus on analysis of the cases that could shed some light on the third one. In (Fig.4.9.) below all the anteriority c-glossemes that are semantically as well as syntactically polyfunctional are listed along with details on the characteristics of both types of overlaps.

anteriority c-glosseme	syntactic overlap	semantic overlap
Lillooet <i>ʔi-</i>	ADPlocative	TAQUO ‘since’ (temporal)
Ndyuka <i>baka di</i>	ADP‘behind’;‘after’(temporal) + NOUN‘back’	SIOVER ‘when’
Santali <i>-kate</i>	ADV‘then’	SIDUR ‘while’
Gola <i>wee</i>	ADP‘against’ + OBL	CAUSE, SIOVER, PURPOSE
Kanuri <i>duwo</i>	ADV‘after’;‘for a while’;‘first of all’ + CONJ‘but’ + REL	general c-glosseme
Mayogo <i>nedhinga</i>	NOUN‘time’	SIDUR-SIOVER-COMPAR
Quechua <i>-pita</i>	CASEablative	CAUSE, COMPAR, CONTRA
Lepcha <i>-ne/-nun</i>	CASEablative	CAUSE
Santali <i>-te</i>	CASEinstrumental;allative	SIDUR, MANNER, CAUSE
Burushaski <i>-ar</i>	CASEdative	PURPOSE
Galo <i>əəm</i>	CASEaccusative	general c-glosseme
Lezgian <i>-(i)z</i>	CASEdative + ADVLZ + INF	SIDUR-MANNER
Khwe <i>nò/nù</i>	CONJ‘and’	general c-glosseme
Burushaski <i>kè</i>	CONJ‘and’(NPs) + COMP	COND, SIOVER
Japanese <i>-to</i>	CONJ‘and’ + ADPcomitative(‘with’)	COND, SIOVER
Japanese <i>-te</i>	CONJ‘and’ + ADP + ASPprogressive	MANNER, CAUSE, CONC
Hindi <i>-kar/-ke</i>	VERB‘do’	CAUSE
Kayah Li <i>ma</i>	VERB‘be so’;‘be true’	general c-glosseme

⁵ The aspectual categories subsequent/continuative/completive that occur in the sample and are listed in (Fig.4.9.) could also be numbered among the members of this group.

Lango <i>àmɛ</i>	REL	POST, SIDUR
Ndyuka <i>di</i>	REL	SIOVER, CAUSE
SE Tepehuan <i>na</i>	PTCP + COMP + REL	general c-glosseme
Quechua <i>-sha</i>	PTCP + NMLZ	general c-glosseme
Tamil <i>-tu</i>	PTCP + COMP	TAQUO, CAUSE, MANNER
Ama <i>-mo</i>	TOPIC marker	COND, SIOVER
Paiwan <i>nu</i>	Partitive marker	COND, SIOVER
Lezgian <i>-na</i>	AOR	MANNER, SIDUR
English <i>-ing</i>	PTCP + NMLZ	CAUSE, MANNER, SIDUR
Khwe <i>-kò</i>	ASPcompletive;continuative;inchoati	general c-glosseme

(Fig.4.9.) Syntactically and semantically polyfunctional anteriority c-glossemes⁶

There are a couple of interesting observations that we can make by looking at the table. Starting from the top, we have 3 examples of semantically polyfunctional c-glossemes with scope over ANTE and other temporal relations. All three linkers have syntactic overlaps characteristic of the relation of anteriority – this could suggest that indeed, the circumstantial meaning of ‘after’ is the one that emerged in the first place while the others – ‘since’, ‘when’ and ‘while’ – are secondary developments. Also the two conjunctions quoted in (Fig.4.9.) and the single examples of a topic marker and a partitive marker provide us with interesting material for hypothesizing about pathways of grammaticalization since they all overlap with clause linkers that convey the meaning of anteriority as well as conditionality and temporal ‘when’ (see section 7.3.). Coordinating conjunctions and topic markers have been claimed by several authors to be sources of conditional subordinators (this issue is discussed in detail in chapter 7). Moreover, cases of markers conveying both COND and SIOVER relations as well as markers that have double readings of SIOVER and ANTE and triple readings ANTE-SIOVER-COND are by no means rare in the world’s languages (cf. sections 4.3. and 7.3.). Putting these two bits of evidence together we may suspect that at least some of these overlaps are of epiphenomenal character.

4.2.2. Polymorphemic markers and their internal structure

As reported in section 4.1., 40 of the 150 anteriority c-glossemes in the sample are polymorphemic structures. Since they have developed through more or less advanced fossilization of a group of morphemes, they are less likely than the monomorphemic linkers to overlap with other syntactic categories. Hence, as explained in sections 3.2.5. and 3.2.6., the insight into the processes of their grammaticalization may be obtained in the majority of cases only through decomposition of their structure.

⁶ The term *general c-glosseme*, as defined in section 3.3.1., refers to a c-glosseme which has 5 and more circumstantial readings.

The internal make-up of all the polymorphemic c-glossemes for which full relevant information has been obtained has been depicted schematically in (Fig. 4.10.). Each row of the table corresponds to one c-glosseme. If the language name is put in italics it means that the c-glosseme is distributed. The rest are linear. The number next to a language name indicates the number of morphemes incorporated in a given c-glosseme. The dots and ‘+’ symbols indicate the types of incorporated morphemes (+ means that the morpheme may act on its own as an anteriority linker in the particular language). Specific meanings within categories have been in some cases gathered into groups (eg. ADV ‘then’/‘later’/‘afterwards’) for ease of presentation; the same has been done with some categories (ADP/NOUN and ADP/ADV). The final column – labelled ‘other’ – contains morphemes that occurred only once or morphemes that are syntactically polyfunctional as well as those whose function has not been reported.

The summary reveals that adpositions and case markers, as well as syntactically polyfunctional adposition/adverb and adposition/noun items, are the most common building blocks of polymorphemic c-glossemes. This set of categories is strikingly similar to the one that emerged from the analysis of the patterns of syntactic polysemy (cf. Fig.4.8.). Moreover, not only the categories themselves but also their specific meanings/functions reported in the previous subchapter reoccur as most frequent in the complex linkers. Amongst adpositions the most commonly incorporated ones are those with the spatial/temporal (‘after’) and locative meaning, among adverbs we find exclusively adverbs of time of the ‘then/afterwards’ type. The range of nouns is limited to those meaning ‘back’ or ‘side’. Among case markers locatives are the most common ones. They are followed by those marking instrumental and instrumental/allative functions⁷. Finally, among verbs we have 2 cases of the verb ‘to finish’ and one of ‘go out’. The first one also occurred in the discussion on syntactic overlaps which supports Heine and Kuteva’s (2002) observations that verbs denoting motion often get grammaticalized into relational or adpositional or subordinating concepts. We should also add that among the incorporated material there are several other verb-specific categories indicating accomplishment of an event: 4 instances of items categorized as ‘completive markers’, as well as a perfective and a completive/continuous/inchoative aspect markers.

15 out of the 39 c-glossemes depicted in (Fig.4.10.) incorporate in their structure a morpheme that on its own can be used as an anteriority c-glosseme (marked by +). Nine of these complex linkers incorporate just one independent ANTE linker and the remaining six markers incorporate two. Among the latter ones four complex c-glossemes are made up exclusively of monomorphemic anteriority c-glossemes.

⁷ Cf. Aikhenvald’s (2008:594) statement on instrumental case-markers as common sources of temporal markers.

It is clear, that the material incorporated in polymorphemic markers is not random. The particular meanings/functions of the constituents are cognitively related to the concept of anteriority. In many cases it is also self-evident why a particular combination of morphemes became reanalysed as an anteriority c-glosseme. A good example here would be the combinations of nouns and locative or ablative case markers ('from back' in Galo, 'in back' in Basque), or the distributed c-glosseme in Akan (ADV'then' + DEM) as well as in Kanuri (combination of noun and adposition 'after back').

Language	ADP					ADV	ADP/ADV			ADP/ NOUN		NOUN	CASE					VERB		ASP		COMP/REL	DEM	NMLZ	COMPLET	C-GLOSS.			OTHER					
	'after'	locative/'behind'	'after'/locative	source	polyfunctional	'then'/'later'/'afterwards'	'after'/'then'	'behind'/'then'	'after'/ADV	'behind'/'back'/'side'	'after'/'back'	'back'	instr. and instr./all.	locative	ablative	comitative	dative (d)/genitive (g)	'go out'	'finish'	perfective	completive, continuative, inchoative						SIOVER	SOVER/COND	ANTE	other	polyfunctional	not known		
Akan (2)						•																	•											
Akan (2)						•																											•	
Arabic (2)								+																									•	
Basque (2)							•					•																						
Basque (2)											•		•									•												•
Basque (2)									•s				•																					
Basque (2)									•s								• G																	
Basque (3)							•					•										•												
Burushaski (2)																																		•
Burushaski (2)	+															+																		
Burushaski (2)	+																																	
Burushaski (2)	+																																	
Cubeo (2)							•																											
Estonian (3)	•																						•				•							
Galo (2)																	• D																	
Galo (2)													•																					
Galo (2)											•						• D																	
Galo (2)											•		•																					
Hausa (2)										+												•												
Hausa (2)											•						• G																	
Hindi (2)	•	•																																

Language	ADP					ADV	ADP/ADV			ADP/NOUN		NOUN	CASE					VERB		ASP		COMP/REL	DEM	NMLZ	COMPLET	C-GLOSS			OTHER		
	'after'	locative/'behind'	'after'/locative	source	polyfunctional	'then'/'later'/'afterwards'	'after'/'then'	'behind'/'then'	'after'/ADV	'behind'/'back'/'side'	'after'/'back'	'back'	instr. and instr./all.	locative	ablative	comitative	dative (d)/genitive (g)	'go out'	'finish'	perfective	completive, continuative, inchoative					SIOVER	SOVER/COND	ANTE	other	polyfunctional	not known
Ilocano (2)																		•													•
Japanese (2)	+																													•	•
Kanuri (2)		•									•																				
Ket (2)						•							•																		
Khwe (6)																		•			+				•	•			•	•	•
Khwe (6)																		•			+				•				+	•	•
Konso (2)																														+	
Konso (3)																											•			+	
Kryiol (2)	•				•																										
Lepcha (2)																															•
Maale (2)		•																													
Nez Perce (4)		•																													
Polish (3)			+																												•
Retuarã (2)		•																		•											
Santali (2)																															•
Swahili (2)					•			+																							
Thai (2)				•																											
	7	4	1	1	2		4	2	1	3	1		3	5	2	1	4	2	1	1	2				3	1	4		6	4	4
	15					3	7			4		5	15					3		3		3	3	4	4	8			14		

(Fig.4.10.) Material incorporated in polymorphemic anteriority c-glossemes

The content of the sub-column 'other' in column OTHER refers to: partitive marker for Basque; impersonal marker for the first Khwe c-glosseme and impersonal marker and CONJ 'and' for the second one; setting particle for both Konso linkers.

4.3. SEMANTIC POLYFUNCTIONALITY AND COGNITIVE AFFINITY

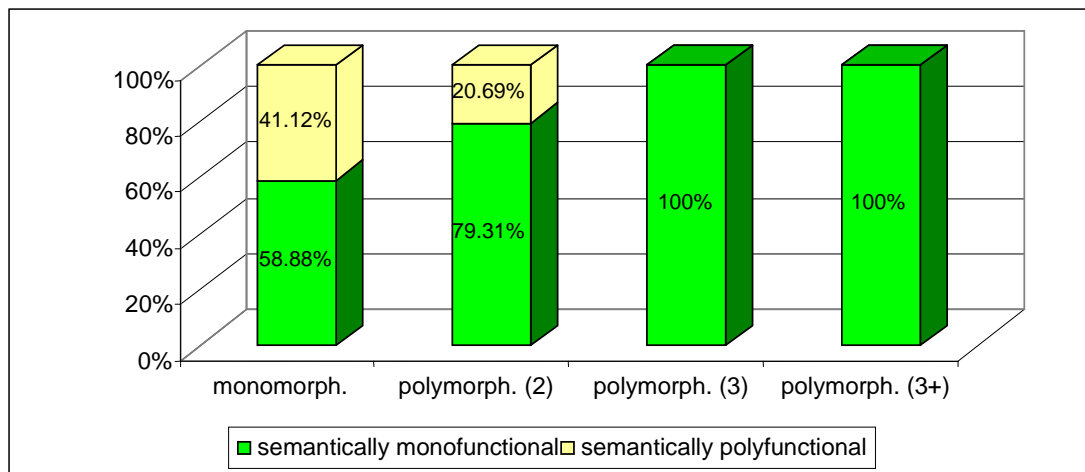
The last element of the analysis of origin of anteriority c-glossemes, which, as argued in section 3.3., can also provide us with interesting material for the study of cognitive organization of circumstantial concepts in our mind, concerns the phenomena of various circumstantial readings of clause linkers.

Among the 150 anteriority markers in the database, 50 have been classified as semantically polyfunctional. For 143 linkers there is full information available regarding their semantic mono/polyfunctionality as well as morphological complexity. The summary of the findings is presented in (Fig.4.11.). The data in the table show clearly that the ratio of semantically mono- and polyfunctional markers changes when morphological complexity changes.

(Fig.4.12.) illustrates the results graphically indicating the percentages of particular clusters of values. The proportion of semantically polyfunctional markers of anteriority is highest for the monomorphemic markers (41.12%). The underspecified c-glossemes constitutes also over 20% of the bimorphemic linkers but are absent from the more morphologically complex ones. Markers which consist of 3 and more morphemes are exclusively monofunctional.

	semantically monofunctional		semantically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	63	44.06%	44	30.76%	107
2 morphemes	23	16.08%	6	4.20%	29
3 morphemes	4	2.80%	-	-	4
3+ morphemes	3	2.10%	-	-	3
TOTAL	93	65.04%	50	34.96%	143

(Fig.4.11.) Distribution of c-glossemes of anteriority according to their semantic mono/polyfunctionality and morphological complexity



(Fig.4.12.) Ratio of semantically monofunctional and polyfunctional markers in anteriority c-glossems with different morphological complexity

The *complexity* → *monofunctionality* tendency (the more morphologically complex a marker is the less likely it is to be semantically polyfunctional) which holds here can be easily explained by the fact that polymorphemic linkers being fossilized but still transparent structures consisting of lexical (or grammatical) elements whose meaning/function is related to the concept of anteriority are less likely to undergo semantic changes as easily as monomorphemic markers do.

Having discussed the frequency and types of individual overlaps, it is also worthwhile to look at the overall *degree of semantic polyfunctionality* of the 50 recorded anteriority c-glossems i.e. at the number of binary, ternary, quaternary etc. semantic overlaps we encounter among them. As shown in (Fig.4.13.), the number of c-glossems conveying a variety of circumstantial meanings is high – over 70% of all overlaps are cases when a linker covers 3 or more relations (including anteriority). Those anteriority markers which overlap with just one relation contribute 28% and so do those that have scope over 5 and more relations and have been classified as general c-glossems.

	scope over 2 relations	scope over 3 relations	scope over 4 relations	scope over 5+ relations	TOTAL
count	14	15	7	14	50
%	28.00%	30.00%	14.00%	28.00%	100%

(Fig.4.13.) Degree of semantic polyfunctionality of anteriority c-glossems

The findings may be linked to the observation of the correlation between semantic mono/polyfunctionality and morphological complexity reported in (Fig.4.11.). If we classify the semantically polyfunctional markers according to the scope of their

meanings we discover that the morphologically simpler a marker is, the more likely it is to be used to encode a variety of circumstantial meanings. As evident from (Fig.4.14.), the semantically polyfunctional bimorphemic c-glossems of anteriority cover up to three circumstantial relations⁸ while over 42% of the monomorphemic ones have scope over more than three relations.

	scope over 2 relations		scope over 3 relations		scope over 4 relations		scope over 5+relations	
	count	%	count	%	count	%	count	%
monomorphemic	9	20.00%	11	24.45%	5	11.11%	14	31.11%
polymorphemic (2 morphemes)	4	8.89%	2	4.44%	-	-	-	-

(Fig.4.14.) Distribution of c-glossems of anteriority according to their degree of polyfunctionality and morphological complexity

The 50 semantically polyfunctional markers of anteriority overlap in total with 12 other circumstantial relations. Altogether there are 65 cases of overlap⁹ as well as 14 general c-glossems whose scope includes anteriority. Among the polyfunctional anteriority linkers there are 14 cases of *binary semantic overlaps* (i.e. c-glossems that in addition to anteriority have only one more additional circumstantial reading). This amounts to 21.5% of all the overlaps if we exclude the general c-glossems and 17.7% if we include them. All the details are presented in (Fig.4.15).

	occurrences of overlaps	occurrences in binary overlaps	
		count	% of the total number of binary overlaps (14)
SIOVER	15	5	35.7%
SIDUR	12	3	21.5%
CAUSE	12	3	21.5%
COND	7	1	7.1%
MANNER/MEANS	6	-	-
TAQUO	4	1	7.1%
PURP	2	1	7.1%
POST	2	-	-
COMPAR	2	-	-
RESULT	1	-	-
CONC	1	-	-

⁸ As was shown in (Fig.4.11.) the trimorphemic and more complex c-glossems of anteriority are exclusively monofunctional.

⁹ The number takes into account each overlap individually. Hence for instance one semantically trifunctional marker that covers the relation of anteriority as well as simultaneity overlap ('when') and conditionality ('if') contributes two entries to the table: SIOVER and COND.

CONTRA	1	-	-
TOTAL	65	14	100%
general c-glossemes	14		
TOTAL including general c-glossemes	79		

(Fig .4.15.) Summary of semantic overlaps of anteriority c-glossemes

The data should be interpreted as follows: the first row indicates that there are 15 instances of overlaps of anteriority and simultaneity overlap ('when'), among which there are 5 cases of binary ANTE-SIOVER overlaps. Those 5 cases constitute 35.7% of all binary overlaps in the sample.

The most common semantic overlaps of the relation of anteriority are, as can be seen from the table, overlaps with two temporal relations: 'when' (SIOVER), 'while' (SIDUR) and causality. SIOVER is the most common one but it outstrips the other two only by 3 occurrences. These three relations make 60% of all anteriority overlaps in the sample. They are also the ones that most commonly occur as the only overlaps of anteriority i.e. as constituents of *binary semantic overlaps*. On the other hand, it needs to be noted that half of the 12 relations listed in (Fig.4.15.) contribute just one or two overlaps (which, in turn, gives almost 14% of all the overlaps). In (Fig 4.16.) below all of the semantic overlaps observed in the sample with exclusion of the overlaps brought by general c-glossemes are depicted graphically.¹⁰

It has been already evident from (Fig.4.15.) that anteriority overlaps frequently with other temporal relations and that the most common overlap in this group is the overlap with temporal 'when' (SIOVER), which is among all the temporal relations clearly the most general one. As Cristofaro puts it:

When relations (...) imply that the main and dependent SoAs overlap in their realization, even though the exact extend of the overlapping is unspecified and subject to variation. For instance a sentence like
/When the Nazis came to power/, Georg Grosz left Germany
 does not mean that the two SoAs take place at exactly the same temporal point. There might be an interval of some days, or even month or years between the two. On the other hand, in a sentence like
/When he entered the room/, she went out.
 it is normally assumed that the two SoAs are either simultaneous or separated by a very short interval. (2003:159)

It does not come as a surprise than that the SIOVER-ANTE as well as SIOVER-SIDUR-ANTE overlaps are not rare. Simultaneity duration (SIOVER) is, as a matter of fact, the second most common relation overlapping with anteriority – it contributes 12 overlaps three of which are binary. The other two temporal relations in the network of affinities of anteriority are posteriority ('before') and terminus a quo ('since'). They are far less

¹⁰ Depiction involving general c-glossemes is impossible on a two-dimensional surface.

frequent occurring 2 and 4 times respectively. An example of a c-glosseme that may be used to express several temporal relations is the Lezgian *-la* suffix that reveals a triple overlap of SIOVER (4.4a), ANTE (4.4b) and SIDUR (4.4c) or the Lango linker *àmé* which expresses ANTE, SIDUR and POST (Noonan 1992).

(4.4a) Lezgian (Haspelmath 1993:383)

Marf qwa-da-j-la *nük'-er-ni* *wiri čüniix že-da*
rain fall-FUT-PTCP-{SIOVER} sparrow-PL-also all hide ANTICAUS-FUT
‘When it rains, even sparrows all hide’

(4.4b) (ibidem:382)

Institut-ar kütäh-aj-la *abur xaji škola.di-z* *k'walax-iz xta-na*
Institute-PL finish-AOPTCP-{ANTE} they native school-DAT work-INF return-AOR
‘After finishing college, they returned to their native school to work.’

(4.4c) (ibidem:383)

Xürek ne-zwa-j-la *Sabir.a wiči-n* *buba.di-waj xabar q'u-na*
Meal eat-IMPRF-PTCP-{SIDUR} Sabir.ERG self-GEN father-ADEL news hold-AOR
‘While they were eating, Sabir asked his father.’

The question that arises here is whether in the above mentioned cases we may draw conclusions as for the directionality of semantic changes that lead to the emergence of the semantic polyfunctionalities. As already discussed in section 3.2.4., the combined analysis of semantic and syntactic overlaps allows us to conclude that at least in some cases a clause linker was first used to express the meaning of anteriority and only then developed other temporal readings. In other cases, however, there is not enough readily available evidence that would allow us to conclude beyond any doubts about the exact pathways of grammaticalization.

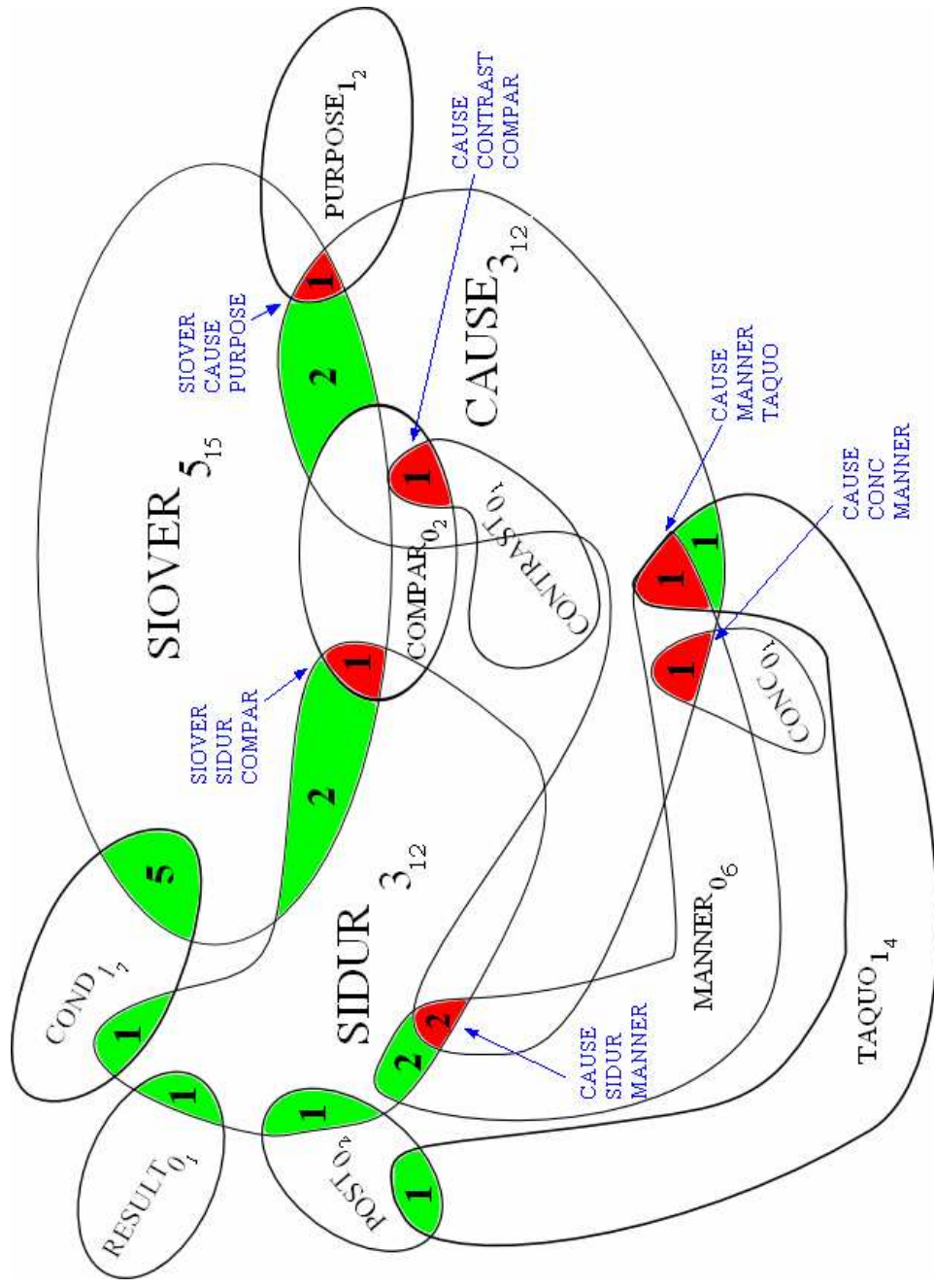
As frequent as the ANTE-SIDUR are also ANTE-CAUSE overlaps as exemplified by the Chukchi *-(i)neju* marker in (4.5a) and (4.5b.):

(4.5a) Chukchi (Dunn 1999:246)

ʒeqe-nijw-e pəkir-ineju *n-in-iw-qin*
bad-uncle-ERG approach-{ANTE} HAB-TR-say-3SG
‘The bad uncle after he approached said (...)’

(4.5b) Chukchi (ibidem:244)

ləγ ʃre-plətku-neju *γ-ekwet-lin jara-γtə*
really race-finish-{CAUSE} PRF-leave-3SG home-ALL
‘Since (he) finished racing he set off homewards.’



(Fig.4.16.) Semantic polyfunctionality of anteriority c-glossemes

The following convention is applied here: green indicates triple overlaps (i.e. anteriority and two more circumstantial relations), red indicates quadruple overlaps (anteriority and three more circumstantial meanings). Numbers in the overlapping areas indicate the number of overlaps. A bigger number next to the name of a relation indicates how many of the overlaps of the given relation with the relation of anteriority are binary overlaps. Subscripts indicate the total number of the overlaps involving the relation. Hence **PURP 1₂** should be read as: "There are 2 overlaps of anteriority and purpose in total and among them 1 overlap is a binary ANTE-PURPOSE overlap". The second overlap, as shown on the diagram, is an overlap with **CAUSE** which means that there is one c-glosseme that has scope over 4 different meanings: anteriority, purpose, causality and simultaneity overlap. For clarity in cases of quadruple overlaps there are descriptions listing the three relations that anteriority overlaps with in a particular case

Notably, beside three binary ANTE-CAUSE overlaps, there are multiple overlaps involving these two relations (as well as some other ones): ANTE-CAUSE-SIOVER (2 cases), ANTE-CAUSE-TAQUO (2 cases) and ANTE-CAUSE-SIDUR (2 cases). This proves clearly the affinity between causality and temporal order of events and the iconic motivation behind grammaticalization. As Thompson, Longacre and Hwang explain:

Two events which are mentioned together as being simultaneous or adjacent in time are often inferred to be causally related. (2007:247)

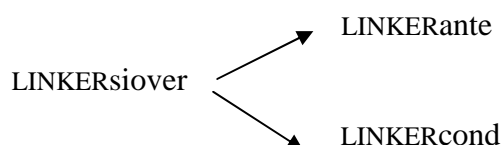
In cases of LINKERtemporal-CAUSE overlaps we may quite confidently conclude that it is the temporal meaning(s) that got reanalysed into the causal one. Such a pathway fits the less abstract meaning > more abstract meaning grammaticalization tendency and has been reported in other works on grammaticalization (cf. Hopper and Traugott 2003:187). However, in cases where a clause linker in addition to being a marker of anteriority and causality is used to convey other temporal meanings (such as the aforementioned TAQUO, SIDUR and SIOVER) it is impossible to propose any more detailed scenarios on the basis of synchronic material only.

Two other frequently occurring overlaps of anteriority are conditionality and manner. We could attempt to explain the affinity between ANTE and COND by referring to the fact that in both relations the SoA expressed in the main clause follows the SoA expressed in the supporting clause. It has to be acknowledged, however, that 5 of the 7 ANTE-COND overlaps involve also SIOVER. In the light of the fact that both ANTE-SIOVER and SIOVER-COND binary overlaps are very frequently attested, this might suggest that at least in some cases conditionality arose as an overlap of anteriority indirectly – i.e. through the extension of meaning of SIOVER or over SIOVER.¹¹ The set of plausible scenarios of development in cases involving ANTE, SIOVER and COND would than include two linear grammaticalization pathways:

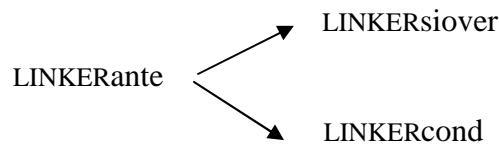
LINKERante → LINKERSiover → LINKERcond

LINKERSiover → LINKERante → LINKERcond

and two pathways in which some of the overlaps are of an epiphenomenal character:



¹¹ The affinity between SIOVER and COND is elaborated on in chapter 7.



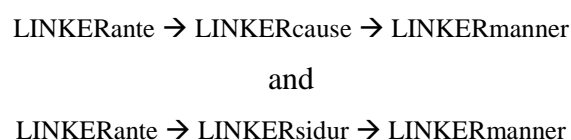
One of the examples of ANTE-COND overlaps comes from Basque where the complex c-glosseme consisting of instrumental case marker *-z* and postposition *gero* serves as an exponent of both the relations – cf. examples (3.4.) and (4.6):

(4.6.) Basque (Hualde, Ortiz de Urbina 2003:744)

Dirua eduki-z gero, baserria erosi-ko nuke
 money have-**{COND}** country.house buy.FUT AUX

‘If I had money, I’d buy a country house’

The explanation of the affinity between MANNER and ANTE is less self-evident. As can be seen in (Fig.4.16.) MANNER does not occur in a binary overlap with ANTE. When the two relations have one c-glosseme assigned to them they have also either a causal or durative (SIDUR) reading. It needs to be emphasized that many of the linkers classified in the grammars that I analysed as exponents of the MANNER relations could be also called exponents of the MEANS relation if we were to distinguish the two types as Thompson and Longacre (1985) or Kortmann (1997) do. Since the relation of MEANS/MANNER seems to have close links with both the causal relation and simultaneity duration (this issue is to be looked at closely in section 5.3.) it seems possible that MANNER came into the picture of anteriority overlaps through the agency of either SIDUR or CAUSE following the pathways:



Summarizing the results presented here we may conclude that:

- a) the relations that are cognitively most closely related to anteriority are: simultaneity overlap ‘when’, simultaneity duration ‘while’ and causality;
- b) terminus a quo ‘since’ and conditionality and manner are less frequent and among these three it seems that manner is the relation which in the majority of cases (if not in all) enters the network of semantic affinities of anteriority indirectly;
- c) the remaining relations depicted in (Fig.4.16.): result, purpose, comparison, contrast and concession should be viewed as more distantly related to anteriority since both the total number of their occurrences and the number of their occurrences in binary

overlaps with anteriority is significantly smaller than in the case of the other relations¹²;

- d) c-glossems of anteriority are often involved in more complex grammaticalization scenarios either due to their direct development into markers of other relations (causality, conditionality) or due to the fact that they may be marginal nodes in multiple (epiphenomenal) grammaticalization pathways;
- e) there is no convincing evidence for development of anteriority c-glossems out of markers of other circumstantial relations.

The final question we can address now is how the findings concerning the strongest semantic affinities of anteriority linkers presented here relate to those reported in Kortmann's study (1997). The tables in (Fig.4.17.) and (Fig.4.18.) present Kortmann's and my results respectively.

	I	II	III	IV	V	VI
ANTE (88 items)	SIOVER (68.2%)	CAUSE (43.2%)	IMMANTE (40.9%)	SIDUR (35.2%)	CONTIN (29.6%)	COND (26.1%)

(Fig.4.17.) The strongest semantic affinities of polyfunctional adverbial subordinators expressing anteriority according to Kortmann (1997:181)

The percentages represent how many of the 88 anteriority subordinators have a further reading of simultaneity overlap ('when'), causality ('because'), immediate anteriority ('as soon as') etc. Since a subordinator can have more than one reading, the numbers obviously do not add up to 100%.

ABSOLUTE NUMBER OF OCCURRENCES (65 overlaps)					
I	II	II	III	IV	V
SIOVER 15 items (23.1%)	CAUSE 12 items (18.5%)	SIDUR 12 items (18.5%)	COND 7 items (10.7%)	MANNER/MEANS 6 items (9.2%)	TAQUO 4 items (6.2%)
OCCURRENCES IN BINARY OVERLAPS (14 overlaps)					
I	II	II	III	IV	IV
SIOVER 5 items (35.7%)	CAUSE 3 items (21.5%)	SIDUR 3 items (21.5%)	COND 1 item (7.1%)	TAQUO 1 item (7.1%)	PURPOSE 1 item (7.1%)

(Fig.4.18.) The strongest semantic affinities of polyfunctional anteriority c-glossems revealed in this study

¹² The relation of posteriority ('before') which occurs in the network two times (in both cases in multiple overlaps) seems to be an exception here. It is not the cognitive distance between posteriority and anteriority but the fact that they are counter-concepts which explains the low number of their overlaps.

The percentages represent how many of the anteriority linkers overlap with other relations and what is the percent of these overlaps in the absolute number of occurrences and occurrences in binary overlaps respectively.

Kortmann has found that the dominating relation in the network of affinities of anteriority is simultaneity overlap ('when') which is followed by causality ('because'), immediate anteriority ('as soon as') and slightly less frequent simultaneity duration ('while'), contingency ('whenever') and conditionality ('if').

In my database the relations of simultaneity overlap also turned out to be the most frequent one both when it comes to the absolute number of overlaps and occurrences in binary overlaps. Similarly as in Kortmann's study causality has taken the second position also in my ranking. The similarities end there, however. The relations of contingency and immediate anteriority have not appeared at all in the analysis of semantic polyfunctionality of anteriority linkers in my study. Instead, several cases of MEANS/MANNER and TAQUO (temporal 'since') relations (as well as PURPOSE in the case of binary overlaps) have been reported. Conditionality, which in Kortmann's study occupies the last position in the table, takes the high third place in the table presenting my results, although its presence can be viewed as hardly significant when it comes to binary overlaps.

On the whole both studies revealed that the relations of SIOVER, CAUSE and SIDUR constitute the core set of the most common semantic overlaps of anteriority. The discrepancies may result from the differences in the designs of the two studies – both when it comes to the language sample and approach to analysis. In my data collection I did not use any pre-established list of circumstantial relations in order not to influence the judgements of the consultants. It seems to me, for instance, that the fact that Kortmann found so many overlaps of ANTE and IMMANTE is at least partially due to the technical problems of distinguishing between the meanings of 'after' and 'as soon as'. Since IMMANTE is nothing else than a subtype of ANTE it comes as no surprise that the differences between these two are not salient and so often go unnoticed. I believe that the high frequency of contingency ('whenever') reported in Kortmann's work may be explained in a similar manner – by referring to the affinity between contingency and conditionality (cf. section 7.3.). The frequency of the other relations that occur in my summary but do not occur in Kortmann's reports (namely means/manner, purpose and terminus a quo) is small and so it does not entitle us to claim that the results presented here significantly alter the picture that Kortmann draw on the basis of his analyses.

4.4. SUMMARY

The subject of analysis in this chapter has been the form and origin of markers of the temporal relation of anteriority and the semantic affinities with other circumstantial relations that the concept displays.

The analysis has revealed a variety of forms used to encode the meaning of anteriority in the world's languages. This includes free words, affixes, combinations of words and affixes, distributed markers and clitics. Almost 30% of the 150 anteriority markers in the database turned out to be morphologically complex. Both the morphologically simple (monomorphemic) and complex (polymorphemic) anteriority markers provided material for the reconstruction of the cross-linguistically most common sources of this group of linkers. The analyses of patterns of synchronic polysemy in section 4.2.1. revealed that the most common sources of anteriority linkers are adpositions, adverbs, nouns, case markers and verbs. The same set of categories emerged from the analysis of the material incorporated in the polymorphemic markers in section 4.2.2. In section 4.3., where results of analysis of cross-linguistic patterns of semantic polyfunctionality have been presented, it has been argued that there is no convincing evidence for development of anteriority c-glossems out of other types of circumstantial clause linkers while the opposite direction is not unusual. In the light of these facts we can treat the diagram in (Fig.4.8.) not only as a depiction of pathways of grammaticalization reconstructed on the basis of analysis of patterns of syntactic overlaps but as a complete picture emerging from all three components of the analysis. The findings are an important contribution to the works on grammaticalization which, so far, have not devoted attention to anteriority linkers. Finally, my analysis of the network of semantic affinities of anteriority linkers revealed its close connection to other temporal concepts as well as to the concept of causality and conditionality, confirming the findings reported by Kortmann (1997).

CHAPTER 5

Causality

5.1. MORPHOLOGICAL COMPLEXITY AND FORMS OF C-GLOSSEMES

In the 84 languages in the sample, 203 c-glossemes encoding the relation of causality (CAUSE henceforth) have been identified. Among them, full information on the morphological make-up has been reported for 186. As shown in (Fig.5.1.), it is, again, the monomorphemic c-glossemes that constitute the majority. At the same time, however, we notice that the bi- and trimorphemic markers are by no means rare: constituting respectively over 25% and over 7% off all the linkers in the dataset.

	count	%
monomorphemic	123	66.13%
2 morphemes	48	25.81%
3 morphemes	14	7.53%
3+ morphemes	1	0.53%
TOTAL	186	100%

(Fig.5.1.) Morphological complexity of causality c-glossemes

As for the morphological status of causality c-glossemes, as depicted in (Fig.5.2.) over three quarters of them are free-words (cf. example 1.10, 1.17) and almost 20% are affixes.

Among the latter ones we encounter 37 suffixes (including a Yup'ik postbase in a non-final position – in 5.1.) and 3 prefixes.

(5.1.) Central Alaskan Yup'ik (Mather, Meade, Miyaoka 2002:97)

Ciki-(rng)a-mki *quay-ut*
give-**{CAUSE}**.1SG-3PL glad-IND.3PL

‘Because I gave (something) to them, they (others) are glad.’

A combination of a free word and an affix occurs only once – in Basque. Discontinuous linkers occur three times in Akan and once in Boko:

(5.2.) Boko (Jones 1998:266)

Má kpá wàé ké ā tɛ ma zi yáí
 1SG.FUT give 3SG {CAUSE} 3SG.STAT follow 1SG.OBJ ADES {CAUSE}

‘I will give it to him, because he follows me’

None of the over 200 c-glossemes of causality is a clitic although one of the polymorphemic ones – the Hatam marker *leuo* – carries an enclitic *-o* (Reesink 1999:128)

	count		%
free word	158		77.83%
affix	40		19.71%
	suffix	prefix	
	37	3	
combination	1		0.49%
distributed	4		1.97%
clitic	0		0.00%
TOTAL	203		100%

(Fig.5.2.) Forms of causality c-glossemes

5.2. INSIGHT INTO ORIGINS

5.2.1. Syntactic polyfunctionality and patterns of polysemy

For the relation of causality, information on syntactic mono/polyfunctionality was available for 188 items. For 4 of the linkers it has not been established whether they are monomorphemic or polymorphemic. Hence, this part of the analysis is based on 184 of the 203 markers.

(Fig.5.3.) presents a detailed overview of the distribution of causality markers when it comes to the number of their polysemes. Polyfunctional c-glossemes constitute over 40% of all the linkers with the monomorphemic ones contributing 55 and polymorphemic ones 21 items in this group. The polymorphemic markers are, as in the case of anteriority, less likely to serve multiple syntactic functions. The difference in proportions between the monomorphemic and polymorphemic linkers in the groups of syntactically monofunctional and polyfunctional markers is smaller than in the case of anteriority. It is, nonetheless, still noticeable with 21 of the polymorphemic markers and 55 of the monomorphemic ones being syntactically polyfunctional.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	62	33.70%	55	29.89%	117
polymorphemic	46	25.00%	21	11.41%	67
TOTAL	108	58.70%	76	41.30%	184

(Fig.5.3.) Distribution of c-glossems of causality according to their syntactic mono/polyfunctionality and morphological complexity

As can be seen from (Fig.5.4), the polymorphemic markers are twice as likely to be monofunctional than polyfunctional. Interestingly, over half of the c-glossems made up of more than 3 markers are syntactically polyfunctional, while the same holds for one third of the bimorphemic markers.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
2 morphemes	35	55.56%	13	20.63%	48
3 morphemes	7	11.11%	7	11.11%	14
3+ morphemes	-	-	1	1.59%	1
TOTAL	42	66.67%	21	33.33%	63

(Fig.5.4.) Distribution of syntactic mono- and polyfunctionality in polymorphic causality c-glossems

Nonetheless we may conclude that also here the *complexity* → *monofunctionality tendency* (cf. section 4.2.1) is maintained: the more morphologically complex markers are less likely to share their form with other syntactic categories. As we might expect, the degree of syntactic polyfunctionality is also higher for the monomorphemic than for the polymorphemic linkers: almost a quarter of the monomorphemic linkers serve two or more additional functions in addition to being c-glossems. The same applies only to one of the 21 polymorphemic markers (Fig.5.5.).

	1 additional function	2 additional functions	3 additional functions	TOTAL
monomorphemic	42	9	4	55
polymorphemic	20	1	-	21

(Fig.5.5.) Degree of syntactic polyfunctionality of causality c-glossems

The analysis of the syntactic overlaps of the 76 syntactically polyfunctional c-glossems of causality revealed that the linkers overlap with over 10 other syntactic

categories (Fig.5.6). Among them the most frequent ones, with 50 items, are adpositions. They constitute over half of all the instances of overlaps and over 65% of all the binary overlaps. Case markers, which occur 10 times, are the second most common group of polysemes. They are closely followed by complementizers but, unlike that group, they occur exclusively in binary overlaps.

		I occurrences	II occurrences in binary overlaps	III occurrences in multiple overlaps
1.	ADPOSITION	50	42	8
2.	CASE	10	10	-
3.	COMPLEMENTIZER	8	3	5
4.	ADVERB	5	1	4
5.	CONJUNCTION	4	2	2
6.	RELATIVE CLAUSE MARKER	3	2	1
7.	VERB	2	1	1
8.	NOUN	2	1	1
9.	PARTICIPLE MARKER	4	-	4
10.	NOMINALIZER	4	-	4
11.	ASPECT MARKER	2	1	1
12.	OBLIQUE MARKER	2	-	2
13.	POSSESSIVE MARKER	1	1	-
TOTAL		97	64	33

(Fig.5.6.) Syntactic overlaps of causality c-glossemes

Application of the first and fourth assumption of the *grammaticalization heuristics* (cf. section 3.2.9.) allows us to conclude initially that these three categories: adposition, case marker and complementizer are, as far as a study limited to synchronic observations may reveal, the most common sources of causality linkers.

In order to apply the remaining assumptions of the heuristics we need to turn to the specifics concerning the types of meanings/functions of the categories listed in (Fig.5.6.) and data concerning semantic polyfunctionality of the polysemous c-glossemes. All the necessary details are listed in (Fig.5.7.).

The summary begins with an overview of ADP-LINKERcausality overlaps. Among the 42 c-glossemes overlapping exclusively with adpositions 23 share their morphological form with adpositions of reason ('because of') and as many as 19 of these are semantically monofunctional. There are also 6 cases of ADPreason-LINKERcausality overlaps where the adpositions is used to encode other meanings/functions as well (instrumental, source, 'for' etc.). The second most frequent

group of adpositions are adpositions of benefit/purpose.¹ However, only in 3 out of 8 cases, where the causality c-glosseme overlaps with ADPbenefit/purpose, the linker is semantically monofunctional. In the 5 other cases it is polysemous with a c-glosseme of purpose. This is unsurprising if we take into account the meaning of the ADP that apparently gave rise to the c-glosseme. Generally, we may expect that, at least in some cases where the overlaps of c-glossemes of causality and ADPpurposive/benefactive occur, the overlapping may be explained by either:

- a) a complex process involving reanalysis of purpose/benefit adposition into a clause linker of purpose which, in turn, extended its meaning also to the relation of causality (in such case the ADPbenefit/purpose-LINKERcausality overlap would be classified as an *epiphenomenon* as discussed in section 3.2.4.)

or

- b) a process initiated by a semantically polyfunctional ADPreason;benefactive;purposive where each of the meanings gave rise to one clause linker (cf. discussion on *polygrammaticalization* in section 3.2.4.).

The third most numerous category of adpositions polysemous with causality linkers are adpositions of source ('from'). Notably, all the linkers with ADPsource overlaps are semantically monofunctional which, according to the grammaticalization heuristics (cf. section 3.2.9.), qualifies them as one of the very likely direct sources of causality markers. The same three types of adpositions as described above occur also in the third column of (Fig.5.7.) which contains information on those c-glossemes that have polysemes in more than one category. This strengthens the proposed scenarios.

The information on multiple cases of polysemy/homonymy involving ADP-LINKERcause overlaps provides us with at least three additional clues as for the routes of grammaticalization. The first two concern the role of nouns and verbs in the emergence of causality linkers. The data from Dagur (Martin 1960:53-54) suggest that we are dealing with a process of grammaticalization of a noun 'reason', 'source' into an adposition of reason and benefit/purpose, which, in turn, gave rise to the clause linker of causality and purpose. The synchronic evidence from Eipo (Heeschen 1998:194) seems to be a trace of a process which was initiated by a grammaticalization of the verb 'think' into an adposition of reason which then gave rise to the c-glosseme of causality (as well as purpose)².

¹ In several cases the grammars I analysed did not specify whether the adposition translated as 'for' serves the function of expressing both the meaning of purpose (as in *He did it for money*) and of benefit (*He did it for her*) or just one of them. In such cases in the table a general gloss 'for' is given.

² For both languages look also at (Fig.5.8.) below.

POLYSEMIOUS/ HOMONYMOUS CATEGORY	CHARACTERISTICS OF POLYFUNCTIONALITY	
ADPOSITION Total:50	BINARY OVERLAPS:42 ADPreason:23 (4) ADPsource:5 ADPbenefit;purpose;'for':5 (3) ADPreason;benefit:2 (1) ADPreason;'for':1 (1) ADPreason;source:2 ADPreason;'according to':1 ADPreason;instrument:1 ADPpossesive:1 ADP'(immediately) after': 1 (1)	MULTIPLE OVERLAPS:8 ADP + ADVERB:3 ADP'since' + ADV'since':2 (1) ADPreason; benefit; directionality + ADVcomparative + COMP:1 (1) ADP + NOUN:1 (1) ADPreason;'for' + NOUN'reason', 'source' ADP + VERB:1 (1) ADPreason + VERB'think' ADP + other:3 ADPsource + NMLZ:1 ADP'against' + OBL:1 (1) ADP + CONJ'and' + ASPprog:1 (1)
CASE Total:10	BINARY OVERLAPS:10 CASEablative:4 (2) CASEdative:1 (1) CASEgenitive:1 (1) CASEadessive:1 (1) CASEinstrumental:1 (1) CASEinstrumental;allative:1 (1) CASEinalative:1	MULTIPLE OVERLAPS:0

COMPLEMENTIZER Total:8	BINARY OVERLAPS: 3 (3)	MULTIPLE OVERLAPS:5 COMP + ADVcomparative + ADPreason; benefit; directionality:1 (1) COMP + ADV'also' + NMLZ:1 (1) COMP + OBLIQ:1 (1) COMP + PTCP:1 (1) COMP + REL + PTCP:1 (1)
ADVERB Total:5	BINARY OVERLAPS:1 ADVdegree;manner;quality:1 (1)	MULTIPLE OVERLAPS:4 ADV'since' + ADP'since':1 (1) ADVcomparative + ADPreason; benefit; directionality + COMP:1 (1) ADV'also' + COMP + NMLZ:1 (1) ADV'then' + CONJ'but':1 (1)
CONJUNCTION Total:4	BINARY OVERLAPS:2 CONJ'and':1 CONJ'or':1 (1)	MULTIPLE OVERLAPS:2 CONJ'and' + ASPprog + ADP:1 (1) CONJ'but' + ADV'then':1 (1)
RELATIVE CLAUSE MARKER Total:3	BINARY OVERLAPS:2 (1)	MULTIPLE OVERLAPS:1 (1) REL + COMP + PTCP
VERB Total:2	BINARY OVERLAPS:1 (1) VERB'be so';'be true'	MULTIPLE OVERLAPS:1 VERBthink + ADPreason

NOUN Total:2	BINARY OVERLAPS:1 NOUN'payment'	MULTIPLE OVERLAPS:1 (1) NOUN'reason', 'cause' + ADPreason;'for'
PARTICIPLE MARKER Total:4	BINARY OVERLAPS:0	MULTIPLE OVERLAPS:4 <i>PTCP + COMP:1 (1)</i> <i>PTCP + COMP + REL:1 (1)</i> <i>PTCP + NMLZ:2 (2)</i>
NOMINALIZER Total:4	BINARY OVERLAPS:0	MULTIPLE OVERLAPS: 4 NMLZ + COMP + ADV'also': 1 (1) NMLZ + ADPsource: 1 <i>NMLZ + PTCP: 2 (2)</i>
ASPECT MARKER Total:2	BINARY OVERLAPS:1 ASPcontinuative;inchoative;completive:1 (1)	MULTIPLE OVERLAPS:1 ASPprog + CONJ'and' + ADP:1 (1)
OBLIQUE MARKER Total:2	BINARY OVERLAPS: 0	MULTIPLE OVERLAPS:2 OBL + ADP'against':1 (1) <i>OBL + COMP:1 (1)</i>

(Fig.5.7.) Details of syntactic overlaps of causality c-glossemes

The conventions of data presentation are the same as described for (Fig.4.7.) in chapter 4.

Acknowledging that only historical data could confirm these suspicions, we have to admit that the pathways described here would fit into and complement the findings presented by Heine and Kuteva (2002). On numerous occasions the authors emphasize that two groups of verbs: process verbs ('go to', 'give', 'follow', 'leave') and verbs denoting location or motion on the basis of some salient semantic property give rise to markers of case relations or adpositional concepts which develop further into markers of grammatical relations between clauses. As for the markers of causality they have also listed the verb 'say' (cf. Fig.3.4.) as a common source of clause linkers. If we add to it the verb 'think', described above, we could conclude that not only verbs of motion and process verbs but also verbs of intellectual activity give rise to linkers of causality. Similarly, in the case of nouns Heine and Kuteva (2002) conclude that certain generic ('matter', 'fact') and relational nouns (including nouns for body parts e.g. 'back') develop over time to grammatical markers including clause linkers. The nouns 'reason' and 'cause', which I identified in the sample as polysemous with causality linkers, have not been mentioned by Heine and Kuteva in this context. I believe that the evidence presented in this and the following section give reasons strong enough to admit these lexical items to the group of the most common sources of causality markers.

The third clue coming from the observation of the Dagur *tuale* and Eipo *tennen* markers concerns grammaticalization processes operating on c-glossemes themselves. As depicted in (Fig.5.8.) both the linkers are semantically polyfunctional. The first one is used to encode the meaning of causality, purpose and concession, the second one expresses interclausal relations of causality and purpose. In Dagur it is already the adposition that carries the meaning of both reason and benefit/purpose and hence it might have evolved directly into markers of two different interclausal relations. If this is what happened then we could say that we deal here with polygrammaticalization. Conversely, the Eipo clausal marker *tennen* has two polysemes whose semantic properties are characteristic of reason rather than purpose and, hence, in this case we have grounds to assume that the use of *tennen* as a purpose marker came about as an effect of extension of the meaning of the marker of causality rather than of direct grammaticalization of the adposition of reason into the marker of purpose. All the findings presented above suggest that the semantic overlap of purpose and cause, which is elaborated on in section 5.3, is an important clue for the quest for sources of causality markers.

The second most common syntactic category overlapping with c-glossemes of causality, as depicted in (Fig.5.6.) and (Fig.5.7.), is that of case markers. However, even if we follow all the assumptions of the grammaticalization heuristics it is difficult to indicate the specific function/meanings of case markers that the c-glossemes of causality most often become grammaticalized from. The reason for that is that almost all of the c-glossemes which are polysemous with case markers are also semantically polyfunctional (cf. entries with numbers in brackets in Fig.5.8.) expressing in majority of cases the meanings of purpose and various temporal relations. The most frequent case marker overlapping with causality is the ablative one which occurs three times. The ablative meaning indicates, naturally, contains semes of both location and source. The concept of ‘source’ has been already mentioned above, while discussing adpositions, as the one that apparently quite often triggers grammaticalization and leads to the development of a causal linker. This parallel, however, does not seem a good enough reason to make sound conclusions about direct grammaticalization of the ablative case marker into a causal c-glosseme. Apart from the relatively low number of occurrences, two of the three ablative markers in the sample are used also as markers of anteriority and *terminus a quo* (‘since’). This makes the situation quite complicated but if we apply the assumption of the development of more abstract meanings out of the less abstract ones we can conclude that it is the temporal meanings that preceded the causal one – an inference about causal relation between two SoAs is, by definition, possible only if the causal SoA is chronologically earlier than the SoA that is its consequence. Such inferences are a part of reasoning – based in the mental space, while temporal order of events is perceptual – it is a part of the external world. In other words, I propose to treat cases of CASEablative-LINKERtemporal_meaning-LINKERcausality polysemy as epiphenomenal results of a clash of two phenomena: grammaticalization of case markers into markers of temporal relations between these clauses and cognitive affinity between these relations and causality.³

causality linker	syntactic overlap	semantic overlap
Sango <i>ngbangati</i>	ADPreason	PURPOSE
Hausa <i>sòbo dà/sòbodà</i>	ADPreason	PURPOSE
Hausa <i>dòmin/don</i>	ADPreason	PURPOSE
Japanese <i>tame</i>	ADPreason;purpose	PURPOSE
Konso <i>mallá</i>	ADPreason	RESULT
Dagur <i>tuale</i>	ADPreason;‘for’ + NOUN‘reason’;‘source’	PURPOSE, CONC
Eipo <i>tennen</i>	ADPreason + VERB‘think’	PURPOSE
Kanuri <i>-ro</i>	ADPbenefit;reason;directionality + ADVcomparative + COMP	PURPOSE, COCOND, SUBSTI

³ See section 5.3. for more discussion.

Eipo <i>ate</i>	ADPreason;'for'	PURPOSE
Arabic San'ani ^ˁ <i>ala sibb</i>	ADP'for'	PURPOSE
Arabic San'ani ^ˁ <i>asibb</i>	ADP'for'	PURPOSE
English <i>for</i>	ADPbenefit;purpose	PURPOSE
Sango <i>tɛnɛti</i>	ADP'for'	PURPOSE
Polish <i>skoro</i>	ADP'(immediately) after'	COND
English <i>since</i>	ADPtemporal('since') + ADVtemporal('since')	TAQUO
Gola <i>wee</i>	ADPagainst + OBL	ANTE, SIOVER, PURPOSE
Japanese <i>-te</i>	ADP + CONJ'and' + ASPprogressive	MANNER, CONC, ANTE
English <i>as</i>	ADVdegree;manner;quality	SIMIL, MANNER, SIOVER
Apache Jicalrilla <i>-go/-o</i>	ADV'also' + COMP + NMLZ	SIOVER, COND
Au <i>-te</i>	ADV'then' + CONJ'but'	RESULT, PURPOSE, COND
Ket <i>-dinal</i>	CASEablative	TAQUO
Lepcha <i>-nu/-nun</i>	CASEablative	ANTE
Quechua Huallaga <i>-pita</i>	CASEablative	ANTE, COMPAR, CONTRA
Ket <i>-diŋta/-dita</i>	CASEadessive	PURPOSE
Konso <i>é</i>	CASEdative	PURPOSE
Krongo <i>má-</i>	CASEgenitive	PURPOSE
Tamil <i>-aal</i>	CASEinstrumental	SIOVER, COND
Santali <i>-te</i>	CASEinstrumental;allative	SIDUR, MANNER, ANTE
Estonian <i>et</i>	COMP	general c-glosseme
Warlpiri <i>yunga/yinga/yunu</i>	COMP	PURPOSE
Galo <i>ɛ̀mlà(a)</i>	COMP	PURPOSE
Paiwan <i>tu</i>	COMP + OBL	PURPOSE, MANNER
Tamil <i>-tu</i>	COMP + PTCP	MANNER, ANTE, TAQUO
SE Tepehuan <i>na</i>	COMP + REL + PTCP	general c-glosseme
Yami <i>ta</i>	CONJ'or'	PURPOSE
Kayah Li <i>ma</i>	VERB'be so'; 'be true'	general c-glosseme
Ndyuka <i>di</i>	REL	SIOVER, ANTE
Khwe <i>-kò</i>	ASPcontinuative;inchoative;complet.	general c-glosseme
English <i>-ing</i>	PTCP + NMLZ	CAUSE, MANNER, SIDUR

(Fig.5.8.) Syntactically and semantically polyfunctional causality markers

Among the case markers in (Fig.5.7.), we also find one example of dative and one of genitive case. This draws, once again, certain semantic parallels with the adpositions discussed above where we have come across numerous benefactive functions as well as one example of a possessive adposition.⁴ Similarly among both adpositions and case markers we find single examples of instrumental as well as locational/directional functions (cf. the allative/adessive/inlative case markers and the

⁴ Note also that in (Fig.5.6.) there is a category of “possessive marker” listed separately from adpositions. This marker comes from the Maale language (Amha 2001:189).

reason/benefit/directionality Kanuri adposition *-ro* in Fig.5.8. also listed in Fig.5.7.). Some of these semantic polyfunctionalities have been mentioned by Aikhenvald (2008) in the passage quoted earlier in section 3.2.8. Nonetheless, due to a small number of occurrences of the items listed here, it is only the first two – genitive and dative case markers – that thanks to their salient semantic properties suggest strongly a pathway of grammaticalization: first to the clausal markers of purpose and only from there to causality linkers.

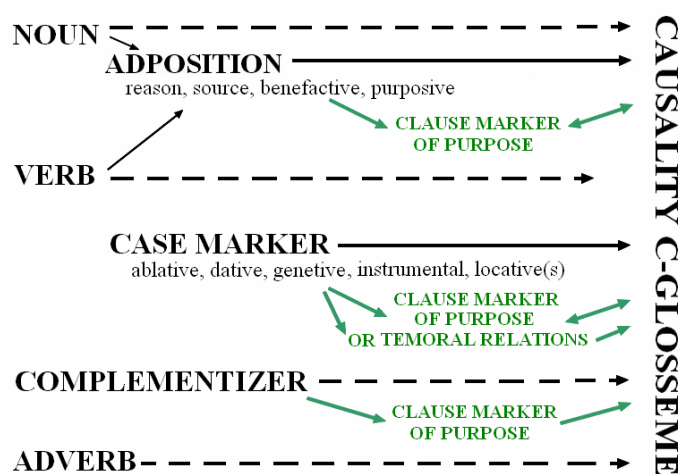
As I mentioned before, the rank table in (Fig.5.6.) suggests initially that complementizers are the third most common source of causality markers. However, a closer look at this syntactic category (cf. Fig.5.7.) weakens this hypothesis since all the items with polysemes in the group of complementizers are semantically polyfunctional c-glossemes. Moreover, as has already been emphasized, only 3 of the 8 markers in this group occur in binary overlaps. In (Fig.5.8.) we read that four of the items displaying the COMP-LINKER_{causality} overlap are also used to express the relation of purpose, two are used as general c-glossemes, and among the meanings encoded by the remaining two we find temporality as well as manner and conditionality. As discussed later on in chapter 6., complementizers quite often occur as the only syntactic overlaps of the c-glossemes of purpose and hence we may hypothesize that at least in some cases the COMP-LINKER_{causality} overlap is an epiphenomenon of a process where the marker of purpose clause, having developed from a complementizer, gave rise to marker of causal clause (and possibly also other clauses).⁵

Adverbs, which occur quite frequently as syntactic overlaps of the relation of anteriority, constitute only 5% of the overlaps of causality linkers. Moreover, like complementizers, all these adverbs are polysemous/homonymous with semantically polyfunctional circumstantial markers. The variety of meanings these markers cover and the fact that they display multiple syntactic overlaps in 4 out of 5 cases, forces us to retain the ADV-LINKER_{causality} grammaticalization pathway merely as possibly common cross-linguistically. The same can be said about many other syntactic categories listed in (Fig.5.6.) and (Fig.5.7.) – conjunctions, relative clause markers, aspect markers etc.

We may conclude that the two categories that can be viewed as cross-linguistically most frequent sources of causality c-glossemes are adpositions and case markers, as depicted in (Fig.5.9.). The data collected suggest also that at least in some cases these two categories may be indirect sources of causality linkers – i.e. they might have developed into the markers of interclausal relations of purpose or temporal relations (marked by green arrows in the diagram) and only then into the markers of causality. It is also possible that in other cases these two categories might have developed into various types of clause linkers independently (polygrammaticalization) or that it was the causality marker that developed into the marker of purpose (this is

⁵ Cf. discussion in section 3.2.8.

marked by the double-sided arrows). Although the evidence that nouns, verbs, adverbs and complementizers are direct sources of causality markers is weaker, it cannot be excluded. Hence the dotted lines in (Fig.5.9.). The data collected (especially when analysed together with the evidence presented in the chapter 7. where the encoding of purpose is discussed) suggests that it is more likely that complementizers could have developed into the markers of purpose which, in turn developed into the markers of causality then the other way round. Finally, although the development of nouns or verbs into adpositions and then into causality markers finds its confirmation only in two cases of syntactic polyfunctionality I have decided to include these pathways in the diagram since, as discussed in the following section, these two categories often appear as building blocks of synchronically polymorphemic causality c-glossems.



(Fig.5.9.) Reconstruction of the most common sources of causality markers

Interestingly, in the material I collected, the nouns that Heine and Kuteva (2002, cf. section 3.2.8.) have listed as giving rise to causality markers (i.e. ‘back’ – body part, ‘matter’, ‘place’) as well as adverbs of place (‘here’) and temporal adpositions are virtually absent. On the other hand, the results of my study suggest that the list of the most common sources of causality linkers may be much longer and more diversified than the authors of WLoG presented it.

Overall, on the basis of analysis of the patterns of polysemy, it may be stated that causality linkers find their sources in a variety of syntactic categories with meanings/functions related to the concepts of reason, origin (source), purpose and benefactiveness, location (and directionality), possessiveness as well as categories with temporal meanings. The search for the sources is continued, from slightly different angle, in the following section.

5.2.2. Polymorphemic markers and their internal structure

As reported in section 5.1., among the 84 languages included in this study 63 polymorphemic markers of causality have been identified. The linkers may be synchronically viewed as phrases: demonstrative, as in Khwe in (5.3.) or prepositional, as in Polish in (5.4.), or as non-phrasal strings of morphemes – see the Hatam example in (5.5.).

(5.3.) Khwe (Kilian-Hatz 2008:336)

Ti #’óm-á-tè vé **kóáná-hɛ** k’án n|gónó-á-tè
 1SG sleep-1SG-PRS NEG {reason-3SG.F→CAUSE} very be.noisy-1SG-PRS

‘I cannot sleep because it is very noisy.’

(5.4.) Polish

Z *powod-u* opóźni-eni-a pociąg-u nie
 {from reason-GEN→CAUSE} delay-NMLZ-GEN train-SG.GEN NEG
zdaqży-l-i na *samolot*
 make.on.time-M.PST-3.PL for airplane.ACC

‘Because the train was delayed they missed the flight.’

(5.5.) Hatam (Reesink 1999:128)

Paulus lene ngat Disyon cig **leu=o** mai *big-yo*
 Paulus then see Disyon father {from=or→CAUSE} die NEG-yet

‘Paulus then has seen Disyon’s father for he hadn’t died yet.’

An inflected noun as an exponent of the causal relation occurs only once in the sample – in Galo (5.6), while Lezgian provides us with three examples of verbs in their converbal form, one of which is *kilinga* (aorist converb of *kiligun* ‘look’ – see 1.17) and the other two are converbs of *luhun* ‘say’: *lahana* and *luhuz* (5.7).

(5.6.) Galo (Post 2007:792)

Hogó=əí *dó-rə-kú* *əm-nam* **lagàa=bə**
 SPRX.LOC=EMP eat-IRR-COMP say-NMLZ-REAL {reason=DAT→CAUSE}

‘Because of saying “I’ll eat it right here” (...)...

(5.7.) Lezgian (Hespelmath 1993:390)

Bazar.di-n *juǰ* *ada-z*, *tars-ar* *awa-č*
 Sunday-GEN day he-DAT lesson-PL be.in-NEG

luhu-z *tak’an* *xâ-nwa-j*
 {say-DAT/INFCONV/IMPRFCONV→CAUSE} hateful become-PRF-PST

‘He hated Sunday because there were no lessons.’

As already mentioned and exemplified in section 5.1., we also encounter four discontinuous markers in which one of the elements is optional. Details concerning the internal structure of all the polymorphemic markers are depicted in (Fig.5.10.).

The analyses reveal that adpositions and nouns are by far the largest groups providing building materials for causality markers. The specific meanings/function of the adpositions that have been discussed in the previous section (reason, locative, benefactive, possessive, source) reoccur here, too. Nouns, which occurred only twice as polysemes of causality c-glossemes, appear 23 times as material incorporated in polymorphemic linkers. Almost half of these nouns convey the meaning of reason/cause. Interestingly, the semantic polyfunctionality of adpositions covering the meanings of reason, source, benefit and purpose is observable also among nouns (cf. Galo and Ilokano in Fig.5.10). Apart from these, the nouns incorporated in the structure of polymorphemic markers convey such generic meanings as ‘matter’ and ‘thing’ as well as a variety of other meanings listed in the captions under the table (majority of them come from English polymorphemic linkers).

The other categories frequently occurring in polymorphemic causality c-glossemes are verbs and case markers. Among the first ones, it is the verb ‘say’ that occurs most frequently. Heine and Kuteva (2002, cf. Fig. 3.4.) have listed the verb as one of the sources of causal clause markers indicating that it often develops first into a complementizer and purpose marker. The data I have collected provide some more examples of causal markers employing the verb ‘say’ that support this pathway of grammaticalization: in Lezgian the verb occurs twice in its converbal form, in Galo it is accompanied by a non-finiteness marker, in Thai and in Sango it is polysemous with complementizer and in the last one also with the noun ‘word’, ‘speech’.

Apart from the verb ‘say’, in the dataset we find also ‘owe’, ‘able’, ‘look’ and ‘be’ and although they are represented by single examples, it is clear that they are part of the wider tendency where concepts related to perception, possession and processes serve as a basis for development of causality markers. Similarly, the same categories of case markers that have been discussed in the previous chapter occur among the morphemes incorporated in the polymorphemic causality linkers. This includes genitive, dative, instrumental as well as those that encode the role of source or direction and location (mutative, adressive, ablative). Interestingly, only in two cases, coming from Polish and Galo, these markers form causality c-glossemes by combination with a noun. In both cases the noun encodes the meaning of ‘reason’.

Yet another category, whose presence cannot be left unmentioned, comprises interrogatives. Their role in formation of clause linkers in Indo-European languages has already been recalled in section 3.2.8. In my dataset they occur as material incorporated in polymorphemic causality morphemes in Basque (isolate), Supyire and Swahili (both Niger-Kordofanian), Santali (Austic) and Hindi (Indo-European). This

may suggest that interrogatives are, in fact, an important element in formation of causality markers also outside the Indo-European circle.⁶ 4 of the 7 interrogatives are interrogative pronouns ‘what’, and 3 interrogatives of reason (‘why’). And here we may note again quite an interesting fact – none of these interrogatives occur with either noun or verb. They combine with a case marker (in Basque), complementizer (in Hindi), conjunction (in Santali), locative adposition and question particle (in Supyire), and with adposition ‘for’ (in Swahili).

As for the other categories, we notice several cases of demonstratives, complementizers, relativizers and complementizer/relativizer/participle polysemes as well as four conjunctions (this includes two disjunctive and two adversative but no coordinating conjunctions). Worth attention is also the fact that 12 of the 63 polymorphemic causality c-glossems listed in (Fig.5.10.) contain elements that can be used on their own as markers of causality. This includes 3 morphemes that find their polysemes in other syntactic categories (marked by the ‘+’ symbol) and 9 that do not have synchronically identifiable polysemes/homonyms (listed in the ‘causality c-glossems’ column of the table).

As a final remark we may add that among the total number of 148 morphemes making up the polymorphemic markers that are listed in (Fig.5.10.) three are reported to be borrowings. This includes the Spanish subordinator *porque* in *porque nagu*’ in Southeastern Tepehuan (the only double-marked causality c-glossems in the sample), the Lao noun *jǎn* ‘thing’ in Sapuan, and the Arabic *sababu* ‘cause’ in Swahili.⁷

From the material discussed here we can conclude that the meanings/functions of the morphemes incorporated in the structure of polymorphemic markers in the majority of cases come from the semantic space built around the concepts of reason, origin (source), purpose and benefactiveness, location (and directionality), possessiveness and perception. This, again, proves that the processes of idiomatization/fossilization are not random. On the contrary, even a synchronic analysis, as presented here, reveals certain clues as to the metaphorical extensions and motivations that lead to the emergence of causality markers. I will not attempt, however, to enter into this wide and complicated topic here.

⁶ We cannot exclude here the possibility that the grammaticalization in these cases is externally motivated (see section 3.2.7.) – all of the languages (apart from Hindi which belongs to the Indo-European family) incorporating interrogatives in the structure of polymorphemic causality markers are in close contact with major Indo-European languages.

⁷ There is also the Shilha linker *cala khatir* which is known to contain an Arabic Tunisian loanword but, since no further details are available, it has not been added to the list in (Fig.5.10.).

Language	ADP						ADP + other	NOUN						COMP	REL	COMP/REL/PTCP	DEF/REL	DEM	CAUSALITY C-GLOSSEM	VERB			CASE	INTER	CONJ	OTHER	NOT KNOWN	
	reason	possesive	locative	benefactive	other	not classified		'reason'/'cause'	'reason'/'source'	reason/purpose/'benefit'	'matter'	'thing'	other							'say'	other	'say' + other						
Japanese (2)							•	•																				
Japanese (2)							•	•																				
Japanese (2)							•	•																				
Kanuri (2)																		•										•
Kanuri (2)							+											•										
Ket (2)																												•
Ket (2)																												•
Khwe (2)								•																				•
Lango (2)						•																						•
Lango(2)																												•
Lavukaleve (3)																												•
Lezgian (2)																												•
Lezgian (2)																												•
Lezgian (2)																												•
Lillooet (2)																												•
Mayogo (2)								•																				•
Ndyuka (2)																												•
Polish (2)																												•
Polish (3)								•																				•
Polish (3)																												•
Sango (2)		•																										•
Sango (2)		•																										•
Santali (2)																												•
SE Tepehuan (2)																												•
SE Tepehuan (2)																												•
Sapuan (3)																												•
Supyire (2)			•																									•
Supyire (2)			•																									•
Swahili (2)								•	▼																			•
Swahili (2)								•	▼																			•

5.3 Semantic polyfunctionality and cognitive affinity

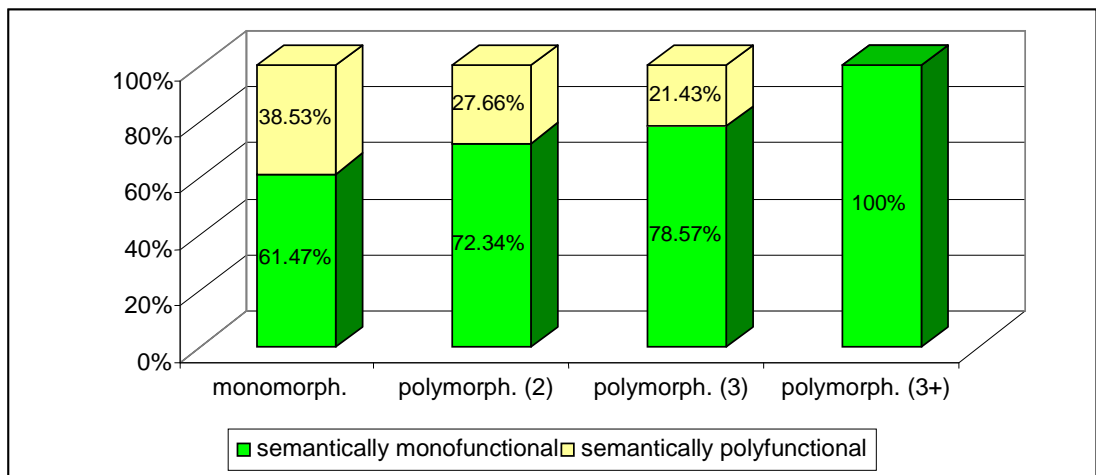
Out of the 203 markers of causality in the sample, information on their semantic scope is known for 184. 63 of them are also used to express circumstantial meanings other than causality (and are hence called, as have been already explained, semantically polyfunctional). 12 of these have been classified as general c-glossemes since they can serve as exponents of 5 or more circumstantial relations. Fig.5.11. presents general information on the semantic mono/polyfunctionality and morphological complexity of the analysed items. The data show clearly that the share of semantically polyfunctional markers decreases with the increase in morphological complexity.

	semantically monofunctional		semantically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	75	40.76%	47	25.54%	122
polymorphemic	34	18.48%	13	7.07%	47
polymorphemic	11	5.98%	3	1.63%	14
polymorphemic	1	0.54%	-	0.00%	1
TOTAL	121	65.76%	63	34.24%	184

(Fig.5.11.) Distribution of c-glossemes of causality according to their semantic mono/polyfunctionality and morphological complexity

By looking at particular categories separately (Fig.5.12.) the differences in ratios become even more evident – the linkers which consist of three morphemes are semantically polyfunctional in just over 20% of the cases, while the same is true about almost 40% of monomorphemic markers and almost 28% of c-glossemes made of two morphemes.⁸ Although the differences are not as striking as the ones observed for anteriority, the *complexity* → *monofunctionality tendency* holds also in the domain of semantic polyfunctionality of causality markers.

⁸ As shown in (Fig. 5.11.), there is only one c-glosseme of causality made of more than 3 morphemes and, unsurprisingly, it is semantically monofunctional.



(Fig.5.12.) Ratio of semantically monofunctional and polyfunctional markers in causality c-glossemes with different morphological complexity

Following the same path as in the earlier chapter, let us look also at the data on the number of circumstantial relations the c-glossemes of causality have scope over – i.e. at their degree of polyfunctionality (Fig.5.13.).

	scope over 2 relations	scope over 3 relations	scope over 4 relations	scope over 5+ relations	TOTAL
count	33	7	10	13	63
%	52.38%	11.11%	15.87%	20.64%	100%

(Fig.5.13.) Degree of semantic polyfunctionality of causality c-glossemes

As the table illustrates, over 52% of all cases of overlaps are binary overlaps, while the linkers with scope over 3, 4, and more than 5 relations, are considerably less frequent. By splitting and categorizing the findings according to the morphological complexity of the linkers (Fig.5.14.) we can see that monomorphemic c-glossemes, with just one exception, have a monopoly on expressing three and more relations. The polymorphemic and semantically polyfunctional markers are in 15 out of 16 cases bifunctional. The same is the case only for slightly over 42% of the monomorphemic markers. These more detailed observations once again indicate the presence of the complexity → monofunctionality tendency.

	scope over 2 relations		scope over 3 relations		scope over 4 relations		scope over 5+ relations	
	count	%	count	%	count	%	count	%
monomorphemic	19	31.15 %		9.84%	10	16.39%	10	16.39%
2 morphemes	13	21.31%	-	-	-	-	-	-
3 morphemes	2	3.28 %	1	1.64%	-	-	-	-

(Fig.5.14.) Distribution of c-glossemes of causality according to their degree of polyfunctionality and morphological complexity

The semantic network in which the 63 polyfunctional markers exist covers 15 circumstantial relations. Altogether, as shown in (Fig.5.15.), this amounts to 85 instances of semantic overlaps.⁹ The overlaps of CAUSE and PURPOSE significantly outnumber all other overlaps. This concerns both the binary overlaps and all the overlaps taken together. In fact, the CAUSE-PURPOSE overlaps amount to almost 60% of all the 37 binary overlaps.

The next two frequent semantic overlaps of causality linkers are those with temporal relations of anteriority and simultaneity overlap ('when') each of which contributes 8.11% of the total number of identified binary overlaps. Over 70% (16 items) of the c-glossemes displaying CAUSE-SIOVER and CAUSE-ANTE overlaps are, however, used also to express other circumstantial meanings. The two other temporal relations in (Fig.5.15.) – TAQUO and SIDUR – are far less frequent and occur only several times with no significant contribution to the group of binary overlaps. Among the remaining relations those with the highest score are MANNER, COND and RESULT the first one contributing almost twice as many cases as the two latter ones.

	occurrences of overlaps	occurrences in binary overlaps	
		count	% of the total number of binary overlaps (37)
PURPOSE	30	22	59.45%
ANTE	12	3	8.11%
SIOVER	10	3	8.11%
MANNER	9	2	5.41%
RESULT	5	3	8.11%
COND	5	2	5.41%
TAQUO	3	1	2.70%
SIDUR	3	-	-
CONC	2	-	-
PLACE	1	1	2.70%
SUBST	1	-	-
SIMIL	1	-	-

⁹ This does not include the relations covered by general c-glossemes since it is difficult to establish the exact scope of their meanings.

COCOND	1	-	-
CONTRA	1	-	-
COMPAR	1	-	-
TOTAL	85	37	100%
general c-glossemes	10		
TOTAL including general c-glossemes	95		

(Fig.5.15.) Summary of semantic overlaps of causality c-glossemes.

The convention of presentation is the same as for (Fig.4.15.) in chapter 4.

The links between the various relations listed in the table above and the reasons for the identicalness of their exponents can be better understood after analysing (Fig.5.16.) which depicts the scope of the semantically polyfunctional markers of causality as reconstructed from the data analysed.

The central position of PURPOSE in the network of semantic affinities of the relation of causality does not come as a surprise. As Thompson, Longacre and Hwang aptly point out:

Purpose and reason clauses can be seen as providing explanations, or accounts, for the occurrence of a given state of affair or action (...) They differ in that purpose clauses express a motivating event which must be unrealized at the time of the main event, while reason clauses express a motivating event which may be realized at the time of the main clause event. (2008:250-251)

Furthermore, as Dixon (2009:17) emphasizes, a purpose linking may be restated in terms of cause:¹⁰

(5.8a) *John took out a loan, **in order to** be able to buy a new car.*

(5.8b) ***Because** John took out a loan, he could buy a new car.*

Three of the 8 occurrences of PURPOSE in non-binary overlaps are to be found in the CAUSE-RESULT-PURPOSE configuration. These three relations are, as Dixon rightly notices (2009:17), strongly related to the notion of consequence. In the case of causality, it is the main clause that expresses consequence (5.9a). In the case of result

¹⁰ The example given by Dixon requires reorganization of the structure of the clause – in (5.8a) the clause linker introduces different SoA than in (5.8b). This operation is not, however, necessary to prove that purpose can be restated in terms of cause. In fact, wherever a clause which is a satellite of another clause expresses purpose, the two SoA can be paraphrased by a combination of a causal marker and modal verb. The reason for that is the very nature of the concept of purpose which embraces the idea of volition and obligation (not necessarily both at the same time). These two ideas can be expressed periphrastically by ‘because X wants’, ‘because X needs’ or ‘because X has to’ – cf. (5.8a) with the following sentence: *John took out a loan, because he wanted/needed/had to buy a new car.*

I do not elaborate further on this topic since it goes beyond the scope of this thesis into purely semantic analyses and philosophical discussions.

and purpose – the clauses marked by the resultative and purposive linkers (5.9b-c) the difference being that in purposive clauses the consequence is always deliberate.¹¹

(5.9a) *I went to the cinema, **because** I wanted to see a movie.*

(5.9b) *I wanted to see a movie, **so** I went to the cinema.*

(5.9c) *I went to the cinema, **in order to** see a movie.*

The polyfunctional marker *-te* in Au is an interesting example of the CAUSE-PURPOSE-RESULT polyfunctionality¹²:

(5.10a) Au (Scorza 1973:205)

Hir neiyim nan te hir nankip
they they.got.them they.came {CAUSE} they they.FUT.kill.him

‘They got them and came because they were going to kill him.’

(5.10b) (ibidem: 208-209)

Keiyik karehiiei wan kaknan te kakwep
he.got.it he.pulled.PL pond it.FUT.comes {PURPOSE} it.FUT.kills.us

‘He got the pond and pulled it in order that it would kill us.’

(5.10c) (ibidem: 208)

Wowirem te niu yapiriue katin kesiikeniik
she.threw.them {RESULT} sago.trees many it.grew it.everywhere

‘She threw them (bones and food scraps) away, and therefore many sago palms grew up everywhere.’

while Konso (5.11a-b) provides a clear example of the binary CAUSE-RESULT overlap:

(5.11a) Konso (Mous and Oda 2009:347)

Oorí sekkammaayé keltayta ishó yoytá
then afterwards baboon and wolf
gootáá kappin-oppá porá ann-áá-n maalá
about forest-at road go-IMPRF-PL {CAUSE}

‘After this Baboon and Wolf were thinking, because they were going on the Savannah.’

(5.11b) (ibidem:348)

Haá-rakkootá sédé in-gap-á ka,
INTERJ-problem this 1SG-have-IMPRF and

maalá takmoo-sé kanní
{RESULT} honey-DEM sell.CONT

‘I have this problem and therefore I am selling this honey.’

¹¹ In the case of RESULT the SoA expressing consequence may be either deliberate as in (5.9b), natural as in *It has been raining all night so the garden furniture is completely wet*, or unintended as *I broke my leg so I could not ice-skate any more that winter*. The same is true of causality.

¹² The examples are quoted as they are presented in the reference grammar – without exact glosses.

The second group of relations that often share the form of their exponent with causality marker is, as mentioned before, the temporal relation of anteriority (12 cases) and overlap (10 cases). The reason for the high frequency of these overlaps is the same: for two SoAs to be causally related it is necessary that one of them precedes the other one in time and, since both SIOVER and ANTE are concepts of temporal adjacency, the extension of meaning is self-evident.¹³ The same rule appears to be the motivation for reanalysis of TAQUO ('since') markers into markers of causality. Two of the four cases of TAQUO-CAUSE overlaps in the sample are binary (cf. English *since* in 5.12a which has an ambiguous causal/temporal reading and 5.12b where it is clearly causal). The two others involve also the relation of anteriority (cf. the Tamil *-tu* linker and its allomorphs in 5.13a-c). I have already emphasized that I believe that in the majority of cases (if not in all) it is the temporal markers that give rise to the causal ones and not the other way round.¹⁴

(5.12a) *I have not seen him **since** he moved out of town.*

(5.12b) *I have to go to London **since** I promised my friend to visit her this weekend.*

(5.13a) Tamil (Lehmann 1993:273)

Aintu naal kazi-ttu·k Kumar inkee va-nt-aan
 five day pass-**{ANTE}** Kumar here come-PST-3SG.M

'After five days passed, Kumar came here.'

(5.13b) (ibidem)

Kumar inkee va-ntu muunru varusam aay-ir-ru
 Kumar here come-**{TAQUO}** three year become-PST-3SG

'Since Kumar came here, three years have passed.'

(5.13c) (ibidem)

Mazai pey-tu payir nanr-aaka valar-nt-atu
 Rain fall-**{CAUSE}** crop goodness-ADV LZ grow-PST-3SG.N

'Because it rained, the crops grew well.'

The last temporal relation in the network depicted in (Fig.5.17.) is the relation of simultaneity duration (SIDUR, 'while') which occurs only two times: in Boko, where the *kε* marker is used as an exponent of the relations of causality, simultaneity overlap,

¹³ This affinity between SIOVER and ANTE has already been discussed in section 4.3.

¹⁴ Cf. also the development of English *since* discussed in section 3.2.4. A similar scenario emerges from the observation of the evolution of another English subordinator – *as*. Its first attested causal usage comes, according to OED, from around 1400. As a subordinator of time, manner and degree it was already in use in the early 13th century. Its first attested function was, however, that of an adverb of degree (around 1000) and of quality or manner (circa 1175).

temporal duration and purpose (Jones 1998:257-267), and in Santali where the c-glosseme *-te* which displays CAUSE-SIDUR polyfunctionality (5.14a-b) is used also as an exponent of the relation of anteriority (5.14c) and manner (5.14d):

(5.14a) Santali (Neukom 2001:189)

ʒkɛ-h̄ɔ̄ ba-ko badae-te sanam
 anyone-also NEG-3SG know-{\CAUSE} all
hɔ̄-ko apaj-kan-a
 person-3SG call.one.another-IMPV-IND

‘As nobody knew (how to do it), they all asked each other (to act).’

(5.14b) (ibidem:187)

Cala-k'-calak-te mit'-taj tayo-ko
 go-MID-RPD-{\SIDUR} one-CLF jackal-3SG

ɲɛl-tiok'ked-e-a
 see-reach-PST.ACTIV-3SG.OBJ-IND

‘While they were walking along, they caught sight of a jackal.’

(5.14c) (ibidem:188)

Nui iə sala-dɔ̄ oka-khɔ̄n cɔ̄-e
 this.ANIM PART scoundrel-TOP where-ABL ever-3SG.SBJ

odok-gɔ̄'-en-te iɲ bəhu-dɔ̄-e
 come-out-PST.MID-{\ANTE} my wife-TOP-3.SG.SBJ

ɔ̄r-ruər-ed-e-kan-a
 pull.return-IMPRF.ACT-3SG.OBJ-IMPV-IND

‘After the dirty scoundrel came out from who know where, he is pulling my wife back.’

(5.14d) (ibidem:189)

Khange-kin lahagɔ̄'-en-te-kin
 then-3DU.SBJ precede-PST.MID-{\MANNER}-3DU.SBJ

əyur-idi-ked-e-a
 lead.take-PST.ACT-3SG.OBJ-IND

‘Then the two (jackals) went ahead leading him (a leopard) along.’

Since the affinity between causality and temporal duration is less self-evident than between ANTE and CAUSE and SIOVER and CAUSE, and since the three temporal relations: ANTE, SIOVER and SIDUR are related to each other on their own rights, it is difficult to hypothesize about any pathway of grammaticalization. The polyfunctionalities may be an effect of polygrammaticalization or the meanings might be vague due the semantic underspecification that the linkers inherited from the sources they developed from. Similar possibilities exist for the explanation of the

origin of the CAUSE-COND–SIOVER overlaps found in the data collected.¹⁵ As Dixon notices:

In some languages ‘when’ and ‘if’ are marked in the same way and in some ‘when’ and ‘because’ are. These can go together – there is one marker which can be used for ‘when’, ‘after’, ‘because’ and ‘if’ in Warekana (...) and also in Jarawa. (2009:20)

On the other hand, however, we should take into account that a c-glosseme with CAUSE-COND functions might have gained its conditional meaning due to the causal overtones that often license conditionality (or the other way round). The first scenario seems to be an especially interesting field for more detailed research since, as Dancyngier (1993, 1998) has aptly noticed and as has been repeated in Dancyngier and Sweetser:

Causal readings enter the interpretation of content conditionals via conditionals’ primary function, *prediction*. Prediction in conditionals is a type of reasoning which consists in setting up a hypothetical (typically future) mental space and attempting to predict its consequences based on knowledge of typical cause-effect chains and general world-knowledge. (2009:122)

The three other c-glossemes serving the functions of exponents of both causality and conditionality that I found in the material gathered would fit this hypothetical scenario well. In Apache Jicarilla, for instance, the linking element *-go* (cf. Jung 2009:7-10) can be used to express the meaning of cause and conditionality as well as temporal simultaneity (SIOVER); in Au (Scorza 1973:203-2010) the *tu* c-glosseme is an exponent of causality, conditionality and purpose, and in Fur the word *asi* (Beaton 1968:165-166) is an example of a c-glosseme with a binary CAUSE-COND function.

MANNER is the final relation whose presence is noticeable in the network of affinities of causality. Being tangled up in a web of polyfunctionalities (which involve also PURPOSE, SIMIL, CONC, ANTE and SIDUR), MANNER remains an intriguing element of the jigsaw that I aim to solve here. Neither the encoding of the relation of manner nor its meaning has received so far serious linguistic interest, let alone special attention in any cross-linguistic study. It would be far beyond the scope of this thesis to attempt to fill in this gap and, thus, I restrict myself to presenting some general arguments that could (at least partially) account for the origin of MANNER-CAUSE, as well as some minor, yet related, overlaps.

Recently Dixon (2009:35) has proposed to distinguish between two types of MANNER. The first type, which has been labelled *real manner* (RMANNER henceforth), covers those relations where the action described by the focal clause is done in the manner described by the supporting clause as in:

¹⁵ In the literature analysed one example has been mentioned for Tamil (Lehmann 1993:36-37;276) and two for Pitjantjatjara (Eckert and Hudson 1994:264-270).

(5.15a) *He wears his hat as his father did.*

The second type – *hypothetical manner* (HMANNER henceforth) – refers to those cases where the supporting clause describes what the activity encoded in the focal clause pretends to be, or what it might be but is not:

(5.15b) *She was spending money as if she was a millionaire.*

or to cases where “the focal clause may depict a state, with the supporting clause then describing some imaginary event which might have given rise to the state” (Dixon 2009:36):

(5.15c) *He screamed as if he saw a ghost.*

It seems to me that it is in the hypothetical cases of manner, and more precisely in the element of consequence which hypothetical manner shares with causality, where the meanings of MANNER and CAUSE meet. We may explain it with the following example: the sentences in (5.15b-c) make sense only if we have the knowledge of certain more or less stereotypical rules governing human behaviour such as: ‘X is a millionaire and so X can spend lots of money not worrying about spending too much’ or ‘X sees a ghost so X screams’. The knowledge is nothing else than a set of cause-result scenarios. Obviously, it also applies to situations not involving human activity such as:

(5.15d) *The sky became so dark as if it was night.*

which, again, make sense because we are aware of a simple (cause-consequence) principle: ‘It becomes dark every single day because(/when) night comes’. Furthermore, it seems reasonable to conclude that various meanings related to RMANNER, such as similarity, comparison and contrast, may enter into the range of meanings covered by causality indirectly – by agency of MANNER (and so the overlaps may be viewed as epiphenomena). Since CAUSE itself is closely related to many other circumstantial concepts, the network of affinities becomes even more dense.

The MANNER-CAUSE polyfunctionality is exemplified by the Nivkh linker *-r*:

(5.16a) Nivkh (Gruzdeva 1998:54)

T'olf₂ kav-katn-gu-ř čnyř mařka-d
summer be.hot-INT-CAUS-{CAUSE} grass pour-FIN
‘Because the summer is very hot, [I] water grass.’

(5.16b) (ibidem:35)

Hajmnař t'axkyř n'aχ-kis nloŋbloŋ-d'i-kavr-r t'yr-d
old.man straight eye-INST blink-INT-NEG-{MANNER} look-FIN

‘The old man looked straight, not blinking eyes.’

In the discussion presented in this section I have analysed to a lesser or greater extent 12 of the 15 semantic overlaps that were reported in the material I have collected for the purpose of this study. The three remaining relations: concession, place and substitution are represented by single examples only and do not have any obvious semantic affinity to the relation of cause.

The conclusions that can be drawn from the analysis come down to the following list:

- a) the relations most closely related to causality are: purpose, simultaneity overlap, anteriority, result and manner;
- b) purpose, result and manner are linked to causality through the fact that they either involve or imply an idea of consequence;
- c) in those cases where causality overlaps with anteriority, simultaneity overlap (‘when’) and *terminus a quo* (‘since’), it is more likely that it is the temporal relations that the causality markers develop from than the other way round;
- d) the relations of simultaneity duration (‘while’) and conditionality in many cases become overlaps of causality linkers most likely due to polygrammaticalization which begun from temporal linkers (most commonly ‘after’ and ‘when’);
- e) relations such as similarity, contrast and comparison enter the network of affinities of causality indirectly – i.e. due to their links with the relation of manner which is related to causality on its own right;
- f) regarding the development of causality linkers out of other semantic types of c-glossems, the data strongly suggest three development pathways: SIOVER → CAUSE, ANTE → CAUSE and PURPOSE → CAUSE (all of them have been depicted in the diagram in Fig.5.9.);
- g) the collected material suggest also that causality markers may develop into purpose markers and conditionality markers.

We may now turn, as we did in chapter 4, to the question on how the findings of this study map onto the network of affinities established for the relation of causality in Kortmann (1997). In order to answer it let us look at the summaries presented in (Fig.5.17.) and (Fig.5.18.).

	I	II	III	IV	V	VI	VII
CAUSE (172 items)	SIOVER (26.7%)	ANTE (22.7%)	PURPOSE (22.7%)	COND (19.8%)	IMANTE (19.2%)	CONTIN (18.0%)	RESULT (17.4%)

(Fig.5.17.) The strongest semantic affinities of polyfunctional adverbial subordinators expressing causality according to Kortmann (1997:198)

(cf. explanation of conventions under the table in Fig.4.17. in chapter 4)

ABSOLUTE NUMBER OF OCCURRENCES (85 overlaps)					
I	II	III	IV	V	V
PURPOSE 30 items (35.29.1%)	ANTE 12 items (14.11%)	SIOVER 10 items (11.76%)	MANNER 9 items (10.59%)	RESULT 5 items (5.88 %)	COND 5 items (5.88 %)
OCCURRENCES IN BINARY OVERLAPS (37 overlaps)					
I	II	II	II	III	III
PURPOSE 22 items (59.45%)	ANTE 3 items (8.11%)	SIOVER 3 items (8.11%)	RESULT 3 items (8.11%)	MANNER 2 items (5.41%)	COND 2 items (5.41%)

(Fig.5.18.) The strongest semantic affinities of polyfunctional causality c-glossems revealed in this study

The most striking observation that emerges from the comparison concerns the frequency of the identified overlaps. In Kortmann’s work SIOVER and ANTE have been discovered to overlap with CAUSE as often as PURPOSE. In my study PURPOSE-CAUSE overlaps are by far the most frequent ones (both when it comes to the total number of overlaps and occurrences in binary overlaps) and SIOVER-CAUSE and ANTE-CAUSE polyfunctionalities do not exceed their number, even if counted together. This may suggest that the results of Kortmann’s study (which, as has been already said, is focused exclusively on the languages of Europe) reveal a geographic/genetic tendency for CAUSE markers to overlap more frequently in form with temporal markers of SIOVER and SIDUR than the world’s average emerging from my summaries.

COND and RESULT have been counted in both studies as important elements in the network of semantic affinities of CAUSE although in my study they are far less significant categories than in Kortmann’s monograph.

Comparison of (Fig.5.17.) and (Fig.5.18.) reveals also certain differences in the set of the most common overlaps: the relation of manner which occurs 9 times in my sample is absent from the list of the 7 most common semantic overlaps of causality in Kortmann’s study.¹⁶ On the other hand, CONTIN (contingency: ‘whenever’) and IMANTE (immediate anteriority: ‘as soon as’), which in Kortmann’s study are even more common than the CAUSE-RESULT overlaps, have not been reported in any of the languages I have looked at.

¹⁶ While discussing MANNER, however, the author has indicated that almost 1/3 of subordinators of manner have as their further reading CAUSE. The same applied to 20% of subordinators expressing the meaning of similarity.

5.4. SUMMARY

In this chapter I have presented the results of analysis of over 200 items which have been reported in the 84 languages included in the sample as capable of conveying the meaning of causality. The study revealed that over 66% of the causality c-glossemes are monomorphemic items. 77.8% of the linkers are classified as words and 19.7% as affixes.

The analysis has also shown that slightly over 40% of causality c-glossemes have polysemes in other syntactic categories (i.e. they are syntactically polyfunctional) and over 34% are used to express more than one circumstantial relation (i.e. they are semantically polyfunctional). Both semantically and syntactically polyfunctional causality linkers, although identified on the basis of exclusively synchronic observations, have served as a window into the origin and functioning of this group of markers. The identification of patterns of polysemy and analysis of material incorporated in the structure of those of the linkers that are polymorphemic revealed that causality markers find their sources most commonly in categories with meanings/functions related to the concepts of reason, origin (source), purpose and benefactiveness, location (and directionality), possessiveness, perception as well categories with temporal meanings/functions. This includes adpositions, case markers, nouns, verbs as well as complementizers, c-glossemes (of purpose and of temporal relations) and adverbs – a range far more diverse than the one presented by Heine and Kuteva (2002).

The analysis of patterns of semantic polyfunctionalities have confirmed close links between the relation of cause and purpose as well as between cause and relations of simultaneity overlap, anteriority, *terminus a quo* and conditionality, all of which have been reported in the literature (Thompson and Longacre 1985, Dixon 2009). Moreover, the affinity between causality and purpose has been found to be far more salient than we could conclude from Kortmann (1997). The study has also revealed that the c-glossemes that convey the meaning of causality are often used as exponents of the relation of manner and a variety of meanings related to manner: similarity, comparison, contrast. Initial attempts at explanations of these overlaps have been added to the analysis.

Finally, the *complexity* → *monofunctionality tendency* has been observed for both aspects of the analysis (i.e. the more morphologically complex a marker is the less likely it is to be used in a variety of syntactic functions and to express more than one circumstantial relation).

CHAPTER 6

Purpose

6.1. MORPHOLOGICAL COMPLEXITY AND FORMS OF C-GLOSSEMES

The database of c-glossemes of purpose on which this chapter is based consists of 165 items. For 12 of those linkers it was not possible to obtain information on their morphological complexity and so they have been excluded from the summary in (Fig.6.1.). Monomorphemic linkers, as can be seen from the table, constitute just over 65% of all the purpose markers. Bimorphemic c-glossemes make up almost a quarter and the remaining ones contribute 16 items.

	count	%
monomorphemic	100	65.36%
2 morphemes	37	24.18%
3 morphemes	12	7.84%
3+ morphemes	4	2.62%
TOTAL	153	100%

(Fig.6.1.) Morphological complexity of purpose c-glossemes

As for the form of the markers, purpose c-glossemes are in 60% of cases free words, in 32.3% affixes and in as 6.4% discontinuous markers. Among affixes we find 45 suffixes and 4 prefixes. The latter ones come from just two languages: Krongo (see example 3.2. in chapter 3 and 6.1a-b below) and Hualapai. While in Krongo the three prefixal c-glossemes are used almost exclusively for expressing purposive relations, in Hualapai they have been classified as general c-glossemes.

(6.1a) Krongo (Reh 1985:351)

m-áa caw ð-múnó-ŋ éekwàarà
 CL.F-COP INF.go {PURPOSE}-INF.call-TR chief
 ‘She is going in order to call the chief’

(6.1b) (ibidem)

n-éediyá *áǎǎj k-óofù-ŋ* *kí-tí*
 1/2-IPFV.FR.come I {PURPOSE}-INF.sleep LOC-it

‘I come often in order to sleep on it (a tree).’

The discontinuous markers occur 10 times in the sample. In some cases one of the elements is a suffix, and the other is a word (cf. example 3.6. in chapter 3 from Yanyuwa where the first element – *nyala* – is optional and 6.2. from Suppyire where both morphemes are obligatory). In others, the markers consist of two or more words (cf. the English example in 6.3.),

(6.2.) Suppyire (Carlson 1994:588)

U a kàrè lwǎhɔ tá-cya-ge *e*
 she PRF go water {PURPOSE1}-seek-G.SG {PURPOSE2}

‘She went to fetch water.’

(6.3.) *In order for* him *to* win, he has to get at least nine votes.

Combinations of free words and affixes occur only twice in the sample and clitics are completely absent.

	count		%
free word	95		60,51%
affix	50		31.85%
	suffix	prefix	
combination	2		1.27%
discontinuous marker	10		6.37%
clitic	-		-
TOTAL	157		100%

(Fig.6.2.) Forms of purpose c-glossemes¹

6.2. INSIGHT INTO ORIGINS

6.2.1. Syntactic polyfunctionality and patterns of polysemy

The information on syntactic mono/polyfunctionality of purposive c-glossemes, on which one may draw quite reliable conclusions regarding the most common grammaticalization pathways of the markers, is available for 151 of the 165 linkers in the sample. Initial details on distribution of the markers in particular categories of morphological complexity and syntactic polyfunctionality are depicted in (Fig.6.3.).

¹ Information on the form of the c-glossemes has been confirmed for 157 items and so not all of the 165 items are considered in the table.

Purpose, interestingly, is the first of the analysed relations where markers are more often syntactically polyfunctional than monofunctional. At the same time the number of monomorphemic purpose linkers is twice as high as that of the polymorphemic ones. As for both the relations described in the earlier sections, also in the case of purpose the polymorphemic markers are more likely to be monofunctional syntactically.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	29	19.21%	69	45.70%	98
polymorphemic	42	27.81%	11	7.28%	53
TOTAL	71	47.02%	80	52.98%	151

(Fig.6.3.) Distribution of c-glossems of purpose according to their syntactic mono/polyfunctionality and morphological complexity

More specifically (cf. Fig.6.4.), almost 80% of the polymorphemic markers are monofunctional. Among the 11 c-glossems that have polysems/homonyms in other syntactic categories bimorphemic linkers contribute eight items and the trimorphemic ones – three. Once again we observe that the increase in morphological complexity of the linkers is correlated with the decrease in both the total frequency of the markers and the number of polysems/homonyms found among them. The most numerous group are the bimorphemic linkers which constitute over half of all the polymorphemic markers.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
2 morphemes	29	54.7%	8	15.0%	37
3 morphemes	9	17.0%	3	5.7%	12
3+ morphemes	4	7.6%	-	-	4
TOTAL	42	79.3%	11	20.7%	53

(Fig.6.4.) Distribution of syntactic mono- and polyfunctionality in polymorphic purpose c-glossems

The *complexity* → *monofunctionality tendency* is also to be observed for the degree of syntactic polyfunctionality. The polymorphemic markers, as shown in (Fig.6.5.), have been found to have at most one polyseme. At the same time almost one third of the monomorphemic markers (22 items) share their form with two or more categories.

	1 additional function	2 additional functions	3 additional functions	4 additional functions	TOTAL
monomorphemic	47	16	3	3	69
polymorphemic	11	-	-	-	11

(Fig.6.5.) Degree of syntactic polyfunctionality of purpose c-glossemes

The analysis of the 80 syntactically polyfunctional markers reveals that the purpose c-glossemes overlap in form with a variety of syntactic, grammatical (TAM markers) and morphological categories (nominal and verbal suffixes). The most frequent group of overlaps, as depicted in (Fig.6.6.) below, is without any doubts the group of adpositions which occur 34 times and in 21 cases are the only overlaps of purpose linkers. The second group, when it comes to the total number of occurrences, are complementizers.² They are, however, overtaken by case markers, as for their occurrences in binary overlaps. Apart from these three groups, the presence of adverbs, verbs, and nouns is also quite noticeable in both the first and the second column of the table. The group of TAM markers is as numerous as that of nouns but when it comes to binary overlaps the first one is less frequent. Hence, the first 6 groups: adpositions, case markers, complementizers, verbs, adverbs and nouns are most likely to be recognized as the most common sources of purpose markers.

		I occurrences	II occurrences in binary overlaps	III occurrences in multiple overlaps
1.	ADPOSITION	34	21	13
2.	COMPLEMENTIZER	19	6	13
3.	CASE	18	16	2
4.	ADVERB	7	3	4
5.	VERB	7	4	3
6.	NOUN	6	5	1
7.	TAM MARKER	6	2	4
8.	MODALITY SUFFIXES AND PARTICLES	5	2	3
9.	CONJUNCTION	3	2	1
10.	RELATIVE CLAUSE	3	-	3
11.	NOUN SUFFIX	1	1	-
12.	COPULA	1	-	1
13.	INFINITIVE MARKER	1	-	1
14.	ADJECTIVE	1	-	1
15.	PARTICIPLE MARKER	1	-	1
	TOTAL	113	62	51

² This group includes several items that have been categorized as markers of direct speech.

(Fig.6.6.) Syntactic overlaps of purpose c-glossemes

As in the case of anteriority and causality here, too, the grammaticalization heuristics (cf. section 3.2.9.) are to be applied. It is, therefore, necessary to analyse in detail the specific functions/meanings of the categories listed in the table above and add to it information on semantic mono/polyfunctionality of the polysemous purpose linkers. All the necessary details are gathered in (Fig.6.7.)

As shown in the table, among binary overlaps of the ADP-LINKER_{purpose} type the adpositions with benefactive/purpose/'for' meanings contribute 11 items, adpositions of reason 3 and adpositions of reason/purpose/'for' 2. Noticeably, while in the first group 1/3 of the markers are semantically polyfunctional, the same is true about 100% of the other two. In all of these cases (cf. Fig.6.8.) the other circumstantial meaning expressed by these markers is that of CAUSE.³

³ The possible scenarios of the emergence of these overlaps have been discussed in detail in section 5.2.1.

POLYSEMOUS/ HOMONYMOUS CATEGORY	CHARACTERISTICS OF POLYFUNCTIONALITY	
ADPOSITION Total:34	BINARY OVERLAPS:21 ADPreason:3 (3) ADPbenefactive;purpose;'for':11 (4) ADPreason;purpose;'for':2 (2) ADP'for'(temporal):1 ADPdirectional:1 ADP:3 (1)	MULTIPLE OVERLAPS:13 ADP + COMP:2 ADPinstrumental;directional;'for' + COMP:1 ADP'of';'for' + COMP:1 ADP + COMP + other:3 ADPbenefactive+ COMP + MOODObligation:1 ADPreason; benefactive; directional + ADVcomparative + COMP:1(1) ADP'to';'for' + COMP + ADJ'full' + VERB'cause';'give';'make':1 (1) ADP + VERB (+ other):2 ADPreason + VERB'think':1 (1) ADPpurpose + VERB'go' + ASPprospective:1 ADP + other:6 ADPreason;'for' + NOUN'source';'cause':1(1) ADPdirectional + INF:1 ADP'against' + OBL:1 (1) ADP'about' + CASEdative:1 ADPlocative + VERBSUFadhortative:1 ADP'for' + VERBSUFintentional:1
COMPLEMENTIZER Total:19	BINARY OVERLAPS:6 (3)	MULTIPLE OVERLAPS:13 COMP + REL (+ other):3 COMP + REL:2 COMP + REL + PTCP:1(1) COMP + ADP (+ other):5 COMP + ADPinstrumental;directional;'for':1 COMP + ADP'of';'for':1 COMP + ADPbenefactive + MOODObligation:1 COMP + ADPreason; benefactive; directional + ADVcomparative:1(1) COMP + ADP'to';'for' + ADJ'full' + VERB'cause';'give';'make':1 (1)

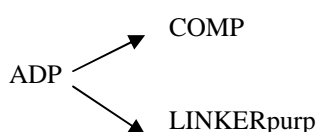
		<p>COMP + other: 5 COMP + VERB'benefit','deserve', 'be equal to','resemble':1 COMP + OBLQ:1 (1) COMP + CASEtranslative:1 COMP + FUT + PASThistoric + MOODexhortative:1 COMP + VERBSUFintentional:1</p>
<p>CASE Total:18</p>	<p>BINARY OVERLAPS:16 CASEdative:7 (3) CASElocative:2 CASEadessive:1 (1) CASEallative:2 (1) CASEinstrumental:1 (1) CASEinelative:1 CASEgenitive:1 (1) CASEoblique:1</p>	<p>MULTIPLE OVERLAPS:2 CASEtranslative + COMP:1 CASEdative + ADP'about':1</p>
<p>TAM MARKER Total:6</p>	<p>BINARY OVERLAPS:2 MOODvolitional/potential:1 ASPECTinchaotive:1</p>	<p>MULTIPLE OVERLAPS:4 MOODpurpose + FUT + MOODadhortative:1 MOODexhortative + COMP + FUT + PASThistoric:1 MOODObligation + ADPbenefactive + COMP:1 ASPprospective + VERBgo + ADPpurpose:1</p>
<p>VERB Total:7</p>	<p>BINARY OVERLAPS:3 VERB'think of','stick to':1 VERB'being in the state of':1 VERB'be so','be true':1 (1)</p>	<p>MULTIPLE OVERLAPS:4 VERB'cause','give','make' + ADP'to','for' + COMP + ADJ'full':1 (1) VERB'benefit','deserve', 'be equal to','resemble' + COMP:1 VERB'think' + ADPreason:1 (1) VERB'go' + ADPpurpose + ASPprospective:1</p>

ADVERB Total:7	BINARY OVERLAPS:4 ADVequative:1 ADVdegree/manner:1 (1) ADV'only then':1 ADV:1	MULTIPLE OVERLAPS:3 ADVcomparative + ADPreason; benefactive; directional + COMP:1 (1) ADV'then' + CONJ'but':1(1) ADP + COP:1
NOUN Total:6	BINARY OVERLAPS:5 NOUN'manner';'way':2 NOUN'intention';'purpose':1 NOUN'similarity':1 NOUN'manner';'likeness';'resemblance':1(1)	MULTIPLE OVERLAPS:1 NOUN'source';'cause' + ADPreason;'for':1(1)
VERBAL SUFFIXES AND MODALITY MARKERS Total:5	BINARY OVERLAPS:2 VERBSUFintentional:1 OPTATIVE(particle):1	MULTIPLE OVERLAPS:3 VERBSUFintentional + ADP'for':1 VERBSUFintentional + COMP:1 VERBSUFadhortative + ADPlocative:1
CONJUNCTION Total:3	BINARY OVERLAPS:2 CONJ'but':1 (1) CONJ'or':1 (1)	MULTIPLE OVERLAPS:1 (1) CONJ'but' + ADV'then'
RELATIVE CLAUSE MARKER Total:3	BINARY OVERLAPS:0	MULTIPLE OVERLAPS:3 REL + COMP: 2 REL + COMP + PTCP:1 (1)

(Fig.6.7.) Details of syntactic overlaps of purpose c-glossemes

The conventions of data presentation are the same as described for (Fig.4.7.) in chapter 4.

The adpositional meaning of purpose, benefactiveness and reason occur also relatively frequently in multiple overlaps (last column of Fig.6.7.). In this group, 5 of the adpositions that express the meaning of purpose and/or benefit overlap with complementizers. The pathway of grammaticalization of complementizers into purpose linkers has already been mentioned in section 3.2.8. I treat the problem in more detail later on in this section and so here I wish to remark only that, according to the unidirectionality hypothesis, we might expect that the ADP-COMP-LINKER_{purpose} overlap came about either through grammaticalization of an adposition separately into a complementizer and purpose marker (polygrammaticalization)⁴:



or through the following grammaticalization chain:

$$\text{ADP} \rightarrow \text{COMP} \rightarrow \text{LINKER}_{\text{purpose}}$$

The analysis of the collected material provides evidence also for other complex scenarios of polygrammaticalization involving purpose linkers. Two of the scenarios – suggested by the configuration of overlaps of the Dagur c-glosseme *tuale* (NOUN ‘source’; ‘cause’-ADPreason; ‘for’-LINKER) and Eipo marker *tennen* (VERB ‘think’-ADPreason-LINKER) – have been discussed in section 6.2.1. Four others include linkers with multiple overlaps including verbs and verbal suffixes. The first to be discussed here is the Akan *ma/ama* word (cf. Fig.6.8.) which, in addition to being a marker of circumstantial relations of purpose and result, has been reported to act also as a verb ‘cause’/‘give’/‘make’, adposition ‘to’; ‘for’, adjective ‘full’ and complementizer.

purpose linker	syntactic overlap	semantic overlap
Sango <i>ngbangati</i>	ADPreason	CAUSE
Hausa <i>sòbo dà/sòbodà</i>	ADPreason	CAUSE
Hausa <i>dòmin/don</i>	ADPreason	CAUSE
Japanese <i>tame</i>	ADPreason;purpose	CAUSE
Dagur <i>tuale</i>	ADPreason;‘for’ + NOUN ‘source’; ‘cause’	CAUSE, CONC
Eipo <i>tennen</i>	ADPreason + VERB ‘think’	CAUSE
Kanuri <i>-ro</i>	ADPbenefactive;reason;directionality + ADVcomparative + COMP	CAUSE, COCOND, SUBSTI
Eipo <i>ate</i>	ADPreason;‘for’	CAUSE

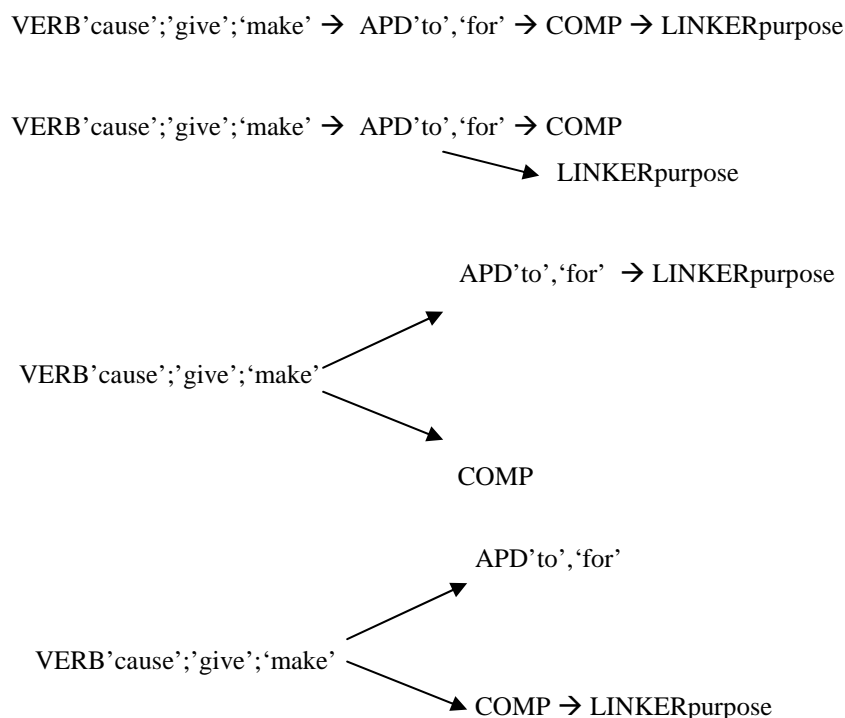
⁴ For discussion on the development of complementizers out of adpositions see Heine and Kuteva 2007:92-93.

Arabic San'ani <i>ʿala sibb</i>	ADP'for'	CAUSE
Arabic San'ani <i>ʿasibb</i>	ADP'for'	CAUSE
Sango <i>tɛnɛti</i>	ADP'for'	CAUSE
English <i>for</i>	ADPbenefactive;purpose	CAUSE
Ndyuka <i>fu/fi</i>	ADP'of';'for' + COMP	COND
Nisga'a <i>ʔa</i>	ADP	SIOVER, SIDUR
Gola <i>wee</i>	ADP'against' + OBL	ANTE, SIOVER, CAUSE
English <i>so</i>	ADVdegree/manner	RESULT
Au <i>-te</i>	ADV'then' + CONJ'but'	RESULT, CAUSE, COND
Akan <i>ma/ama</i>	APD'to','for' + VERB'cause';'give'; 'make' + ADJ'full' + COMP	RESULT
Ket <i>-dir̄ta/-dita</i>	CASEadessive	CAUSE
Basque <i>-ra</i>	CASEallative	COND
Konso <i>é</i>	CASEdative	CAUSE
Burushaski <i>-ar</i>	CASEdative	ANTE
Galo <i>bə̀</i>	CASEdative	MANNER
Krongo <i>má-</i>	CASEgenetive	CAUSE
Dagur <i>-eer/-aar/-ier/-oor</i>	CASEinstrumental	PRERER, SUBSTI
Estonian <i>et</i>	COMP	general c-glosseme
Warlpiri <i>yunga/yinga/yunu</i>	COMP	CAUSE
Galo <i>ə̀mlà(a)</i>	COMP	CAUSE
Paiwan <i>tu</i>	COMP + OBL	CAUSE, MANNER
SE Tepehuan <i>na</i>	COMP + REL + PTCP	general c-glosseme
Yami <i>ta</i>	CONJor	CAUSE
Kayah Li <i>ma</i>	VERB'be so'; 'be true'	general c-glosseme
Japanese <i>noni</i>	CONJ'but'	CONC
Khwe <i>-kò</i>	ASPcontinuative;inchoative;completive	general c-glosseme
Japanese <i>yoo</i>	NOUN'manner';'likeness'; 'resemblance'	COMPAR
Lezgian <i>-wal</i>	NOUNSUFabstract	MANNER

(Fig.6.8.) Syntactically and semantically polyfunctional purpose markers

It is interesting that in Akan (cf. Fig.6.8. and discussion in Balmer and Grant 1942:169-171) the concept of reason ('to cause'), benefactiveness ('to give') and result ('to make') surfaces already at the level of verb. At the adpositional level the purposive/benefactive sense is also clearly present. The complementizer might have developed from either the verb or the adposition. As for the presence of adjective in this overlap, we cannot be sure whether it is accidental or not. What is quite clear, however, is that the verb must have been the first node in this grammaticalization chain (cf. unidirectionality hypothesis discussed in section 3.2.3.) It might have developed into ADP and COMP separately or first into ADP and from there into a complementizer. The function of linking circumstantial clauses may have entered the picture either by the reanalysis of the adposition or the complementizer. If we assume, following the unidirectionality hypothesis that there was a pathway of development which started

from the reanalysis of verb and lead through the development of adpositions and/or complementizers to the clause linkers it still gives us four possibilities as for the further stages, including scenarios of polygrammaticalization⁵:



A polygrammaticalization scenario, as has already been said in 3.2.5, emerges also from the analysis of the Rama morpheme *-bang* which, in addition to being a purpose linker, serves also the function of adposition of purpose, verb 'go', prospective aspect and imperative marker. It is plausible that the process of polygrammaticalization is responsible also for two other overlaps involving adpositions: ADPlocative-VERB_SUFadhortative-LINKERpurpose polyfunctionality of the Lepcha *ká* c-glosseme (Plaisier 2006:125-126) and ADP'for'-VERBSUFintentional-LINKERpurpose overlap of the Santali *lagit'/lagat'* marker (Ghosh 2008). The intentional and adhortative meanings developed most likely from verbs on a verb-to-affix cline as discussed by Hopper and Traugott (2003:111).⁶ It is also possible that the same verbs gave rise to the adpositions of location and purpose/benefit respectively. Whether the c-glosseme developed from the adpositions or from the verbal suffixes is, however, not clear.

With the Lepcha example we have entered the topic of adpositional functions other than purpose/benefit/reason which overlap with linkers of purpose. Apart from

⁵ As Heine and Kuteva notice: "The primary source of prepositions and postpositions is provided by nouns (...). But verbs, as well, are a common source of adpositions" (2007:71). See also further discussion *ibidem* on verbs giving rise to adpositions (pages 71-73) and complementizers (pages 76-77) as well as on the development of complementizers out of adpositions (pages 92-93).

⁶ Note that both the intentional and adhortative functions are often fulfilled by mood markers. However, in the two cases evoked here the two suffixes have been categorized as verbal, derivational suffixes and have been treated separately from modal categories.

the two single adposition with the meaning of ‘about’ and ‘to’/‘for’ there are four occurrences of directional adpositions – one item in the column discussing binary overlaps, and three items on the list of multiple overlaps. Since the concept of directionality is clearly related to that of goal and benefit/recipient but is less abstract than the latter one (cf. less abstract → more abstract development assumption), the reconstruction of the grammaticalization path should probably look as follows:

ADP_{directionality} (→ ADP‘for’) → LINKER_{purpose}

Summarizing the discussion so far we can then conclude that the adpositions that are most common sources of purpose markers are those expressing the meaning of purpose/benefit, reason and directionality. It has to be remembered, however, that the ADP_{reason}-LINKER_{purpose} overlap may be epiphenomena of the LINKER_{causality}→LINKER_{purpose} development (see also section 5.2.1.).

The second most numerous group of overlaps of purpose markers, comprises, as listed in (Fig.6.6.), complementizers.⁷ One of the most general cross-linguistic findings relevant for the topic analysed here is that “both relative and complement clauses can develop into adverbial clauses while a development in the opposite direction is unlikely to happen” (Heine and Kuteva 2007:252). The data I collected provide additional evidence to an already rich body of works mentioning complementizers giving rise to markers of purpose clauses (see section 3.2.8.). Among the 19 COMP-LINKER_{purpose} overlaps listed in (Fig.6.6.) 6 are binary overlaps. Two of them express also the circumstantial meaning of causality (cf. Fig.6.8.) and so we may put forward a hypothesis that the grammaticalization chain looked as follows:

COMP → LINKER_{purpose} → LINKER_{causality}

The situation becomes more complicated in the case of multiple overlaps involving complementizers. The majority of the categories accompanying complementizers in (Fig.6.7.) (especially adpositions, verbs and relativizers) are likely sources of clause linkers on their own but it is equally plausible that they developed into complementizers which in turn became reanalysed as markers of circumstantial clauses.

The third category that the linkers of purpose most often overlap with are case markers. Aikhenvald, in the paper on polyfunctionality of case marker which has been already referred to on numerous occasions so far, notices that “dative or purposive marking on a noun phrase tends to have a purposive meaning when used as a clause linker” (2008:594). Indeed, in my study dative case markers have been found to

⁷ The distinction between a complementizer and an adverbial modifier (subordinator, linker) is not always easy. I choose to follow here the rule of thumb: if a clause is used as an obligatory argument, it is analysed as a complement clause; if it is an optional modifier of the verb in the main clause, it is a circumstantial clause.

contribute almost half of all the cases of binary CASE-LINKER_{purpose} overlaps. In three cases these grammatical overlaps are accompanied by semantic overlaps – the dative case markers are also used as exponents of the relation of anteriority in Burushaski, manner in Galo, and cause in Konso (cf. Fig.6.8.). The second significant contribution of case markers as potential sources of purpose c-glossemes comes from those that express concepts related to location and/or movement, such as locatives, allatives, adessives and inelatives. This phenomena has also been reported by Aikhenvald (see section 3.2.8.).

The fourth group of the most common syntactic overlaps listed in (Fig.6.6.) is the broad category of TAM markers.⁸ The grammaticalization of TAM markers into markers of interclausal relations has not been discussed in cross-linguistic studies on grammaticalization yet. What we do know from typological studies is that aspect markers develop into markers of tense rather than the other way round (see the insightful study by Bybee et al. 1994). Hengeveld, in his recent paper (forthcoming), reveals also that mood markers can develop both from tense markers and aspectual categories, while Bybee and Dahl (1989) as well as Palmer (1986:216-218), indicate that future tense may develop *inter alia* from verbs with the modal content of intention or volition, modality of obligation.

Some of the evidence collected in the course of my research seems to suggest that over time certain TAM markers associated with one proposition (on the basis of their salient semantic properties and through their use in a context of another SoA) become reanalysed as exponents of circumstantial relations. In (Fig.6.7.) five mood classes (volitional, adhortative, exhortative and obligation) as well as two tense classes (future and historical past), and two aspectual classes (inchoative and prospective) have been mentioned. Almost all of these categories mark the activity/state as yet unrealized (future). All the mood markers are, moreover, clearly related to the concept of goal. I believe that these two properties: future reference and underlying meaning of purpose are the basis on which the TAM markers may develop to mark SoA as being a deliberate consequence of another SoA. A good example of such development is the Polish purpose subordinator *by*, which is identical to a volitional/potential mood marker⁹:

(6.4a) Polish

Kasia *posz-t-a-by* *na spacer*
 Kate go-PST-3SG.F-VOLIT for walk.SG.ACC

‘Kate would (eagerly) go for a walk.’

⁸ It is not my intention here to enter the discussion on the distinctions between these three categories and within them – a topic that has been covered by entire volumes (see Comrie 1976 and 1985, Palmer 1986, Dahl 1985 and 2000, Dahl and Bybee 1989, Bybee et al. 1994 and for additional references also Timberlake 2007:332-333) and still is not free of controversy. Here I focus on the meanings and functions of particular categories rather than on the labels which I simply repeat after the authors of grammars.

⁹ For more detailed discussion see Pisarkowa 1984:158-159.

(6.4b) Polish

Tomek zosta-ł z dzieć-mi,
Tom stay-PST.3.M with child-PL.INS

by Kasia mog-ł-a pójść na spacer
{PURPOSE} Kate can-PST-3.SG.F go.INF for walk.ACC

‘Tom stayed with children so that Kate could go for a walk.’

The Vitu marker *kata/koto/kutu/kete/kiti* also provides an interesting example of reanalysis of functions. It may be used to “indicate a future event, usually with deliberate or volitional overtones” (van den Berg and Bachet 2006:115) as in

(6.5a) Vitu (van den Berg and Bachet 2006:116)

Hau kata vano kara ruma kuari
1.SG FUT go to house that

‘I am going to that house.’

in adhortative function:

(6.5b) (ibidem:117)

To kata hani-a!
1DU.INCL ADHORT eat-3SG

‘Let’s eat him!’

and finally, as a purpose linker:

(6.5c) (ibidem)

Ia e varidinga-ni-au kata mai
3SG REAL.3SG force-TR-1SG {PURPOSE} come

‘She forced me to come.’

In the case of multiple overlaps it is, obviously, not easy to indicate with certainty what the exact grammaticalization pathways was, but if we were to use typological arguments as presented in Bybee and Dahl (1989) and Palmer (1986:216-218), we could suggest that the purpose linker developed from the reanalysis of the use of *kata* as a future marker. As for the other examples of multiple overlaps involving TAM markers which have been listed in (Fig.6.7), it is not possible to reconstruct the grammaticalization pathways since what we are dealing with is a complex set of TAM categories and categories which have been attested as common sources of c-glossemes: complementizers and adpositions.

We may refer the ideas concerning TAM markers presented here to another group of purpose overlaps: verbal suffixes of modality and modal particles. It is reasonable to assume that the linkers developed out of these modality morphemes along similar lines as hypothesized for the TAM markers. The examples from Retuarã exemplify the modal and clause linking functions of the *-ẽrã* suffix:

(6.6a) Retuarã (Strom, personal communication)

Parua ki-ba.a-ẽrã baa-yu
 banana 3SG.M-eat-INTEN do-PRS

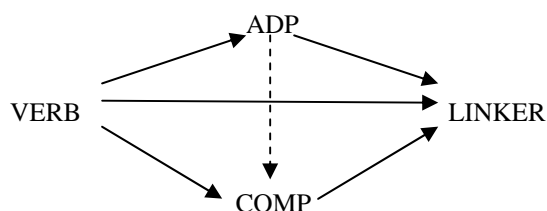
‘He is about to eat the banana.’

(6.6b) (Strom 1992:170)

ki-re dã-wapahĩ-re.ã kopereka ki-ta.a-ẽrã
 3SG.M-TERM 3PL-pay-PST 3SG.M-guard-**{PURPOSE}**

‘They paid him to guard the door.’

The role of verbs (especially of the verb ‘say’) in the grammaticalization of clause linkers of purpose has often been emphasized in the literature on grammaticalization.¹⁰ In the data set analysed here we find, as depicted in (Fig.6.7.), the verb of motion ‘go’, a verb with the stative meaning ‘being in a state of’, two instances of the verb ‘think’, as well as verbs with other meanings cognitively related to goal/purpose: ‘benefit’, ‘give’, ‘cause’ (the verb ‘say’ does not occur as a polyseme of any of the markers, but, as shown in section 6.2.2. it is a common building block of polymorphemic markers). It is reasonable to conclude that in all these cases it was these verbs that triggered grammaticalization processes. The purpose linkers might have developed either directly from the verbs or through the agency of adpositions and/or complementizers. The diagram below illustrates all the possibilities:



Moreover, in some cases (such as the VERB‘think’-ADPreason-LINKER overlap) the meaning of purpose might have arisen from an extension of the meaning of a causality marker as it has been discussed in chapter 5.

The next group of polysemes/homonyms mentioned in (Fig.6.6.) and described in detail in (Fig.6.7.) is the group of adverbs. It is formed by two items with temporal meaning and three labelled ‘equative’, ‘manner/degree’ and ‘comparison’ respectively. The temporal adverbs together with some of the TAM markers described above appear to form one group of sources of purpose markers: temporal concepts (cf. the list of sources of purpose linkers presented in Schmidtke-Bode 2009 quoted in section 3.2.8.) The concepts expressed by the other three adverbs also constitute, as scrutinized in

¹⁰ Cf. discussion in section 3.2.8.

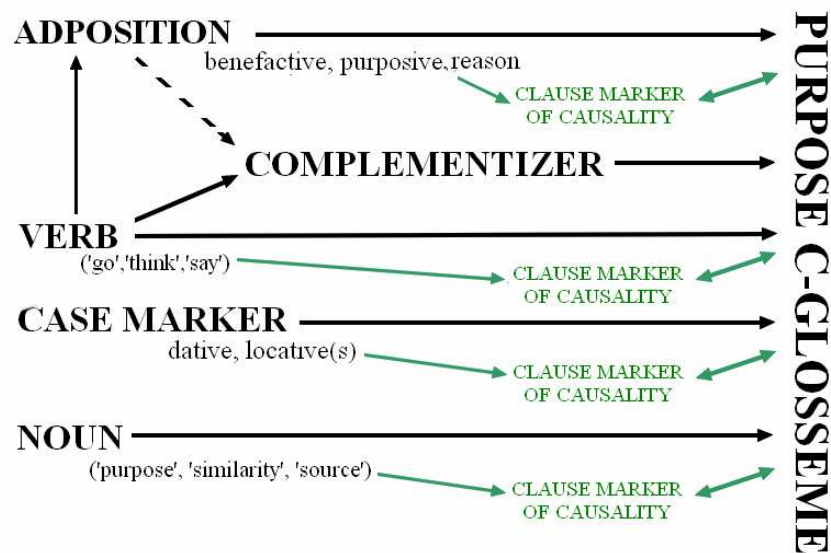
section 6.3, a part of a bigger phenomenon. On the whole, however, it seems that adverbs are rather uncommon sources of purpose linkers.

The last syntactic category that definitely should not be left unmentioned here is nouns. They occur in the sample 6 times and possess a variety of meanings: ‘manner’, ‘way’, ‘intention’, ‘purpose’, ‘similarity’, ‘likeliness’, ‘resemblance’, ‘source’ and ‘cause’. This very interesting and compact set provides us also with examples of affinity between the concept of purpose and manner, and purpose and cause. In 5 of the 6 overlaps there is no synchronic evidence for categories acting as mediators between nouns and purpose linkers. The only example that suggests such mediation is the Dagur *tuale* linker which is polysemous/homonymous with the noun ‘source’/‘cause’ and the adposition reason/‘for’. As pointed out in section 6.2.1. this marker may act as an exponent of the relation of purpose and causality. Although it is quite clear that the order of grammaticalization was

NOUN → ADP → LINKER

it is not obvious whether both the semantic functions of causality and purpose developed from the adposition or whether one of them gave rise to the other one.

The picture of grammaticalization pathways emerging from the analysis here is quite complex. On the one hand it is clear that the categories that most often were reanalysed as markers of purpose are: adpositions, complementizers, case markers, verbs and nouns. On the other hand, in many cases we are not able to determine which of the categories was the immediate predecessor of the linker. The main reason for that is that all these categories exhibit “rich grammaticalization behaviour” (Heine and Kuteva 2007:87) and consequently, even having supported the argumentation presented here with typological evidence reported in other studies, in many cases it is still difficult to predict the order of emergence of particular meanings/functions. Hence, in the diagram below I endeavour to depict the most common sources of grammaticalization of purpose markers as emerging from the analysis comprising the broader grammaticalization perspective at the same time.



(Fig.6.9.) Reconstruction of the most common sources of purpose markers and pathways of grammaticalization suggested by the material analysed

The dotted lines mark a pathway which is only vaguely suggested by the material collected but has been discussed in more detail in other studies. The possibility that some of the purpose markers might have developed directly from causality markers is marked by green arrows, and the possibility that purpose linkers came to mark the relation of causality – by double-sided arrows. Since there was no clear trend for specific verbal or nominal meanings to occur more frequently than others, in these two categories the meanings encountered in the dataset are put in brackets.

6.2.2. Polymorphemic markers and their internal structure

As reported in section 6.1., among the 84 languages in the sample we find 53 polymorphemic markers including phrasal c-glossems (6.7.) and inflected verbs (6.8).

(6.7.) Polish

Zadzwoń-ł na policj-ę w cel-u
 call-PST.3M for police-ACC {in goal-ABL→PURPOSE}

zgłosz-eni-a kradzież-y
 report-NMLZ-GEN theft- GEN.SG

‘He called police in order to report the theft.’

(6.8.) Lezgian (Haspelmath 1993:393)

Wun masa-ur.u-laj usal ta-xu-j
 you.ABS other-SBST.PL-SREL poor NEG-be-OPT

luhuz, za wiri
 {say-DAT/INFCONV/IMPRFCONV→PURPOSE} I.ERG all

zehmet-ar wun patal č’ugwa-zwa
 work-PL you.ABS for pull-IMPRF

‘I am doing all the work for you so that you may not be poorer than others.’ (lit. ‘...saying: May you not be poorer than others’)

The majority of the polymorphemic markers are, however, non-phrasal c-glossemes. Some of them have quite complex internal structure incorporating more than 3 morphemes. One of such complex fossilized c-glossemes has been found in Supyire. In that Mali language there exists the discontinues *bà...mé* marker used in clauses of comparison. It appears also in purpose clauses but, as Carlson explains:

to function as purpose clauses they [*bà...mé* clauses – A.M.] must have a distinctive internal structure consisting of an initial potential clause with the verb *pyi* ‘do’, ‘be’, followed by a same subject subjunctive clause (...). The entire two clauses structure is bracketed by subordinators *bà* and *mé*.’ (1994:587)

The use of *bà...mé* in a comparison clause is exemplified in (6.9a) and in a purpose clause in (6.9b)¹¹

(6.9a) Supyire (Carlson 1994:569)

Bà pi sanmpítí jye mé yì gú m-pyì àmuni
like they OTHERS.DEF be like you.PL POT FUT-be thus

‘Like the others are, you would be like that.’

(6.9b) (ibidem:587-588)

Pi ba wyīge tūrù bà pi gú m-pyì
they PROG hole.DEF dig.IMPRF {PURPOSE1} they POT FUT-{PURPOSE2}

sí lwɔhɔ ta mé
SBJV water get {PURPOSE3}

‘They are digging the hole in order to get water.’

The internal structure of all the 53 polymorphemic markers has been depicted schematically in (Fig.6.10.). It is clear from the table that adpositions are the most common group of incorporated elements. As in the analysis of polysemes presented in the previous section, we note here, too, that the frequency is highest for benefactive/purpose and locative adpositions, which are followed by those encoding directionality and possessiveness. Interestingly, only 6 of the 18 adpositions occur with nouns. Hindi, Yanyuwa and Supyire provide us with examples of combinations of adpositions and case markers but only for the first language the c-glosseme is

¹¹ The history behind the origin of this c-glosseme becomes even more interesting when we look at the atomic meanings, as Carlson explains: “the subordinating morphemes *bà* and *mé* which enclose the entire clause are odd to say the least. They seem to be identical in form to the negative identifier *bà* ‘it is not’ and the clause final negative marker *mé*” (1994: 568). The following example illustrates such a usage of these two markers:

Mu wú bà mé
your POSS it.is.not NEG
‘It isn’t yours.’

continuous for the other two the structures are discontinuous (discontinuity is marked by italicizing the language name).

Nouns occur in the table 13 times: 5 of them are nouns ‘reason’ and ‘matter’. As we might expect all of the c-glossemes incorporating these nouns are used also as exponents of the relation of causality. This fact can be used to support the assumption that the purposive meanings of these linkers developed from their original causal meanings. Other nouns listed in (Fig.6.10.) are ‘order’ (occurring three times in English), ‘place’ (in Didinga), ‘side’ (in Lezgian), ‘habit’/‘custom’/‘practice’ (in Hausa) and two whose semantic affinity with purpose is especially interesting: the already mentioned in (Fig.6.8.) Japanese ‘manner’/‘likeliness’/‘resemblance’ noun and the Polish word for ‘goal’.

It has been said in the previous section that verb ‘say’/‘tell’, which is often mentioned in the context of grammaticalization of purpose markers, has not occurred as a polyseme of any of the c-glossemes in the sample. From (Fig.6.10.) we can see, however, that with 8 occurrences this verb clearly is the most common building block of polymorphemic purpose linkers.¹² Notably, the verb occurring in the Sango *tEnɛti* linker is polysemous with the noun ‘word’/‘speech’ and complementizer.¹³ The chain of polysemes is itself an interesting example of grammaticalization where the phonological form of the original source has been preserved despite the reanalysis of its function/meaning. Other verbs that have been reported to be incorporated in the structure of polymorphemic markers include ‘see’ (in Leti), ‘do’/‘be’ (in Supyire) and ‘reach’/‘arrive’ in Khwe. The last one is especially telling example that there exists an affinity between the concept of purpose and that of a motion towards a clearly defined destination.

Meanings/functions related to the concept of purpose are also to be found among the case markers, and TAM and modality markers listed in the table. This includes: dative and locative markers, morphemes with volitional, intentional, optative and potential meanings. Demonstrative and complementizers also occur on several occasions as material incorporated in the polymorphemic purpose markers.¹⁴ Adverbs, on the other hand, are less common – 3 of the 5 occurrences have been contributed by the English markers *so* and *as*. The remaining two (coming from Meyah and Southeastern Tepehuan) are temporal adverbs with the meaning of ‘then’.

Finally, in 4 languages: Boko, Japanese, Konso and Supyire, we encounter c-glossemes incorporating markers which may themselves act as a c-glosseme. In

¹² Haspelmath explains how the verb ‘say’ in its converbal form became reanalysed as purpose marker in Lezgian: ‘the purpose meaning arises in the following way: “A does B, saying: May C do D!” gives rise to “A does B in order for C to do D”’ (Haspelmath 1993:393). The functioning of the idiomatic form is exemplified in (6.8.) above.

¹³ This fact was pointed out to me by Christina Thornell.

¹⁴ This includes the English polysemous word *that*, which serves both the functions.

Japanese and Konso, these morphemes, if used on their own, would be semantically polyfunctional. When they form a part of a polymorphemic c-glosseme, the complex linker unambiguously encodes the relation of purpose.

The analysis of patterns of polysemy in the previous sections has been concluded with a list of categories that appear to be the most common sources of purpose markers. The same categories: adpositions, nouns, verbs, TAM and modality markers as well as case markers and complementizers, reappear frequently also in the material analysed in this section.¹⁵ The particular meanings/concepts of these categories also reoccur. We may therefore conclude that purpose markers arise from a variety of sources whose conceptual content is organized around the ideas of goal, purpose/benefit, reason, intention, location, directionality and movement towards a goal.

In comparison to the set presented in Heine and Kuteva (2007, see section 3.2.8.) my list of sources of purpose linkers is more diverse and the pathways reconstructed are, on many occasions, more complex. The outcome of the analysis has also certain advantages over the list of “developmental trajectories of purpose clauses” given in Schmidtke-Bode (2009) and summarized in section 3.2.8. Firstly, it identifies the specific syntactic categories from which the markers develop in the world’s languages. Secondly, it presents a more exhaustive list. And thirdly, it gives us an idea about the origin of clause linkers themselves, not mixing the issue with the development of purposive constructions in general.

¹⁵ Note also, that in both the analysis of polysemy and the incorporated material adverbs turned out to be rather a marginal category.

	ADP					ADP+other	ADV		NOUN					NOUN + VERB + COMP 'say'/'word'/'speech'	VERB		TAM AND MODALITY MARKERS	CASE	POSSparticle/marker	DEM	COMP	REL	C-GLOSSEME (c=causality, p=ppurpose,g=general)	OTHER	NOT KNOWN		
	'for'/benefactive/purpose	locative	directionality	possessive	other		degree/manner/quantity	'then'	'order'	'reason'	'matter'	other	'say'/'talk'		other												
Arabic (2)					•																						•
Baure (3)																										•••	
Basque (2)																		•								•	
Boko (2)																								• (c.)			
Didinga (2)													•										•				
English (3)		•				+			•																		
English (2)							•																				•
English (3)		•							•																		•
English (4)	•	•				+			•																		
English (3)						+	••																				
Estonian (2)																											
Galo (2)																											
Galo (2)																											+
Hausa (2)													•														
Hatam (2)																											
Hatam (4)						•							• ('talk')														•
Hatam (3)						•							• ('talk')														••
Hatam (3)	•												• ('talk')														•
Hindi (3)	••																										
Japanese (2)																											+
Japanese (2)																											+
Japanese (2)						+																					+
Japanese (2)												+															+
Ket (2)																											•
Ket (3)																											•
Khwe (2)												•															
Khwe (2)																											•+
Konso (2)																											
Lezgian (2)																											•
Lezgian (2)																											
Lezgian (2)													• ('say')														
Lezgian (2)													• ('say')														

1. Category ADP column *other*: ADP without a meaning reported for Arabic; ADPreason/purpose for Japanese; ADPsource for Meyah; ADD'after'/locative/purpose for Polish.
2. Category ADP + OTHER: ADPdirectionality and INF for all the English markers; ADP and NMLZ for both Hatam markers.
3. Category NOUN column *other*: NOUN'place' for Didinga; NOUN'habit'/'practice'/'custom' for Hausa; NOUN'manner'/'likeness'/'resemblance' for Japanese; NOUN'side' for Lezgian; NOUN'goal' for Polish.
4. Category VERB column *other*: VERB'arrive'/'reach' for Khwe; VERB'see' for Leti; VERB'do'/'be' for Supyire.
5. Category TAM AND MODALITY: optative particle for Ket; VERBSUFintentional + COMP for the first Khwe linker, and VERBSUFintentional + COMP and ASPcontinous/completive/inchoative for the second one; aorist suffix or aorist converb/infinitive for Lezgian; MOODvolitional for all 4 Polish markers; FUT and AUXpotential for Supyire.
6. Category OTHER: ABS,EMP,ATTR for Baure; NMLZ for Basque; ANA for the first Hatam marker, ANA and INS for the second one, ANA for the third one and NMLZ for the last one; ADV + COP for all Japanese markers; CASEtranslative + COMP for the first Ket marker and INANIM for the second one; SEQ for Leti; CASEdative + INF + ADVLZ for Lezgian, FOC, DEF for Lillooet; ABS for Maale; COP and NMLZ + REL for Mandarin; NMLZ for Nivkh; CONJ'and';'but' for both Polish markers; two negators for Supyire; COMP + PTCP for Tamil; COMP + REL for SE Tepehuan.

6.3. SEMANTIC POLYFUNCTIONALITY AND COGNITIVE AFFINITY

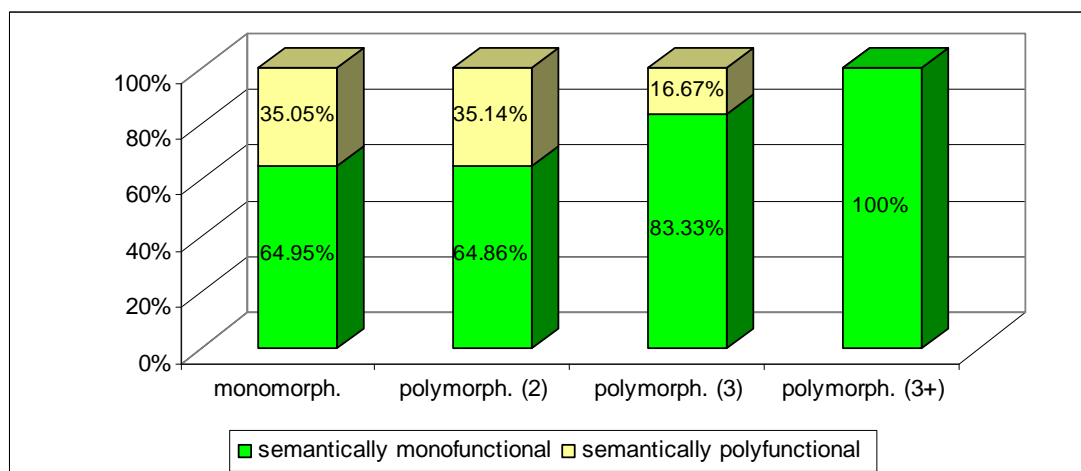
Among the 150 purpose c-glossems which have been given codes for their semantic mono/polyfunctionality in the database, 49 are used to express more than one circumstantial relation. The majority of them, as evident from (Fig.6.11.), are monomorphemic linkers. The bimorphemic semantically polyfunctional items are almost three times less frequent and the trimorphemic markers exhibit the polyfunctionality only in 2 of 12 cases.

	semantically monofunctional		semantically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	63	42.00%	34	22.67%	97
2 morphemes	24	16.00%	13	8.66%	37
3 morphemes	10	6.67%	2	1.33%	12
3+ morphemes	4	2.67%	-	-	4
TOTAL	101	67.34%	49	32.67%	150

(Fig.6.11.) Distribution of c-glossems of purpose according to their semantic mono/polyfunctionality and morphological complexity

Again, the data show that the complexity → monofunctionality tendency holds both for syntactic (see section 6.2.1.) and semantic polyfunctionality of c-glossems.

As depicted in (Fig.6.12.), the ratio of monofunctional and polyfunctional markers in monomorphemic and bimorphemic c-glossems is almost identical. The share of polymorphemic linkers, however, decreases significantly for the trimorphemic markers and is completely absent from the group of markers made of 4 and more markers.



(Fig.6.12.) Ratio of semantically monofunctional and polyfunctional markers in purpose c-glossems with different morphological complexity

When we move to the analysis of degree of polyfunctionality (Fig.6.13.), however, one element may surprise us – the fact that there is a sudden increase in the degree of polyfunctionality in the middle part of the table. The c-glossemes with a scope over 4 relations constitute over 22% of all the polyfunctional markers while those with a scope over three relations are significantly less common.

	scope over 2 relations	scope over 3 relations	scope over 4 relations	scope over 5+ relations	TOTAL
count	27	3	11	8	49
%	55.10%	6.12%	22.45%	16.33%	100%

(Fig.6.13.) Degree of semantic polyfunctionality of purpose c-glossemes

10 of these 11 markers with scope over 4 relations are monomorphemic and the remaining one is bimorphemic as the detailed table below (Fig.6.14.) shows. In fact, among the markers that have scope over purpose and more than one other relation, this is the only example of a polymorphemic marker used as an exponent of a variety of circumstantial glossemes. All the general c-glossemes (markers with scope over 5 relations) are, unsurprisingly, monomorphemic. Thus, although the numbers do not decrease gradually for the monomorphemic linkers, when we put together the data presented in (Fig.6.11.) and (Fig.6.14.), it is clear that the complexity → monofunctionality tendency remains valid for polymorphemic markers of purpose too.

	scope over 2 relations		scope over 3 relations		scope over 4 relations		scope over 5+ relations	
	count	%	count	%	count	%	count	%
monomorphemic	13	26.53%	3	6.12%	10	20.41%	8	16.33%
2 morphemes	12	24.49%	-	-	1	2.04%	-	-
3 morphemes	2	4.08%	-	-	-	-	-	-

(Fig.6.14.) Distribution of c-glossemes of purpose according to their degree of polyfunctionality and morphological complexity

In the materials analysed, the purpose markers have been found to share their form with a total of 14 other circumstantial relations. The rank table (Fig.6.15.) begins with the relation of causality which overlaps with purpose 30 times. It contributes more than half of all the overlaps and almost 65% of the binary overlaps. The following three relations – result, manner and conditionality are far less frequent occurring only 5 to 3 times. Moreover, even taken together they comprise only one third of the binary overlaps that causality contributes. Finally, we need to emphasize that there are also 8

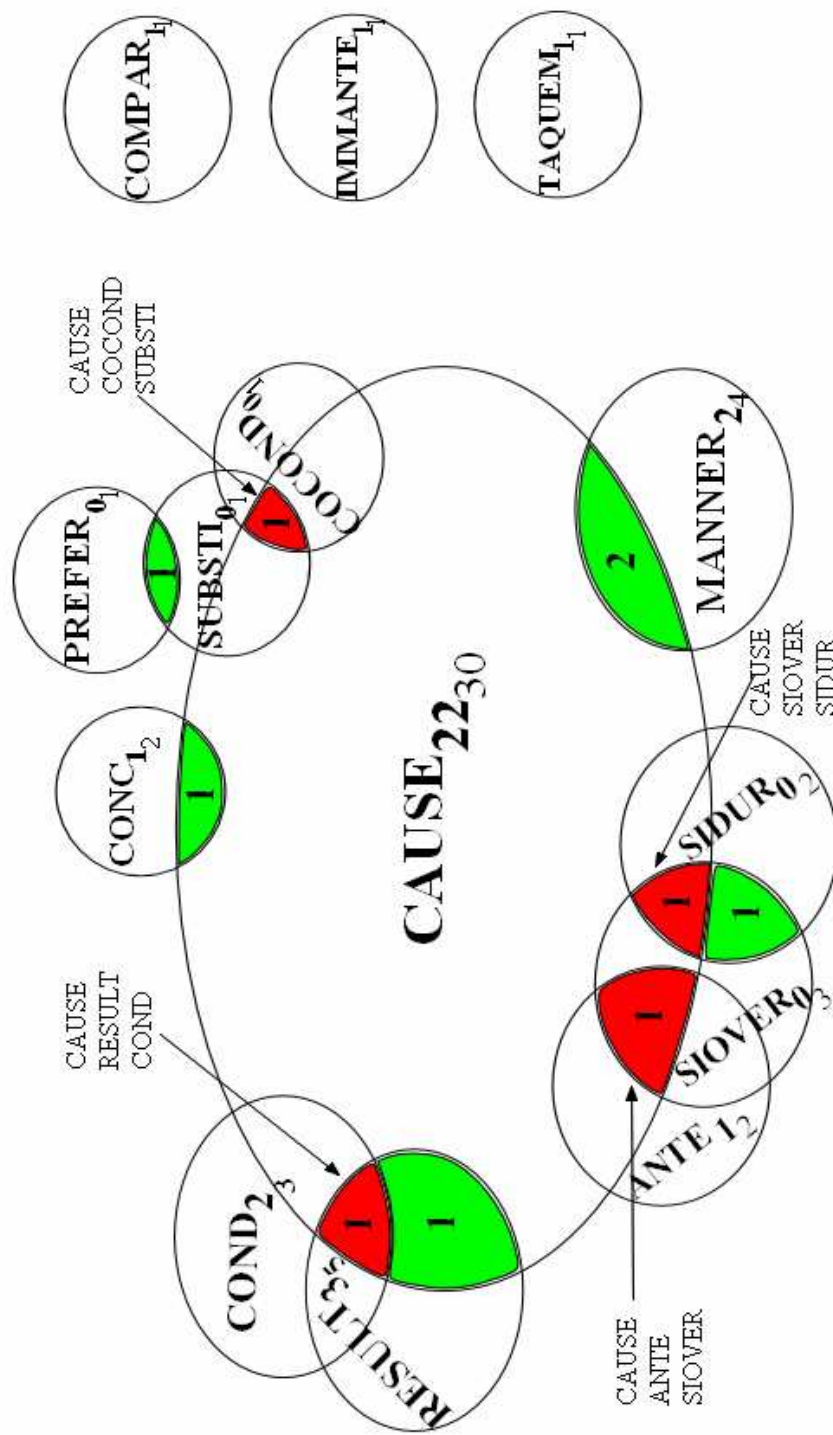
general c-glossemes reported to have scope over the relation of purpose. They have been added at the bottom of the list.

The fact that the last 10 relations in the table above (from SIOVER downwards) contribute on average only 1.6 overlaps, and that none of them occurs in binary overlaps with purpose more than once is a good enough reason to assume that they are only remotely related to purpose. For this reason those which occur at the end of the list are not to be considered a significant element of the network which is reconstructed in (Fig.6.16.).

	occurrences of overlaps	occurrences in binary overlaps	
		count	percentage of the total number of binary overlaps (34)
CAUSE	30	22	64.71%
RESULT	5	3	8.83%
MANNER	4	2	5.88%
COND	3	2	5.88%
SIOVER	3	-	-
ANTE	2	1	2.94%
CONC	2	1	2.94%
SIDUR	2	-	-
SUBSTI	2	-	-
COMPAR	1	1	2.94%
IMMANTE	1	1	2.94%
TAQUEM	1	1	2.94%
COCOND	1	-	-
PREFER	1	-	-
TOTAL	58	34	100%
general c-glossemes	8		
TOTAL including general c-glossemes	66		

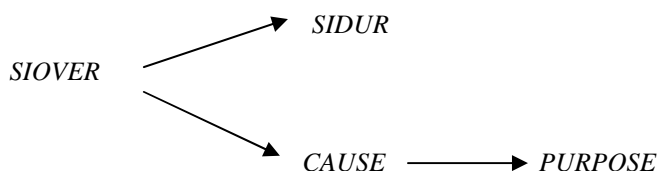
(Fig.6.15.) Summary of semantic overlaps of causality c-glossemes

The convention of data presentation are the same as for (Fig.4.15.) in chapter 4.



(Fig .6.16.) Semantic polyfunctionality of purpose c-glossemes
 The conventions of presentations are the same as for (Fig.4.16.) in chapter 4.

One of the thoughts that come to mind immediately when we look at (Fig.6.16.) is that the network of semantic affinities of purpose markers appears to be far less complicated than that of anteriority and causality. Besides a relatively low number of overlaps, almost all of the circumstantial meanings are organized around the concept of cause.¹⁶ The fact that many of them do not occur in binary overlaps with purpose strongly suggests that they might have entered the semantic space of purpose markers indirectly. This is most probably the explanation for the presence of temporal relations of anteriority and simultaneity overlap (SIOVER) in the diagram. Both of them have been discussed in detail in section 6.3. as cognitively very close to causality and it is most likely through the overlap with causality that they entered the semantic space of purpose. I believe that the relation of simultaneity duration (SIDUR) also entered the space indirectly – through the affinity with the relation of temporal overlap (for this point see section 4.3.). This is to say that the presence of SIDUR-PURPOSE overlap can be explained by the affinity chain where SIOVER is related to both SIDUR and CAUSE but PURPOSE is cognitively close only to the second one¹⁷:



The second cluster of overlaps involves relations of cause, result and conditionality. The affinity between the first two and purpose has been discussed in the previous chapter. It has been said (see section 5.3.) that it is the more general concept of consequence that brings these two relations together. The same concept licenses the presence of conditionality in the RESULT-CAUSE-COND-PURPOSE cluster, as well as in binary PURPOSE-COND overlaps. Let us look at the following two sentences:

(6.10a) *If you want to open the door you have to press the button.*

(6.10b) *In order to open the door you have to press the button.*

The SoA expressed in the apodosis in (6.10a) is understood as a condition for the SoA in the protasis to occur and the event of the doors opening is to be seen as a consequence. Through the use of the modal verb of volition (‘will’) the conditional clause can be read exactly as purpose clause – as expressing deliberate consequence. The same is true about negated clauses:

¹⁶ The relationships of cause and purpose and the motivation behind many of the overlaps depicted in (Fig.6.16.) have already been scrutinized in sections 4.3. and 5.3.

¹⁷ The affinity chain is not the same as the grammaticalization chain. I do not hypothesize here about any specific chronological order of emergence of particular meanings/functions depicted in the chain.

(6.11a) *If you don't want to be late, better take a taxi.*

(6.11b) *In order not to be late, better take a taxi.*

The *fu/fi/efi/ofu* marker from Ndyuka is an example of a polyfunctional c-glosseme capable of expressing the relations of both purpose and conditionality:

(6.12a) Ndyuka (Huttar and Huttar 1994:115)

Mi o soi en wan moi sani
1SG FUT show 3SG.OBL INDEF nice thing

fu a sa fika mi
{PURPOSE} 3SG IRR leave 1SG
'I will show her a fine thing so that she would leave me alone.'

(6.12b) (ibidem:119)

Efu mi tyai en te doo a boto, mi o siki
{COND} 3SG carry 3.SG.OBL until arrive INDEF boat 1SG FUT sick

'If I carry it all the way to the boat, I'll be sick.'

An example of an Au polyfunctional CAUSE-PURPOSE-RESULT marker has been given in the previous chapter (examples 5.10a-c)

The overlaps of PURPOSE with MANNER are quite intriguing elements of the jigsaw puzzle. This topic has not been, to my knowledge, scrutinized in any linguistic study yet. I shall attempt to propose a hypothesis which would account for at least some of the overlaps.

Certain, apparently unconnected, findings mentioned on several occasions in the discussion so far seem to suggest that MANNER, as well as concepts related to it (such as similarity and comparison), may be directly linked with purpose.¹⁸ It has been found (see Fig.6.8.) that among polysemes of purpose markers we find a noun with the meaning of 'manner'/'likeness'/'resemblance' in Japanese, noun 'similarity' in Cubeo, comparative adverb in Kanuri, equative adverb in Maale and adverb of degree/manner in English as well as verb 'be equal to'/'resemble'/'benefit'/'deserve' in Akan. Many of these items, as well as the English degree/manner/quantity adverb *so* reoccur also in the material incorporated in the structure of polymorphemic markers. Finally, in the map of semantic affinities in (Fig.6.16.) we also notice that the relation of comparison, known to be related to manner (cf. Kortmann 1997), is one of the binary

¹⁸ In reference grammars of many languages the meanings of manner, means, similarity, comparison etc. are discussed under the general header manner. Since this thesis is focused predominantly on analysis of the linkers of anteriority, causality, purpose and conditionality I shall not engage in revising the classifications. The only consequence of this choice is that the label MANNER should be understood here as a general term denoting a group of closely related meanings of manner/means/similarity and not as one homogenous concept.

semantic overlaps of purpose. This gives us 8 pieces of evidence which come from various parts of the world and which support the claim of affinity between purpose and manner/similarity/comparison.¹⁹ There seem to be just one explanation of the conceptual process that led to the emergence of these overlaps: the point where manner and purpose meet is in these cases where the manner in which an action is performed is expected to lead directly to the intended result. Let us, once again, consider some English examples:

(6.13a) *She was screaming **so that** he could hear her.*

(6.13b) *She was screaming **in such a way that** he could hear her.*

In (6.13a) the meaning of purpose is encoded explicitly without giving information about the manner or volume of the scream. The use of the subjunctive verb form triggers the purpose reading of (6.13b) as well but in this example it is additionally explicitly stated (note the polymorphemic manner c-glosseme *in such a way that*) that there is some special manner in which the activity is performed.²⁰ It is from these kinds of usages that the PURPOSE-MANNER polyfunctionality of clause linkers most probably emerges.

Of course, in the cases of multiple PURPOSE-MANNER-CAUSE overlaps (see Fig.6.16.) it is possible that the PURPOSE-MANNER overlap emerged as epiphenomena of the fact that both manner and purpose are closely related to causality (see section 5.3.). An example of such polyfunctionality for which it will never be possible to determine the exact grammaticalization pathway, is the Nivkh marker *-r* (cf. examples 6.13 and 5.16a-b)

(6.13.) Nivkh (Gruzdeva 1998:52)

Čax kuryjo-ɣavr-gu-iny-ř n'in mu lylu-d-ɣun
 water purl-NEG-CAUS-MOD-{PURPOSE} we boat caulk-FIN-PL

'We caulked the boat in order that the water does not purl.'

We may also suspect that the relations of similarity and comparison have entered the semantic network of purpose as indirect nodes – through their affinity with the relation of manner. The data collected do not provide evidence for such claims but it must be remembered that they are based only on synchronic observations and thus some elements of the grammaticalization chains which have undergone phonological

¹⁹ Güldemann (2001) mentions in his paper that similitive verbs (like 'be equal to', 'resemble') are one of the sources of complementizers. According to the author the process typically leads from similitive, through quotative to complement markers but the nature of this pathway is not entirely clear. In the light of the fact that complementizers are common sources of purpose markers it would be fruitful to look at a wider choice of languages seeking for evidence for more complex pathways involving these various categories.

²⁰ The manner is left underspecified but in a given communicative context, such as big physical distance between the two persons, it will be perfectly clear that the woman screamed loudly.

changes, are not captured. More evidence and more detailed studies would be needed to verify the hypothesis and to explain the origin of the less frequent affinities depicted in (Fig.6.16.).

The main conclusions emerging from the analysis of semantic overlaps of purpose c-glossemes presented here can be summarized as follows:

- a) the network of semantic affinities is organized in three main clusters: causality-temporal relations (ANTE, SIDUR, SIOVER); causality-conditionality-result and causality-manner;
- b) among these relations causality is cognitively most closely related to purpose and it is this relation that the majority of the overlaps are gathered around;
- c) the temporal relations are rather insignificant group in the network and their presence is most likely of epiphenomenal character;
- d) the affinity between purpose and the relations forming the biggest cluster: RESULT, COND and CAUSE are best explained by reference to the general concept of consequence in which all four concepts are encapsulated;
- e) the relation of manner seems to enter the semantic space of purpose through those usages where one action is deliberately performed in such a way that it leads to the emergence of the second one;
- f) among clause linkers the only group that may be surely viewed as cross-linguistically common source of purpose markers are c-glossemes of causality, although it seems also that in some cases it is the purpose linkers that gave rise to causality markers (both these directions have been depicted in Fig.6.9.).

The findings concerning the strongest semantic affinities presented here resemble those reported by Kortmann (1997) when it comes to the top of the ranking – both studies have found that PURPOSE overlaps most often with the relation of CAUSE and RESULT (cf. Fig.6.17. and Fig.6.18.) although there are significant statistical differences between the two studies. My results suggest that CAUSE as an overlap of PURPOSE, is 6 times more frequent than RESULT in the absolute number of occurrences and over 7 times more frequent in binary overlaps, while in Kortmann’s study it is RESULT that occupies the leading position.

	I	II	III	IV	V
PURPOSE (80 items)	RESULT (52.5%)	CAUSE (46.3%)	COND (21.3%)	SIOVER (13.8%)	SIMIL (13.8%)

(Fig.6.17.) The strongest semantic affinities of polyfunctional adverbial subordinators expressing purpose according to Kortmann (1997:198)

ABSOLUTE NUMBER OF OCCURRENCES (58 overlaps)				
I	II	III	IV	IV
CAUSE 30 items (51.72%)	RESULT 5 items (8.62%)	MANNER 4 items (6.90%)	COND 3 items (5.17%)	SIOVER 3 items (5.17%)
OCCURRENCES IN BINARY OVERLAPS (34 overlaps)				
I	II	III	III	
CAUSE 22 items (64.71%)	RESULT 3 items (8.83%)	MANNER 2 items (5.88%)	COND 2 items (5.88%)	

(Fig.6.18.) The strongest semantic affinities of polyfunctional c-glossems of purpose revealed in this study

In Kortmann's work conditionality has been reported to be the third most common overlap of purpose leaving behind the concepts of simultaneity duration (SIOVER) and similarity (SIMIL). My analysis has revealed that the frequency of overlaps with COND is almost identical to that of RESULT, MANNER and SIOVER (the last relation does not occur as a binary overlap of PURPOSE). The PURPOSE-MANNER overlap has not been mentioned by Kortmann at all which may simply mean that in the author's database MANNER and PURPOSE overlap with each other in less than 10% the cases (the author set the 10% threshold as a criterion for a relation to enter the table). In my sample manner contributes only 7% of all the overlaps but, in fact, neither of the other relations, except causality, goes beyond 10%.

6.4. SUMMARY

The analysis of the 165 purpose markers has provided us with several interesting findings. Firstly, it has been revealed that over 60% of the markers are free words and almost 1/3 are affixes. Secondly, it has been found that in both the semantic and syntactic domains the complexity → monofunctionality tendency holds. The third group of interesting conclusions has come from the analysis of patterns of polysemy which has provided us with numerous hypotheses as to the pathways of grammaticalization. The hypotheses, considered in the light of findings of the typological studies mentioned in chapter 3, have revealed a rather complex picture of

the sources and channels of development of purpose linkers. Along with categories such as nouns, verbs, adpositions, case markers and complementizers, more unusual ones – such as TAM markers and modality markers – have been depicted in the diagram summarizing the results of reconstructions. In general, the analysis has also revealed that purpose markers arise from a variety of sources whose conceptual content is organized around the ideas of goal, purpose/benefit, reason/cause (including c-glossemes of causality), intention, location, directionality and movement towards a goal. Finally, it has been shown that the network of semantic affinities of purpose is relatively neatly structured and uncomplicated with the dominant position of causality and significant contribution from relations such as result, conditionality and manner. It has been emphasized that the concept of consequence embraces all the major relations overlapping with purpose and, consequently, it should be viewed as the most likely explanation for the origin of many of these overlaps. This apparent simplicity of the network of affinities of purpose markers stands in direct contrast to the complicated picture of the sources of purpose c-glossemes. On the whole the data discussed here prove that the relation of purpose and its linguistic exponents are really an intriguing topic which would deserve more attention and further, detailed studies.

CHAPTER 7

Conditionality

7.1. MORPHOLOGICAL COMPLEXITY AND FORMS OF C-GLOSSEMES

In the analysed sample 171 conditionality c-glossemes have been identified and included in the database. For 8 of them it was not established whether they were morphologically complex or simple. Among the remaining 163, as presented in (Fig.7.1.), almost 67% consist of just one morpheme. Bimorphemic markers contribute 19% and the trimorphemic ones over 10%.

The majority (over 63%) of the conditionality linkers, as shown in (Fig.7.2.), are free words (for examples see 3.1.,7.1b, 7.5a etc.). Affixes constitute almost a quarter of all the 171 items and, unsurprisingly again, we find that it is suffixes that are far more frequent than either prefixes or infixes (see examples 1.30. from Apache Jicarilla and 3.10a. from Mantauran Rukai). There are also three cases of clitics (see example 3.7. from Yindjibarndi) but what seems most interesting is the relatively high number of distributed conditionality linkers (see examples 3.5. and 3.8. from Khwe). They occur in the sample as many as 20 times constituting almost 12% of all the markers.

	count	%
monomorphemic	109	66.87%
polymorphemic	31	19.02%
polymorphemic	17	10.43%
polymorphemic	6	3.68%
TOTAL	163	100%

(Fig.7.1.) Morphological complexity of conditionality c-glossemes

	count			%
free word	108			63.16%
affix	40			23.39%
	suffix 36	prefix 3	infix 1	
combination	-			-
distributed	20			11.70%
clitic	3			1.75%
TOTAL	171			100%

(Fig.7.2.) Morphological complexity of conditional c-glossemes.

7.2. INSIGHT INTO ORIGINS

7.2.1. Syntactic polyfunctionality and patterns of polysemy

For 160 of the 171 markers it was possible to elicit information on both morphological complexity and syntactic mono/polyfunctionality. As (Fig.7.3.) shows, over 70% of the conditionality c-glossemes are syntactically monofunctional but it is the remaining 46 polyfunctional linkers that are the subject of detailed analysis in the following section.

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	65	40.62%	41	25.63%	106
polymorphemic	49	30.63%	5	3.12%	54
TOTAL	114	71.26%	46	28.75%	160

(Fig.7.3.) Distribution of c-glossemes of conditionality according to their syntactic mono/polyfunctionality and morphological complexity

Before we move to the core analysis, let us first look more closely at the distribution of the c-glossemes in the particular groups of markers determined on the basis of their internal complexity and syntactic polyfunctionality.

As it is clear from (Fig.7.3.), among the polyfunctional c-glossemes 41 are monomorphemic and 5 polymorphemic. The discrepancy clearly supports the existence of the complexity → monofunctionality tendency between these two groups. However, unlike in the case of the other three circumstantial relations analysed in this thesis, it is not possible to observe the tendency within the group of polymorphemic markers

themselves. The reason for that is the small total number of items in this group. Among the 5 polymorphemic and syntactically polyfunctional linkers two are bimorphemic, two are trimorphemic and one consists of more than three morphemes:

	syntactically monofunctional		syntactically polyfunctional		TOTAL
	count	%	count	%	
2 morphemes	29	53.70%	2	3.70%	31
3 morphemes	15	27.78	2	3.70%	17
3+ morphemes	5	9.27%	1	1.85%	6
TOTAL	49	90.75%	5	9.25%	54

(Fig.7.4.) Distribution of syntactic mono- and polyfunctionality in polymorphic conditionality c-glossemes

When it comes to the degree of syntactic polyfunctionality (Fig.7.5.), the data reveal that 34 of the 46 syntactically polyfunctional forms that act as c-glossemes of conditionality serve only one additional syntactic function. This includes all 5 polymorphemic markers. Monomorphemic forms with two additional functions occur 9 times, while those with 3 and more functions – 3 times.

	1 additional function	2 additional functions	3 additional functions	4 additional functions	TOTAL
monomorphemic	29	9	2	1	41
polymorphemic	5	-	-	-	5

(Fig.7.5.) Degree of syntactic polyfunctionality of conditionality c-glossemes

The 46 polyfunctional conditionality c-glossemes have homonyms/polysemes in a variety of syntactic and grammatical categories which, for the purpose of this study, have been put into 20 categories listed below in (Fig.7.6.). Unlike in the case of the three previously discussed relations (anteriority, causality and purpose), there is no single category of polysemes/homonyms that would outnumber the remaining ones in the sample.

		I occurrences	II occurrences in binary overlaps	III occurrences in multiple overlaps
1.	ADVERB	10	3	7
2.	COMPLEMENTIZER	9	3	6
3.	ADPOSITION	9	5	4
4.	CASE MARKER	5	5	-
5.	TOPIC MARKER	5	3	2
6.	CONJUNCTION	5	1	4
7.	VERB	4	4	-
8.	TAM MARKERS AND MODALITY MARKERS	3	3	-
9.	NOMINALIZER	3	1	2
10.	DEMONSTRATIVE	3	2	1
11.	PARTICIPLE MARKER	2	-	2
12.	INTERROGATIVE	2	-	2
13.	RELATIVE CLAUSE	1	1	-
14.	NEGATION MARKER	1	1	-
15.	PARTITIVE MARKER	1	1	-
16.	PREDICATIVE MARKER	1	1	-
17.	SEQUENTIALITY MARKER	1	1	-
18.	DIRECTIONALITY MARKER	1	-	1
19.	EQUATIVE LINKER	1	-	1
20.	COPULA	1	-	1
TOTAL		68	35	33

(Fig.7.6.) Syntactic overlaps of conditionality c-glossemes

Statistically the most common category is the category of adverbs closely followed by complementizers and adpositions. These 3 categories make up over 40% of the 68 identified overlaps. Case and topic markers, as well as conjunctions occur 5 times each. Verbs occur 4 times and various TAM and modality markers 3 times. Interestingly, all of these occurrences are in binary overlaps.

Judging only from (Fig.7.6.) we would be tempted to say that the categories that may be pretenders to the title of the most common sources of conditionality linkers are categories 1 – 8 with the exclusion of conjunctions. To verify this initial observation and to be able to apply other of the grammaticalization heuristics we need to look at the specific meanings/functions of the categories (Fig.7.7.) and control for their semantic polyfunctionality (cf. entries with numbers in brackets in Fig.7.7. and Fig.7.8. listing the identified semantic overlaps).

Among adverbs overlapping in form with conditionality c-glossemes the most numerous are temporal adverbs with the meaning of ‘then’. They occur 6 times, in 54

cases in multiple overlaps. Only two of these markers are semantically monofunctional. The other 6, as depicted in (Fig.7.8), display semantic overlaps with relations such as SIOVER, SIDUR, RESULT, as well as CAUSE, PURPOSE and MANNER. Those that occur in binary overlaps are semantically polyfunctional with scope over SIOVER and SIDUR. This may indicate that the temporal meanings got reanalysed into conditional ones¹. Interestingly, the Vitu marker *kini/kunu/kene*, classified by van den Berg and Bachet as sequentiality marker rather than adverb (and glossed as ‘then’ or ‘and then’), is also used as a marker of conditionality:

(7.1a) Vitu (van den Berg and Bachet 2006: 113)

Hita ta kalinga-ni-a boro kini zahe kara polok-a mugomugo
 1PL REAL chase-TR-3SG pig then go.up to inside-3SG forest

‘We chased the pig and then it went up into the bush.’

(7.1b) (ibidem: 114)

Kini kuzabarae, ia ti kemi
 {COND} like.that 3SG PRF.3SG good

‘If it is like that, it is fine.’

The authors have stated that “given the relatively large semantic distance between the notions of sequentiality and conditionality, we treat the conditional use as a separate homophonous morpheme” (2006:114). The cross-linguistic frequency of overlaps of conditionality markers with the temporal adverbs observed in my investigation would suggest that the distance is not as large as it might have been thought.² There are two possible explanations for the fact that conditionality markers overlap in form with temporal adverbs. The first one would assume that the adverbs gave rise to temporal clause markers first and the markers then extended their scope over the relation of conditionality. Thus, the ADV-LINKERconditionality would be an epiphenomenon of the LINKERtemporal-LINKERconditionality overlap. The proposed grammaticalization pathway would than look as follows:

ADV‘then’ → LINKERtemporal → LINKERconditional

¹The affinity between temporal and conditional concepts has already been mentioned in section 4.3. and is elaborated on below in section 7.3. Cross-linguistic data quoted in other works on grammaticalization (Traugott 1985, Heine and Kuteva 2002 – cf. section 3.2.8.) strongly suggest that it is the temporal markers that develop into the conditional ones, rather than the other way round. This follows the tendency of less abstract meanings developing into more abstract ones discussed in section 3.2.9.

² Note also that Traugott (1985:292) lists temporal expressions of duration and temporals ambiguous between duration and punctuality as sources of conditional markers (see section 3.2.8).

POLYSEMOUS/HOMONYMOUS CATEGORY	CHARACTERISTICS OF POLYFUNCTIONALITY	
ADVERB Total:10	BINARY OVERLAPS:3 ADV'then':1 ADV'sometime';'then':1(1) ADV'at that time';'then':1 (1)	MULTIPLE OVERLAPS:7 ADV + COMP (+ other):3 ADVequative + COMP + INTER'how':1 (1) ADVequative + COMP:1 (1) ADV'also' + COMP + NMLZ:1 (1) ADV+ other: 4 ADV'then' + INTER'how':1(1) ADV'then' + CONJ'but':1 (1) ADV'then';'well' + DEM:1 (1) ADV'then' + ADP: 1
COMPLEMENTIZER Total:9	BINARY OVERLAPS:3 (1)	MULTIPLE OVERLAPS:6 COMP + ADV (+ other):3 COMP + ADVequative + INTER'how':1 (1) COMP + ADVequative:1 (1) COMP + ADV'also' + NMLZ:1 (1) COMP + ADP:1 (1) COMP + ADP'of';'for' COMP + CONJ:2 COMP + CONJ'and'(NPs):1 (1) COMP + CONJ'or':1
ADPOSITION Total:9	BINARY OVERLAPS: 5 ADPsource:1 (1) ADP'about';source:1 ADPlocative:2 (2) ADP'(immediately after)':1 (1)	MULTIPLE OVERLAPS:4 ADP'of';'for' + COMP:1 (1) ADPcomitative + CONJ'and':1 (1) ADPcomitative + TOP + PTCP + DIR:1 (1) ADPcomitative + TOP:1

CASE MARKER Total:5	BINARY OVERLAPS:5 CASEallative:1 (1) CASElocative:1 (1) CASEinstrumental:1 (1) CASEgenetive:1 (1) CASEnominative:1	MULTIPLE OVERLAPS:0
TOPIC MARKER Total:5	BINARY OVERLAPS:3 (1)	MULTIPLE OVERLAPS:2 TOP + ADPcomitative + PTCP + DIR:1 (1) TOP + ADPcomitative:1
CONJUNCTION Total:5	BINARY OVERLAPS: 1 CONJ'and':1 (1)	MULTIPLE OVERLAPS: 4 CONJ'and'(NPs) + COMP:1 (1) CONJ'or' + COMP:1 CONJ'and' + ADPcomitative:1 (1) CONJ'but' + ADV'then':1
TAM MARKERS AND MODALITY PARTICLES Total:4	BINARY OVERLAPS:4 FUT:2 VERBSUFadhortative:1(1) IRR:1	MULTIPLE OVERLAPS:0

VERB Total:4	BINARY OVERLAPS:4 VERB'do':2 VERB'chance upon':1 VERB'say';'tell':1 (1)	MULTIPLE OVERLAPS:0
NOMINALIZER Total:3	BINARY OVERLAPS:1	MULTIPLE OVERLAPS:2 NMLZ + PTCP:1 (1) NMLZ + ADV'also' + COMP:1 (1)
DEMONSTRATIVE Total:3	BINARY OVERLAPS:2	MULTIPLE OVERLAPS:1 (1) DEM + ADV'then';'well'
PARTICIPLE MARKER Total:2	BINARY OVERLAPS:0	MULTIPLE OVERLAPS:2 PTCP + NMLZ:1 (1) PTCP + TOP + ADPcomitative + DIR:1 (1)
INTERROGATIVE Total:2	BINARY OVERLAPS:0	MULTIPLE OVERLAPS:2 INTER'how' + ADV'then':1(1) INTER'how' + ADVequative + COMP:1 (1)

(Fig.7.7.) Details of syntactic overlaps of purpose c-glossems

The conventions of data presentation are the same as described for (Fig.4.7.) in chapter 4.

The second scenario that comes to mind arises from the simple fact that cross-linguistically the temporal adverb ‘then’, marking the consequence clause, is used often as an explicit marker of apodosis (cf. *If you are planning to go for a walk, then you should better take an umbrella*). As Podlesskaya notices for Russian, “when the apodosis is marked explicitly, the protasis is often (though not always) self-sufficient in that it may contain no device signalling that its content will be somewhat utilized in the further discourse” (2001:1003). The same holds for many of the world’s languages. Taking into account these two observations it is not unreasonable to conclude that at least in some cases the adverb ‘then’, as a marker of protasis, may become reanalysed as a marker of apodosis.³

As for other adverbs in the sample, equative adverbs (‘as’) occur twice and in both cases they overlap in form with clause linkers encoding the meaning of comparison, similarity or manner (cf. Fig.7.8). It is clear that these three relations are semantically closer to the meaning of the adverb than the concept of conditionality. This most likely indicates only that the ADVEquative-LINKERcond overlap is of epiphenomenal character if not accidental.⁴

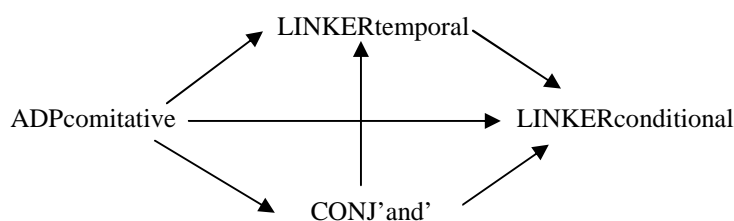
The second category listed in (Fig.7.7.) is complementizers. The data does not yield up a clearly possible explanation for the 9 COMP-LINKERconditionality overlaps. There are only two examples of binary COMP-LINKER overlaps where the linker is not used to express other circumstantial relation apart from conditionality. We know that complementizers may become markers of conditional clauses as epiphenomena of PURPOSE-COND overlap (see section 3.2.8. where the implicational hierarchy of grammaticalization of complementizers proposed by Saxena has been quoted) and this is the most likely explanation for the origin of the COMP-LINKERconditionality overlap in Ndyuka (cf. Fig.7.8.). In Japanese and Burushaski, as we read from (Fig.7.8.), the complementizer which overlaps in form with a conditionality linker has the same phonological shape as a coordinative conjunction. I am not aware of any work that gives an example of an attested grammaticalization pathway of complementizer into conjunction (or the other way round) and so I will not suggest here the existence of such a chain. In five other cases of COMP-LINKERconditionality overlaps our only clue to the explanation of the polyfunctionalities is the fact that the complementizers overlap in form also with markers of temporal relations and in one case with the marker of causality. The three remaining COMP-LINKERconditionality overlaps (contributed by

³ An example of an identical form marking both protasis and apodosis, comes from the Australian language Yindjibarndi where the determiner clitic *-yhu* is used in both these functions – cf. example 3.7. in chapter 3.

⁴ Note also that the first equative adverb listed in (Fig.7.7.) is identical in form with complementizer and interrogative ‘how’. It is the latter one that reveals an especially close link with the concept of manner. See also Heine and Kuteva (2007:242-244) for more details on a cross-linguistic pattern of grammaticalization interrogative → subordinator mentioned earlier in section 3.2.8.

Apache Jicarilla *-go/-o*, Sango *tongana* and Polish *jak*) do not offer any reliable clues regarding the order of development of the particular meanings/functions. On the whole, complementizers do not seem to be an important direct source of conditionality linkers. In some cases they appear to be at most candidates for indirect sources of these c-glossemes in pathways of grammaticalization that involve markers of purpose and possibly also temporal relations and causality.

The next two groups listed in (Fig.7.7.) – adpositions and case markers – do not remain unproblematic either since it is difficult to identify the meaning/function of these categories that would reveal a clear semantic affinity with markers of conditionality and would ultimately provide us with an explanation for the origin of the overlaps (cf. Fig.7.7.). Only in 2 of the 9 cases of items with ADP-LINKER overlap the marker does not encode any other circumstantial meaning apart from conditionality and only one of these markers occurs in a binary overlap. The remaining ones overlap in function with temporal markers and in one case (the already mentioned Ndyuka example), with purpose (see Fig.7.8.). Although among the adpositions listed in (Fig.7.7.) the comitative adpositions are the most numerous group, it is difficult to assign to them a significant role in the formation of conditionality markers. Firstly, because they occur only in multiple overlaps (including topic markers, which have been known to be sources of conditional linkers, and in one case also coordinating conjunction). Secondly, because the items displaying these overlaps are used to mark temporal relations as well. We know from cross-linguistic studies that comitative markers develop *inter alia* into markers of temporal inter-clausal relations and into both phrasal and sentential coordinating conjunctions (Heine and Kuteva 2002:78-90 and references therein). We are also aware of cases where a comitative preposition is identical with an ‘if’ particle (Heath 2004:74). Finally, it has been noticed that coordinating conjunctions, such as ‘and’ may come to be used as subordinating conjunctions, including conditional and temporal markers (Harris and Campbell 1995:290). All these findings indicate, again, quite a complicated chain of grammaticalization possibilities which can be depicted as follows:



The temporal meaning occurs also as a common semantic overlap of conditionality linkers displaying polysemy/homonymy with case markers. This tendency has been

noticed by Aikhenvald (2008:504, see also section 3.2.8.) with respect to locative case markers but in my sample it has been identified also for genitive and instrumental markers. The CASEinstrumental-LINKERconditionality overlap has been found in Tamil where the *-aal* marker encodes also the meaning of SIOVER and CAUSE.

Another case marker – allative – contributed by the Basque *-ra* suffix, apart from sharing its form with the conditionality *c-glosseme*, may also be used as a purpose linker. In (Fig.7.7.) there is only one instance of a CASE-LINKER overlap that does not involve meanings other than conditionality. Taking into account the semantic polyfunctionality of the linkers displaying overlaps with case markers and the types of circumstantial relations over which they have scope, it appears that the case markers are not a cross-linguistically common direct source of conditionality linkers.

The link between topic markers – the fifth group depicted in (Fig.7.7.) – and conditionals has been described in detail by Haiman (1978). The author based his strong claim that topics are conditionals on three facts:

- a) the protasis almost always proceeds apodosis (cf. Greenberg’s universal number 14);
- b) in many languages the topic marker and conditionality marker are identical;
- c) conditionals and polar interrogatives often use identical or very similar marking.

Since the publication of Haiman’s observations, some linguists have challenged his claim suggesting instead that conditionals are only partially topics in function (cf. Akatsuka 1986, Ferguson et al. 1986:10). Nonetheless, as Zaefferer (1991a:218) remarks, when it comes to the identicalness of topic and conditionality markers in unrelated languages: “there is something to be accounted for”.

In my sample examples of TOPIC-LINKERconditionality overlaps have been found in Ama, Hatam, Kanuri (2 examples) and Batak Karo. Three of these five syntactic overlaps are binary but the Ama marker is also semantically polyfunctional expressing the meanings of COND, SIDUR and ANTE (cf. Fig.7.8.). In Kanuri, apart from being a topic and conditionality marker, the *-ga* marker may also be used as a comitative adposition. This introduces some more confusion since, as has been already said, comitative adpositions themselves are a potential source of conditionality markers. Without more detailed analysis it is difficult to propose here any convincing grammaticalization scenario so we can only repeat here the pathway originally depicted by Haiman:

TOPIC MARKER → LINKERconditionality

As already mentioned, the fact that markers of conditionality may develop from coordinating conjunctions has been emphasized by Harris and Campbell (1995:290). The authors give an example of Mingrelian *da* which in proto-language formed coordinated sentences and in modern Mingrelian is used as a marker of conditional

sentences. In my sample, as depicted in (Fig.7.7.) there is only one example of binary CONJUNCTION-LINKERconditionality overlap. It comes from Khwe where the *nò/nù* marker is used in these two functions and as an exponent of variety of other circumstantial meaning (which is not surprising taking into account, as discussed in section 1.5.2, that various circumstantial readings may be drawn from coordinated clauses). In Japanese, on the other hand there is a conditional marker *-to* (encoding the relations of ANTE and SIOVER too) that can also be used as a coordinating conjunction and a comitative adposition. As discussed above, other typological findings suggest that the comitative adpositions give rise to coordinating conjunctions⁵ so putting the data together we may hypothesize the following pathway of grammaticalization:

ADPcomittaitve → CONJUNCTION'and' → LINKERconditionality

The last example of coordinating conjunction sharing its form with conditionality marker comes from Burushaski, where the marker is used also as complementizer. The *kè* conjunction is an NP rather than VP coordinator though. This CONJ-COMP-LINKER overlap is an isolated example for which I have not found any parallels in the grammaticalization literature. In total there are then only two examples of VP'and'-LINKERconditionality overlaps both of which reveal also semantic polyfunctionality.

The remaining conjunctions listed in (Fig.7.7.) are the Au *-te* ('but') and the Baure *apo* ('or'). Dixon points out that

some languages (particularly those in the Oceanic branch of Austranesian) use the same syntactic marker for Disjunction ('or') and for Conditional ('if'). For Disjunction we get two clauses linked by the 'or/if' marker, which comes between them, whereas for Conditional the 'or/if' marker precedes the Supporting clause. (2009:14)

This CONJ'or'-LINKERconditionality overlap, at least in my sample, does not seem to be cross-linguistically common though. Moreover, since both the disjunctive and alternative conjunction mentioned above are semantically polyfunctional (their scope includes temporal markers – cf. Fig.7.8.) we need to accept that it is possible that in these two languages the conjunctions gave rise to conditionality markers indirectly or that we deal here with cases of polygrammaticalization.

Regarding the category of verbs – only four instances of conditionality linkers with polysemes/homonyms in this category have been identified in my sample: the Galo verb *əəm* ('say', 'tell'), the Seediq verb *netun* ('chance upon') and the Swahili *ikiwa* and *iwapo* – words that, apart from being conditionality markers, may act also as verbs with the meaning 'do'. The Galo marker is semantically polyfunctional which, according to the grammaticalization heuristics listed in section 3.2.9., leaves us with only three examples of prototypical candidates for sources of conditionality markers. In

⁵ Cf. also Stassen (2005:258-261) discussing cross-linguistic distribution of languages in which NP coordinator 'and' is identical to 'with' and Haspelmath (2005:262-265) discussing cases of identicalness and non-identicalness of nominal and verbal coordinating conjunctions.

the following section, however, we will see that verbs, as Traugott (1985) has claimed are by no means rare as material incorporated in polymorphemic conditionality linkers.

The same can be said about TAM and modality markers. There are three conditionality c-glossemes identified in the sample as homonymous/polysemous with TAM markers: Kayah Li perfective irrealis *ke*, Ilokano future markers *intono* and *no* and Lavukaleve adhortative verbal suffix *-le*.

Irrealis, which is most commonly regarded as a modal category, and future (tense category) are closely related concepts in that they both indicate that a particular event has not appeared (in the case of future, more specifically, that the event has not appeared yet). Similarly, the idea of adhortativeness is by definition linked to the concept of an event yet unrealized. It is easy to see where the affinity between conditionality and these TAM and modality markers lies – the protasis, in the case of real conditionals looked at in this thesis, always encodes the SoA which is yet unrealized, to occur in the future.⁶ Again, however, since the category of TAM and modality markers includes only three items and since the adhortative Lavukaleve marker is used also in other semantic functions, the category is not be treated as one constituting a cross-linguistically common source of conditionality markers.

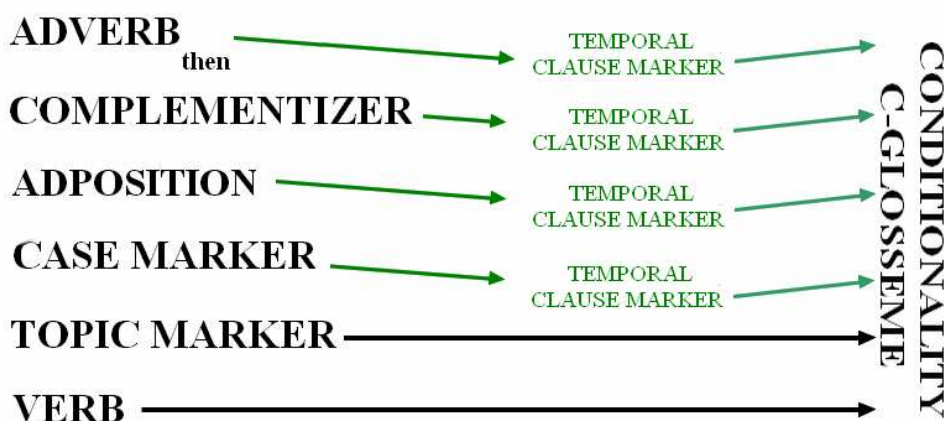
conditionality linker	syntactic overlap	semantic overlap
Krongo <i>-má</i>	ADV'at that time';'then'	SIOVER, SIDUR
Chukchi <i>tite</i>	ADV'sometimes';'then'	SIDUR
Sango <i>tongana</i>	ADVequative + COMP	SIOVER, COMPAR
Yanyuwa <i>namba</i>	ADV'well';'then' + DEM	SIOVER, RESULT
Au <i>-te</i>	ADV'then' + CONJ'but'	RESULT, CAUSE, PURPOSE
Polish <i>jak</i>	ADVequative + INTER'how' + COMP	SIOVER, TAQUO, SIMIL, MANNER
Apache Jicalrilla <i>-go/-o</i>	ADV'also' + COMP + NMLZ	SIOVER, CAUSE
Estonian <i>kui</i>	ADV'then' + INTER'how'	SIOVER, MANNER
Ket <i>-qaka</i>	ADPllocative	SIOVER
Ket <i>-kuuka</i>	ADPllocative	SIOVER
Rama <i>-ka</i>	ADPsource	SIOVER
Ndyuka <i>fu/fi</i>	ADP'of';'for' + COMP	PURPOSE
Kanuri <i>ga</i>	ADPcomitative + DIR + TOPIC + PTCP	SIOVER
Japanese <i>-to</i>	ADPcomitative + CONJ'and'	SIOVER, ANTE
Polish <i>skoro</i>	ADP'(immediately) after'	CAUSE
Basque <i>-ra</i>	CASEallative	PURPOSE
Ket <i>-ka</i>	CASElocative	SIOVER, SIDUR
Tamil <i>-aal</i>	CASEinstrumental	SIOVER, CAUSE

⁶ Hopper and Traugott see the affinity between modalities (including verb with modal meanings) as well as interrogatives and conditionals in general (hence also counterfactual and hypothetical conditionals) in the fact that “Conditionals raise possibilities and cast doubts on propositions; therefore the presence of modalities and interrogatives among the sources of conditionals seems naturally motivated by the function of conditionals” (2003:186).

Yami <i>no/ano/anu</i>	CASEgenitive	SIOVER
Taba <i>polo</i>	COMP	SIOVER
Burushaski <i>kè</i>	COMP + CONJ'and'(NPs)	SIOVER, ANTE
Khwe <i>nò/nù</i>	CONJ'and'	general c-glosseme
Galo <i>æm</i>	VERB'say';tell'	general c-glosseme
Lavakuleve <i>-le</i>	adhortative suffix	COCOND, SIOVER
Ama <i>mo</i>	TOPIC	SIDUR, ANTE
Paiwan <i>nu</i>	PRTT	ANTE, SIOVER
Meyah <i>erek</i>	COP + EQUAT	MANNER
Quechua Huallaga <i>-sha</i>	NMLZ + PARTICIP	general c-glosseme

(Fig.7.8.) Syntactically and semantically polyfunctional conditionality c-glossemes

On the whole we can say that the analysis of synchronic patterns of polysemy does not allow us to identify many direct sources of conditionality linkers. It is also clear that there is no single homonymous/polysemous category that would outnumber others as has been observed for the other three circumstantial relations. In the majority of cases the grammaticalization pathways include also markers of temporal relations between clauses (and in some isolated cases, not marked in the diagram below, also markers of purpose and causal relations). As mentioned earlier, on the basis of other typological evidence and the less abstract → more abstract grammaticalization tendency, in such cases I assume that it is the conditional markers that develop from temporal ones and not the other way round. I mark this route with green arrows in (Fig.7.9.) which depicts that 4 of the 6 categories that we may call most common sources of conditionality linkers: adverbs, complementizers, adpositions and case markers appear to be more often indirect than direct sources. Verbs and topic markers remains candidates for direct sources.



(Fig.7.9.) Cross-linguistically most common sources of conditionality markers reconstructed on the basis of patterns of synchronic polysemy

My summary lacks two of the categories that have been listed by Traugott (1985), as well as by Heine and Kuteva (2002) as commonly giving rise to

conditionality linkers: copulas and interrogatives. Let us see whether the results of analysis of the material incorporated in polymorphemic conditionality markers change the picture.

7.2.2. Polymorphemic markers and their internal structure

The 54 polymorphemic markers of conditionality found in the sample come from a total of 19 languages. The variety of categories incorporated in the structure of the polymorphemic markers is impressive, but it needs to be emphasized that in many cases it is due to one language that the total number of occurrences of a particular category appears high. For instance, locative adpositions that occur 9 times in the incorporated material come from four languages: English, Ket, Basque and Polish, but it is the first two that contribute 7 of these 9 occurrences. Similarly in the case of adverbs 4 of the 5 cases come from English (linkers *as long as* and *so long as*).

Nouns and verbs are good candidates for being labelled as the most common building blocks. As can be seen from the table, however, there is a significant diversity in the meanings of these two groups: among nouns we find ‘condition’, ‘event’, ‘word/speech’ and ‘thing’, but only the first one occurs in two languages (Polish and English) – all the other ones are language specific. Among verbs it is the verb ‘assume’ that occurs in two languages (again Polish and English) while ‘provide’ and ‘give’ are language-specific. It cannot be left unnoticed that the verb/copula ‘be’ (reported in linguistic literature as a common source of conditionality markers – see section 3.2.8.), as well as verb ‘say’ (polysemous with complementizer in Thai), contribute 10 items to the table. Again, however, these two are not very common cross-linguistically – in fact the 10 items come from just three languages: Mayogo, Swahili and Akan. It can be, therefore, said that nouns and to greater extent verbs are commonly incorporated into the polymorphemic c-glossemes but it would be difficult to indicate the specific meanings of these two groups that are cross-linguistically common in the context of idiomatization into conditionality markers.

Similarly, in the group of interrogatives (which, too, have been claimed to be common sources of conditionality markers cf. 3.2.8.), case markers and complementizers, single languages contribute many items. The best example of it is the already mentioned locative case marker that is present in all of the Ket c-glossemes and in those of them that are distributed, it occurs even twice.

	ADP	ADV	NOUN				VERB					VERB + COMP		COMP	COMP+ OTHER	DEM	INTERROGATIVE		NMLZ + PTCP	CASE	C-GLOSSEME	OTHER	NOT KNOWN		
			locative	'at that time/' 'then'	degree/manner/ (quantity)	'condition'	'event'	'word'/speech'	'thing'	'provide'	'assume'	'be'	'give'				'say'	'deserve', 'be fit', 'be equal to', 'resemble'+COMP						'say' + COMP	'how many/much'
Khwe (2)																						+		•	
Khwe (3)																•							+		•
Krongo (2)		+																						•	
Lezgian (2)																							+ • ▼		
Lezgian (2)																							+ • ▼		
Mandarin (2)						•																		•	
Mandarin (3)						•																	+	•	
Mayogo (3)										•							•							•	
Mayogo (4)										•							•						+	•	
Mayogo (3)										•							•						+	•	
Mayogo (2)										•							•								
Nisga'a (2)																•							• ('when')		
Polish (2)																								•	
Polish (3)										•														•	
Polish (4)	•															•							• (Ins.)		
Sango (2)																							• ▼	+	
Sapuan (2)																									
Santali (2)																							+ • ▼		
Santali (2)																							+		
Swahili (3)										•													+	•	
Swahili (4)										•													+	•	
Swahili (2)																•							+	•	
Thai (2)																							+		
Thai (3)																							+ +		

	ADP		ADV		NOUN			VERB					VERB + COMP		COMP	COMP+ OTHER	DEM	'how many/much'	INTERROGATIVE	NMLZ + PTCP	CASE	C-GLOSSEME	OTHER	NOT KNOWN	
	locative	'at that time/' 'then'	degree/manner/ (quantity)	'condition'	'event'	'word'/speech'	'thing'	'provide'	'assume'	'be'	'give'	'say'	'deserve', 'beftit', 'be equal to', 'resemble'+COMP	'say' + COMP											
Thai (2)																						+	+		
Thai (2)																							+		•
Thai (3)													•										+		•
Yup'ik (2)																							+		•
	9	1	4	3	1	2	2	4	3	6	2	1	1	3	3	7	6	1	4	4	11	30	24	5	
		5			8			16					4					5							

(Fig.7.10.) Material incorporated in polymorphemic conditionality c-glossemes

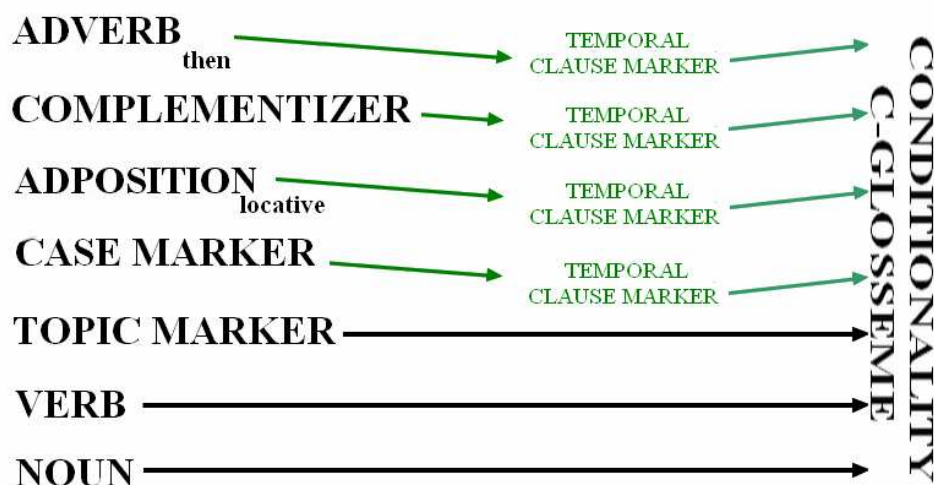
The conventions of data presentation are the same as described for (Fig.4.10.) in chapter 4; the ▼ symbol means the item is a borrowing.

The points in the columns labelled 'other' are described below. The '+' symbol in the description means that the morpheme is polysemous.

1. Category COMP+ OTHER: COMP and DEM for English linkers; COMP and IRR for Nisga'a.
2. Category OTHER: REL for Akan; ADVthen + ADPafter for the first Basque and NMLZ and POSS for the second one; PTCP + PST for the first two English linker, DEF for the third one, PTCP + PST for the following two and ADJ'long' for the last two; NMLZ and PRON for Hatam; AUX for Krongo; NMLZ + REL for Mandarin; ADV + COMP for Mayogo; polyfunctional adposition for the first Polish linker and PTCP for the second one; ADVequative + COMP for Sango; class markers for the first two Swahili linkers and REL for third one.

The most numerous group listed in (Fig.7.10.) are, quite surprisingly, c-glossemes which occur as incorporated material in 11 languages in the sample. The frequency of these morphemes is higher than for any other of the three circumstantial relations described in the previous sections. None of the polymorphemic markers incorporating c-glossemes in its structure is semantically polyfunctional. In Khwe and Nisga'a the addition of the c-glosseme to the other morpheme(s) results in a complex c-glosseme encoding exclusively the meaning of conditionality and can be therefore viewed as a mechanism enhancing explicitness of marking. In Lezgian, Sango and Santali the original, already unambiguous conditionality c-glossemes may be enriched by the addition of borrowed morphemes (marked by the triangles in Fig.7.10.). In this case it is rather the issue of prestige of the loaning language than of increased explicitness that explains the origin of these complex markers (for further discussion see section 9.3.4.).

In the light of the facts presented here, we can conclude that the categories most often incorporated in complex conditionality c-glossemes are: locative adpositions, complementizers, verbs (including verbs/copulas) and nouns. Hence, there are two elements that we may add to the reconstruction presented in (Fig.7.11.): the category of locative adpositions (which has also occurred twice in Fig.7.7.) and the category of nouns:



(Fig.7.11.) Cross-linguistically most common sources of conditionality markers reconstructed on the basis of patterns of synchronic polysemy and material incorporated into polymorphemic linkers.

Copulas and interrogatives, which appeared in both Heine and Kuteva's (2007) and Traugott (1985) summaries, have not made it to the final set in my study since their frequency was much lower than that of the other relations.

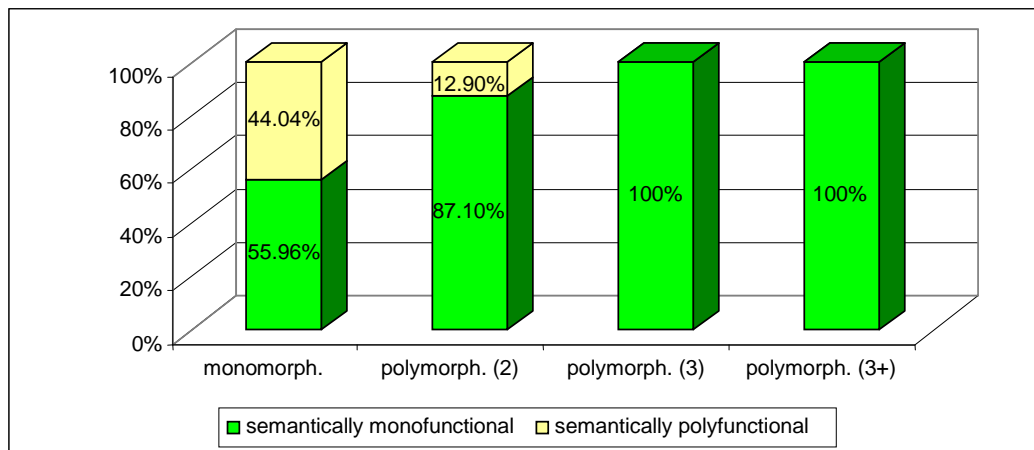
7.3. SEMANTIC POLYFUNCTIONALITY AND COGNITIVE AFFINITY

Quantitative analyses of the material collected (cf. Fig.7.12.) reveal that the semantically monofunctional conditionality c-glossems are over twice as numerous as those which in addition to conditionality express also other functions. Similarly as for the other three circumstantial relations discussed in the preceding chapters also in the case of conditionality we observe the tendency that the number of polyfunctional markers decreases when the internal complexity of the markers increases. None of the 23 markers that consist of 3 or more morphemes is semantically polyfunctional. Among the bimorphemic linkers the ratio of monofunctional to polyfunctional ones is 6.75:1, while for the monomorphemic ones only 1.3:1.

	semantically monofunctional		semantically polyfunctional		TOTAL
	count	%	count	%	
monomorphemic	61	37.42%	48	29.45%	109
2 morphemes	27	16.57%	4	2.45%	31
3 morphemes	17	10.43%	-	-	17
3+ morphemes	6	3.68%	-	-	6
TOTAL	111	68.10%	52	31.90%	163

(Fig.7.12.) Distribution of c-glossems of conditionality according to their semantic mono/polyfunctionality and morphological complexity

The graph in (Fig.7.13.) depicts the share of semantically monofunctional and polyfunctional markers for each of the groups of linkers separately. It shows clearly once again that the data reveal the complexity → monofunctionality tendency. In fact, the tendency is stronger for conditionality than for any other of the relations that are analysed in this thesis. I go back to this issue in chapter 8.



(Fig.7.13.) Ratio of semantically monofunctional and polyfunctional markers in conditionality c-glossemes with different morphological complexity

As for the degree of semantic polyfunctionality, among the 54 markers for which the value has been determined, half are linkers with scope over conditionality and one other relation (Fig.7.14.). The number drops to just under 26% for linkers expressing conditionality and two other relations. There is also a significant number of general c-glossemes (20.37%), but only 2 cases of linkers with scope over conditionality and three other circumstantial relations.

	scope over 2 relations	scope over 3 relations	scope over 4 relations	scope over 5+ relations	TOTAL
count	27	14	2	11	54
%	50.00%	25.93%	3.70%	20.37%	100%

(Fig.7.14.) Degree of semantic polyfunctionality of conditionality c-glossemes

More detailed information on the degree of semantic polyfunctionality and morphological complexity is available for 49 of the linkers. It is evident (see Fig.7.15.) that the complexity \rightarrow monofunctionality tendency holds strongly. There are only 6 polymorphemic markers which are semantically polyfunctional, both of which are bimorphemic. Among the monomorphemic ones, almost 40% have scope over conditionality and one more circumstantial relation. The same holds for over 26% of markers with scope over 3 relations (13 items) and for 4% of those expressing 4 circumstantial meanings. The rest of the monofunctional linkers are those which I classify as general ones.

	scope over 2 relations		scope over 3 relations		scope over 4 relations		scope over 5+ relations	
	count	%	count	%	count	%	count	%
monomorphemic	19	38.78%	13	26.53%	2	4.08%	9	18.37%
2 morphemes	6	12.24%	-	-	-	-	-	-

(Fig.7.15.) Distribution of c-glossemes of conditionality according to their degree of polyfunctionality and morphological complexity⁷

When it comes to the depiction of the network of semantic affinities of the relation of conditionality, there are, as shown in (Fig.7.16.), 13 circumstantial relations that have been reported in the literature to overlap with conditionality. One of them – SIOVER (‘when’) – clearly outnumbers others both in the total number of occurrences (32), and in the occurrences in binary overlaps among which it contributes almost 63% of cases. SIOVER is followed in the rank table by anteriority and causality. These two, however, occur in binary overlaps with conditionality only three times altogether. Other relations are even less frequent and among them only two – PURPOSE and SICOEX (simultaneity co-extensiveness ‘as long as’) occurs in binary overlaps more than once.

	occurrences of overlaps	occurrences in binary overlaps	
		count	percentage of the total number of binary overlaps (27)
SIOVER	32	17	62.96%
ANTE	7	1	3.71%
CAUSE	5	2	7.40%
PURPOSE	3	2	7.40%
MANNER	3	1	3.71%
COCOND	3	1	3.71%
SIDUR	3	-	-
SICOEX	2	2	7.40%
CONC	2	1	3.71%
RESULT	2	-	-
TAQUO	1	-	-
COMPAR	1	-	-
SIMIL	1	-	-
TOTAL	65	27	100%
general c-glossemes	11		
TOTAL including General c-glossemes	76		

(Fig.7.16.) Summary of semantic overlaps of conditionality c-glossemes

The convention of data presentation are the same as for (Fig.4.15.) in chapter 4.

⁷ Data available for 49 items.

The network of semantic affinities of conditionality (Fig.7.17.) resembles the one which has been reconstructed for purpose in section 6.3. in that it is organized neatly and almost exclusively around one dominating meaning – simultaneity overlap (‘when’). On the whole there are only two linkers completely detached from the main loop.⁸ All the other ones are either directly or indirectly bound to SIOVER. The frequent cross-linguistic attestation of linkers used to express both the relation of SIOVER and COND has already been mentioned in section 5.3. This phenomenon has not escaped linguists’ attention. Thompson and Longacre (1985) and later Thompson, Longacre and Hwang notice that:

In some languages, including Indonesian languages of Papua New Guinea, there is no distinction between ‘if’ clauses and ‘when’ clauses. In many of these languages, the neutralization holds, however only for *predictive* conditionals and *future* time clauses. (...) The distinction between English ‘when’ and ‘if’ clauses is simply one of degree of expectability, and is a distinction that many languages do not code. (2007:257-258)

Dixon adds also:

Many languages from all over the world use the same marker for ‘when’ (Temporal linking) and ‘if’ (Conditional linking). In some contexts only ‘when’ is possible, in some only an ‘if’ interpretation, and in others either. Which kind of linking is involved, has, in some languages, to be inferred from the semantics and pragmatics of the discourse in which the linking appears. (2009:14)

It comes as no surprise then that the SIDUR-COND overlap occurs frequently also in the data I collected for the purpose of this thesis where I chose to focus on real conditionals which, of all types of conditionals, are cognitively most closely related to the temporal (or even omnitemporal/habitual) concepts:

(7.2a) *If I press the button the computer will start.*

(7.2b) *When I press the button the computer will start.*

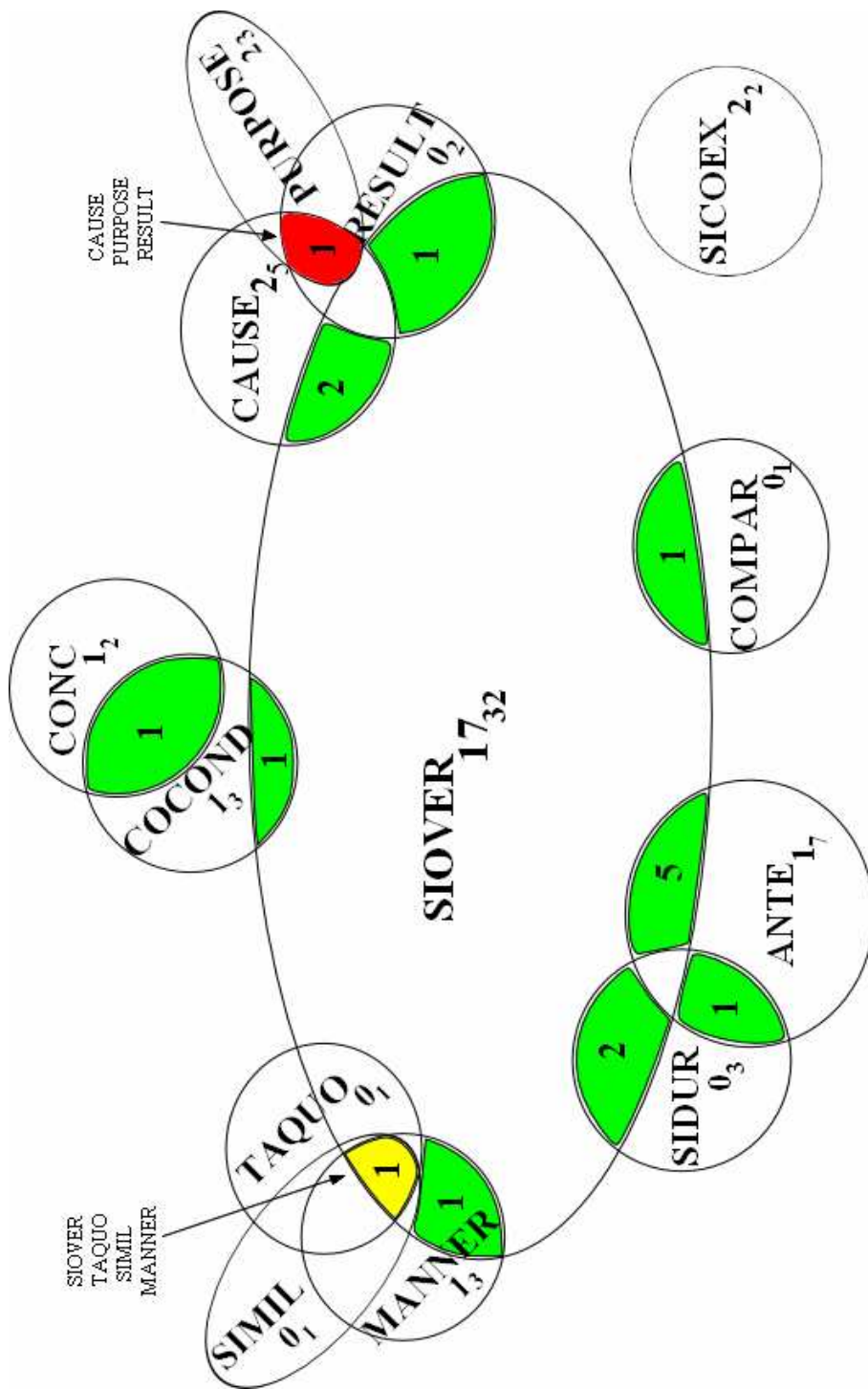
The overlap can be exemplified by the binary examples in Chukchi (7.3) and Taba (7.4a)

(7.3.) Chukchi (Dunn 1999:126)

Tite	<i>mik-ə-ne</i>	<i>ənqen</i>	<i>ɣe-n-ə-mlətj-ew-lin</i>
{COND/SIOVER}	someone-E-ERG	DEM.3SG.ABS	PRF-CAUS-E-wound-CAUS-3SG
<i>ənqen</i>	<i>winə-t</i>	<i>qonpə</i>	<i>ləyi n-ine-ly-ə-qin</i>
DEM.3SG.ABS	track-3PL.ABS	always	know HAB-TR-AUX-E-3SG

‘When/if somebody has wounded him (a bear) then he always knows their tracks.’

⁸ These are two English polymorphemic linkers *as long as* and *so long as*.



(Fig. 7.17.) Semantic polyfunctionality of conditional c-glossemes

The conventions of presentation are the same as applied for (Fig.4.16.) in chapter 4.

(7.4a) Taba (Bowden 1997:457)

K=rasa mapot polo k=oik au
1SG=feel heavy {COND/SIOVER} 1SG=leave.behind 2SG
'I (my heart) feel(s) heavy when/if I leave you.'

The polyfunctionality in Taba becomes evident if we compare (7.4a) with cases where *polo* has a clearly temporal (7.4b) and clearly conditional reading as in the example of counterfactual conditional in (7.4c):

(7.4b) (ibidem:456)

Polo t=cung um li boa me t=ha-osak
{SIOVER} 1PL.INCL=enter house LOC door well 1PL.INCL=CAUS-open
tahate do
be.impossible REAL
'When we entered the house, well we couldn't open the doors.'

(7.4c) (ibidem)

Polo yapyap n=tala John, t=ha-klol John te
{COND} ash 3SG=meet John 1PL.INCL=CAUS-know John NEG
'If John had been covered in ash, we wouldn't have recognized John.'

The next two temporal relations in the network – ANTE and SIDUR in the majority of their occurrences overlap not only with COND but also with SIOVER. The fact that one string of phonemes is used to express more than one circumstantial meaning surely suggest that the polyfunctionality came about in a series of semantic extensions. Without proper diachronic evidence, however, absolutely reliable reconstruction of the grammaticalization chains is not possible. Synchronic evidence in the form of patterns of syntactic polysemy does not give us a satisfactory answer. In Ket, for instance the *-ka* marker is used to express the relations of COND, SIOVER and SIDUR and acts also as the locative case markers (Andrey Nefedov – personal communication). We do know that locative markers often give rise to temporal clause linkers (see Aikhenvald 2008:594) but we cannot tell anything more specific about this particular case. If we apply here the assumption that the more abstract meanings develop from less abstract ones we may only hypothesize that it is the temporal meaning that precedes the conditional one in time.

An interesting example of a c-glosseme that encodes a variety of circumstantial meanings, including SIOVER, SIDUR, ANTE, COND as well as CAUSE comes from the Australian language Pitjatjantjara⁹:

⁹ The c-glosseme, due to the fact that it encodes 5 different circumstantial meanings is classified here as a general c-glosseme. Interestingly, it is not used for encoding the concept of purpose (see Eckert and Hudson 1994:265-270). The examples are quoted as they are presented in the reference grammar – without exact glosses.

(7.5a) Pitjantjatjara (Eckert and Hudson 1994:266)

Tjintu pakan-nyangka-na wankaringu
Sun rising-**{SIOVER}**-I awoke

‘I awoke when the sun came up.’

(7.5b) (ibidem)

Ngayulu kunkunpa ngari-nyangka palutu pitjangu
I asleep lying-**{SIDUR}** he/she came

‘She came while I was asleep.’

(7.5c) (ibidem)

Palumpa kutangu ngurakutu iyanu mai ngalku-nyangka maiangka
His/her senior.brother home.to sent food eating-**{ANTE}** afterwards

‘His brother sent him home after he’d (eaten) a meal.’

(7.5d) (ibidem:268)

Nyarakutu anku-nyangka-nta-ya watingku kuwari wataparara nyanganyi
Yonder.to go-**{COND}**-you-they man now chasing are.seeing

‘If you go over there some men will/might go after you and watch you.’

(7.5e) (ibidem:271)

Kata lirpungku-nyangka-na ananyi ngurakutu kunkunkitja
head aching-**{CAUSE}**-I am.going camp.to sleep.for

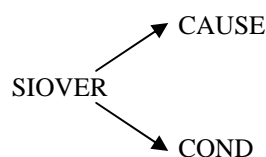
‘As my head is aching I’m going to camp for a sleep.’

The next cluster of relations depicted in (Fig.7.17.) involves meanings of PURPOSE, CAUSE and RESULT. The affinity between these three and COND has already been discussed in section 6.3. and 7.3. where it was pointed out that it is the concept of consequence where all these meanings meet. Whilst causal and conditional clauses express pre-condition for the consequence expressed in the main clause, purpose and result express the consequence itself. It has also been emphasized that in the case of apodosis and purpose clause the consequence is often of deliberate character (cf. examples 7.22., 7.23.). Interestingly, 2 of the 5 markers with scope over CAUSE and COND and 1 of the 2 that have scope over RESULT and COND are used also to express the meaning of SIOVER. Again, we may expect that here grammaticalization too took place in stages but the actual order of emergence of particular meanings remains problematic and cannot be resolved easily on the basis of synchronic evidence.

If we accept the assumption that SIOVER, as a temporal concept, is the least abstract one while COND is the most abstract, we could hypothesize that for the first group the order of grammaticalization was

SIOVER → CAUSE → COND

On the other hand, it would not be groundless to hypothesize that the marker of SIOVER could have given rise to the markers of COND and CAUSE independently.



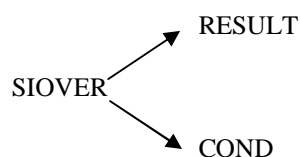
In the case of the SIOVER-RESULT-COND overlap contributed by the Yanyuwa linker *namba* (Bradley and Kirton 1992:234 and personal communication with John Bradley) the situation is also very interesting. If we proposed a SIOVER → RESULT → COND order we would automatically suggest that a marker changed its position from a marker attached to clause X to a marker of clause Y and then back to the marker of clause X which would be quite unusual – cf. the following three sentences:

(7.6a) *When you open the box you will see the spider.*

(7.6b) *You opened the box, (and) so you saw the spider*

(7.6c) *If you open the box you will see the spider.*

It would seem more convincing, therefore, to hypothesize that the marker of SIOVER developed into the markers of RESULT and COND independently.



The final group of overlaps that draws our attention in the affinity network consists of the relations of concession (CONC, ‘although’) and concessive conditionality (COCOND, ‘even if’). The first one contributes 2 and the second one 3 entries to the table of overlaps. CONC and COCOND overlap in one case with each other and the latter one occurs also in an overlap with SIOVER. None of these overlaps comes as a surprise. The relation of COCOND, which is prototypically encoded in English by a polymorphemic c-glosseme *even if*, is without any doubts cognitively related to the semantically simpler concept of conditionality (and in fact in many languages, the COND linker is, as in the English example, incorporated in the COCOND linker). While in conditional clauses the semantically independent SoA is a straightforward consequence of the SoA expressed in the semantically dependent one (cf. example 7.7a), in the case of COCOND the SoA in the semantically independent clause is commonly a counterpart of the consequence that we would normally expect (cf. 7.7b.).¹⁰

¹⁰ I argue that this is the prototypical context in which reanalysis of COND markers into COCOND markers takes place. Of course, the use of COCOND linkers goes beyond the cases of simple

(7.7a) **If** John wants me to stay I will stay.

(7.7b) **Even if** John wants me to stay I will go.

Taking into account the affinity between COCOND and COND on the one hand, and between COND and SIOVER on the other, the reasons for semantic overlap SIOVER-COCOND-COCOND found in Lavukaleve become self-evident. Example (7.8a) below shows the *-le* suffix acting as an exponent of the relations of SIOVER and COND and in (7.8b) as a COCOND c-glosseme.

(7.8a) Lavukaleve (Terill 1999:394)

Me-re-le *iire a-e-re-le* *foiga*
2PL-say-**{SIOVER}** yes 3SG.M.OBJ.-DEP-say-**{COND}** it

‘When you say [this], then if he says “Yes”, okay.’

(7.8b) (ibidem:393)

Kini koa ga ekelei siala ngo-me-le
ACT door SG.ART 3SG.OBJ-near do-EXT 2SG.HAB-**{COCOND}**

fi koa ga o-ala-re
3SG.FOC door SG.ART 3SG.SBJ-open-FUT

‘Even if you were not close to the door yet, the door would open.’

Finally, to account for the presence of the relation of concession (prototypically expressed in English by *although*) in CONC-COND and CONC-COND-COCOND overlaps we shall once again emphasize the close affinity between conditionality and causality. As König (1991) has argued, concessive subordination is the dual counterpart of causal subordination. Moreover, similarly as in the case of COCOND-COCOND overlaps also for conditionality/causality and concession the difference can be brought down to the counterexpected consequence:

(7.9a) **Because** Lucy likes apples she will buy some.

(7.9b) **If** Lucy likes apples she will buy some.

(7.9c) **Although** Lucy likes apples she will buy pears.

The counterexpected consequence appears, therefore, to be the simplest element linking the concepts of concession and concessive conditionality with each other as well as with conditionality (and causality). When it comes to hypothesizing about a specific grammaticalization pathway in the evolution of conditional and concessive meanings, Hopper and Traugott (1993:180) proposed a COND → CONC route but, as Heine and Kuteva (2002:93) notice, more data from language families other than Indo-European

counterexpected consequences – cf. a sentence like *Even if you sell the two cars, you will still have the other three*. In such cases, the reading appears to be more of a concessive than conditional character. Since, however, neither COCOND nor CONC are the focus of this thesis, I do not attempt to pursue this discussion here.

are required to substantiate this hypothesis. In the case of the markers referred to in this thesis more detailed diachronic investigation would be required to confirm the existence of this pathway.

The remaining relations depicted in the network: comparison, manner, similarity, simultaneity co-extensiveness and terminus a quo contribute only a small number of overlaps and so are not elaborated on here. The only observation I shall mention is that 5 of the total number of 8 occurrences of these relations are co-occurrences with SIOVER which, again, clearly suggests that what we deal here with are effects of a multi-stage reanalysis and grammaticalization.

Gathering all the observations on semantic polyfunctionality of conditionality linkers we can conclude that:

- a) the relation of simultaneity overlap (SIOVER) is by far the most common overlap of conditionality which can be easily explained by the temporal connotations of any type of conditionals, and especially the “real conditionals” which have been the subject of analysis here;
- b) the origin of the other overlaps depicted in (Fig.7.17.) can be, in the majority of cases, explained by the fact that the relations share certain temporal, causal and consequential semantic characteristics;
- c) in many cases it is the relation of simultaneity overlap (SIOVER) that seems to be the most likely candidate for the concept through which other circumstantial relation entered the space of semantic affinities of conditionality;
- d) it is clear that among the types of markers discussed here it is the temporal markers (especially SIOVER) that are common sources of conditionality linkers (this fact has been depicted in Fig.7.9. and Fig.7.11.), there is also some evidence (although not as prominent) for causality c-glossemes developing into conditional markers.

Concluding this section, let us look at the strongest semantic affinities of conditionality linkers as described in Kortmann (1997):

	I	II	III	IV	V	VI	VII
COND (131 items)	CONTIN (38.2%)	SIOVER (31.3%)	CAUSE (26.0%)	ANTE (17.6%)	IMMANTE (16.0%)	SIDUR (16.0%)	SICOEX (15.3%)

(Fig.7.18.) The strongest semantic affinities of polyfunctional adverbial subordinators expressing conditionality according to Kortmann (1997:198)

Kortmann has found that the dominating relations in the network of affinities of conditionality are contingency (‘whenever’) and simultaneity overlap (‘when’). They

are followed by the concept of cause and four temporal concepts: anteriority ('after'), immediate anteriority ('as soon as'), simultaneity overlap ('while') and simultaneity co-extensiveness ('as long as'). In my database the relation of contingency has not appeared on its own. It is possible that some of the 'when' linkers mentioned here may occur in contexts in which English would use 'whenever' but the overlap of forms of linkers expressing contingency and conditionality has not been directly reported in the materials I analysed or by any of the consultants. This concerns also the distinction between anteriority and immediate anteriority – the latter, as already mentioned in section 4.3. has not been reported.

One point in which both Kortmann's and my result completely agree is the high position of SIOVER. Although in my investigation markers of ANTE, CAUSE, SIDUR and SICOEX have been also identified as overlapping in form with COND, their frequency is low both in the absolute number of overlaps and in binary overlaps. Similarly low is also the frequency of PURPOSE, COCOND and MANNER linkers (which have not surfaced in Kortmann's summary at all). I would incline towards looking for the explanation of these discrepancies in the genetic/geographic bias in the design of Kortmann's study.

ABSOLUTE NUMBER OF OCCURRENCES (65 overlaps)							
I	II	III	IV	IV	IV	IV	
SIOVER 32 items (49.2%)	ANTE 7 items (10.8%)	CAUSE 4 items (6.2%)	SIDUR 3 items (4.6%)	COCOND 3 items (4.6%)	PURPOSE 3 items (4.6%)	MANNER 3 items (4.6%)	
OCCURRENCES IN BINARY OVERLAPS (26 overlaps)							
I	II	III	IV	IV	IV	IV	IV
SIOVER 17 items (65.4%)	SICOEX 2 items (7.75%)	PURPOSE 2 items (7.75%)	ANTE 1 item (3.8%)	CAUSE 1 item (3.8%)	CONC 1 item (3.8%)	COCOND 1 item (3.8%)	MANNER 1 item (3.8%)

(Fig.7.19.) The strongest semantic affinities of polyfunctional c-glossems of conditionality revealed in this study

7.4. SUMMARY

The concept of conditionality is, without any doubt, a complex one. In this thesis only one of its subtypes – the real conditionals – has been scrutinized. The analysis presented in this section not only confirms many of the findings reported in earlier works but adds also a couple of elements to our understanding of the origin and functioning of markers of protasis.

I argued here that the synchronic, cross-linguistic patterns of polysemy and the material incorporated in morphologically complex conditionality markers, suggest that in many cases conditionality c-glosseemes developed directly from markers of temporal interclausal relations. This seems especially striking in those cases where overlaps of conditionality linkers with adverbs, adpositions and case markers have been identified. The analyses confirmed also what has been argued for in works of other linguists – the affinity between topic markers and conditionals or certain verbs and conditionals or even complementizer and markers of protasis.

The prevalence of temporal concepts emerged also from the analysis of patterns of semantic polyfunctionality of markers of protasis. The relation of simultaneity overlap ('when') is without any doubt the one that dominates in the space of affinities of conditionality although other concepts – such as anteriority, cause or even purpose and concession – also marked their place in the network.

We should recall here also that, as it was the case with the three circumstantial relations discussed before, the investigation revealed existence of quite strong *morphological complexity* → *monofunctionality* tendency at both syntactic and semantic level. When it come to the form of conditionality c-glosseemes the analysis revealed dominating proportion of free words (which outnumber affixes almost 3 times) and very high proportion of distributed markers.

CHAPTER 8

Comparisons and conclusions

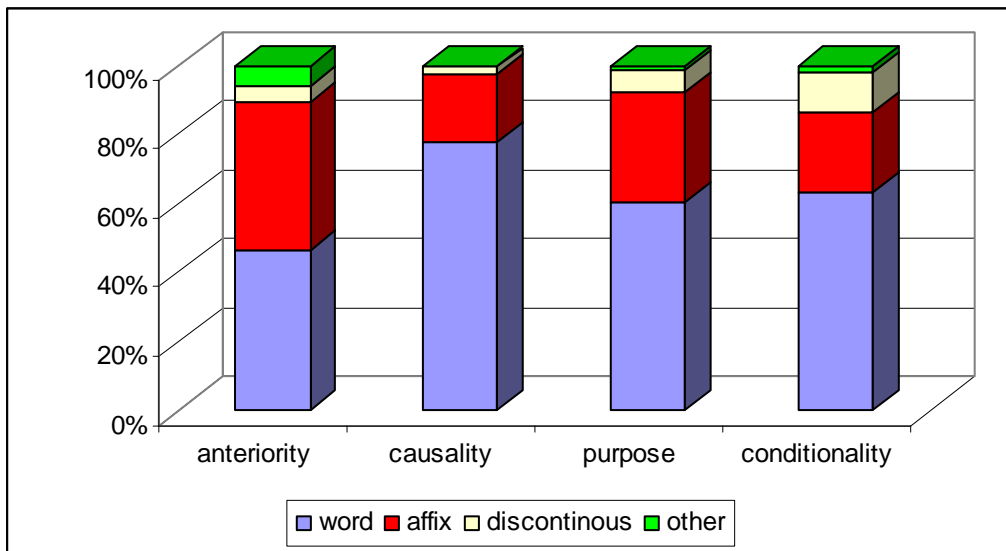
In the four preceding chapters detailed discussion of the form, complexity and origin of c-glossemes of anteriority, causality, purpose and conditionality has been presented together with an insight into the semantic space of circumstantial relations in which these four concepts occupy an important position. The aim of this chapter is to compare the results obtained in chapters 4, 5, 6 and 7 and draw conclusions from the comparisons.

8.1. FORMS OF C-GLOSSEMES

The first major element that has been looked at in each of the analytical chapters is the form of c-glossemes. Continuous linkers classified as words have been distinguished from affixes, distributed markers, clitics and combinations of words and affixes. Recall that for each of the four relations it has been found that the c-glossemes encoding them are only rarely clitics and combinations. For this reason the two groups have been treated jointly and labelled ‘others’ in the summary of results presented in (Fig.8.1.) below and in the diagram depicting the findings graphically (Fig.8.2).

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
words	70	46.67%	158	77.83%	95	60.50%	108	63.16%
affixes	64	42.67%	40	19.70%	50	31.84%	40	23.40%
discontinuous markers	7	4.66%	4	1.97%	10	6.36%	20	11.69%
others	9	6.00%	1	0.50%	2	1.30%	3	1.75%
TOTAL	150	100%	203	100%	157	100%	171	100%

(Fig.8.1.) Comparison of the forms of c-glossemes



(Fig.8.2.) Proportions of the particular forms of c-glossems

One of the elements that we immediately notice when we look at the summaries is that among the four relations causality has the highest proportion of lexical c-glossems (almost 80%) and lowest proportion of affixal linkers (below 20%). The ratio words-to-affixes is also quite high for conditionality and purpose – in both cases over 60% of the linkers are words. For anteriority, on the other hand, the number of affixes and words is almost equal, while the remaining forms constitute just over 10% of the markers. Anteriority is also the relation with the highest percentage of clitics and combinations of words and affixes – 6%. These two groups contribute less than 2% in the case of conditionality and purpose and merely 0.5% in the case of causality. Another interesting observation is the distribution of discontinuous markers across the four relations – while they are quite rare in the case of causality and slightly more common in anteriority and purpose, they constitute over 10% of the c-glossems of conditionality.

Although the sample I have used is far too small to make any significant inferences about the tendencies of encoding of circumstantial relations in particular language families, there are several elements that are worth emphasizing. First of all it is clear from the summary presented in (Fig.8.3.) that while for the Amerind, Indo-Pacific and Australian languages affixal c-glossems are very common, their frequency and significance is far smaller in African or Austric languages (and they are completely absent from the two creole languages in the sample: Ndyuka and Kriyol). This can be easily explained by the typological characteristic of these languages. As Schachter and Shopen concisely put it:

Closed word classes tend to play a more prominent role in analytic languages than they do in synthetic languages. This is because much of the semantic and syntactic work done

by the members of closed word classes in analytic languages is done instead by affixes in synthetic languages. (2007:23)

In other words, where morphology is scant one should expect functional items (including clause linkers) to be lexicalized.

The second interesting observation concerns discontinuous markers which appear to be especially common in Niger-Kordofanian languages and so this may suggest a genetic pattern. The discontinuous linkers occur also in a couple of unrelated languages in Asia but it is unlikely that their presence is an effect of areal diffusion since the languages are geographically far too distant from each other.

language	anteriority	causality	purpose	conditionality
Uralic-Yukaghir languages				
Estonian	w A	W	W	W
Indo-Hittite languages				
English	W A	W a	W disc	W
Polish	W a	W a	W	W
Hindi	W a	W	W A	W
Niger-Kordofanian languages				
Krongo	W	W A	A	A DISC
Boko	A	W disc	W DISC	W
Akan	DISC	w DISC	W	W DISC
Swahili	W	W	W	W A DISC
Sango	-	W	W	W
Suppyire	-	W	w DISC	W
Mayogo	W	W	W	W DISC
Gola	W	W	W	W
Afro-Asiatic languages				
Konso	W DISC CLIT	W	W	W
Shelha	W	W	W	-
Arabic	W	W	W	W
Hausa	W	W	W	W
Maale	w A	A	w A	A
Nilo-Saharan languages				
Kanuri	W DISC	A	A	A
Lango	W	W	W	W
Ma'di	-	W	W	W
Didinga	-	W	W	W
Fur	W	W	W	W
Khoisan languages				
Khwe	W	W	W	W
Amerind languages				
SE Tepehuan	W	W	W	W
Hualapai	A	A	A	A
Seri	A	W A	-	A
Nez Perce	A	W	A	W
Nisga'a	W	W	W	W
Lillooet	CLIT	W	W	CLIT
Ika	-	A	A	A
Rama	A	W	A	A
Cubeo	A COMB	W	A	A
Retuarã	A	A	A	W
Achagua	-	W A	A	A
Baure	-	W	W	W

Apurina	-	W	W	-
Quechua Huallaga	A	A	A	A
Mocovi	-	W	-	W
Macushi	W	W	A	A
Na-Dene languages				
Apache Jicarilla	-	W A	W A	A
Eskimo-Aleut languages				
Yup'ik	A	A	A	A
Australian languages				
Yanyuwa	W	W	DISC A	W
Wambaya	A	-	A	-
Jingulu	-	-	W	W
Yindjibarndi	-	A	-	CLIT
Pitjantjatjara	A	A	A	A
Warlpiri	A	W	W A	W
Arabana	-	-	A	-
Indo-Pacific languages				
I'saka	-	-	-	-
Hatam	-	W	W	W
Lavukaleve	A	W	A	A
Yimas	A	A	-	A
Eipo	W A	W	W a	W
Ama	CLIT	-	-	CLIT
Meyah	-	W	W	W
Au	W	W	W	A
Austic languages				
Sapuan	-	W	W	W
Jahai	W	W	-	W
Seediq	-	W	W	W
Leti	W	W	W	W
Santali	A	W A	W	W disc
Taba	-	W	W	W
Yami	-	W	W	W
Ilokano	W	W	W	W
Batak Karo	W	W	W	W
Thai	W	W	W	W disc
Mantauran Rukai	W A	W	W	W
Paiwan	W	W	W	W
Vitu	-	W	W	W
Altaic languages				
Japanese	w A	W a	W a	A
Dagur	A	W	W A	DISC
Sino-Tibetan languages				
Mandarin	W	W	W	W
Lepcha	A	A	W	W
Kayah Li	W	W	W	W
Galo	w A	W a	W	W A
Elamo-Dravidian languages				
Tamil	W a	w A	W a	A
Caucasian languages				
Lezgian	w A	W a	W a	A DISC
Chukchi-Kamchatkan languages				
Chukchi	A	W A	-	W
Isolated languages				
Nivkh	A	A	A	A
Burushaski	W A DISC	W	W a	W
Basque	W A	W comb a	A	A comb
Ket	W	w A	W A DISC	W A DISC
Creole languages				

Ndyuka	W	W	W	W
Kryiol	W	W	W	W

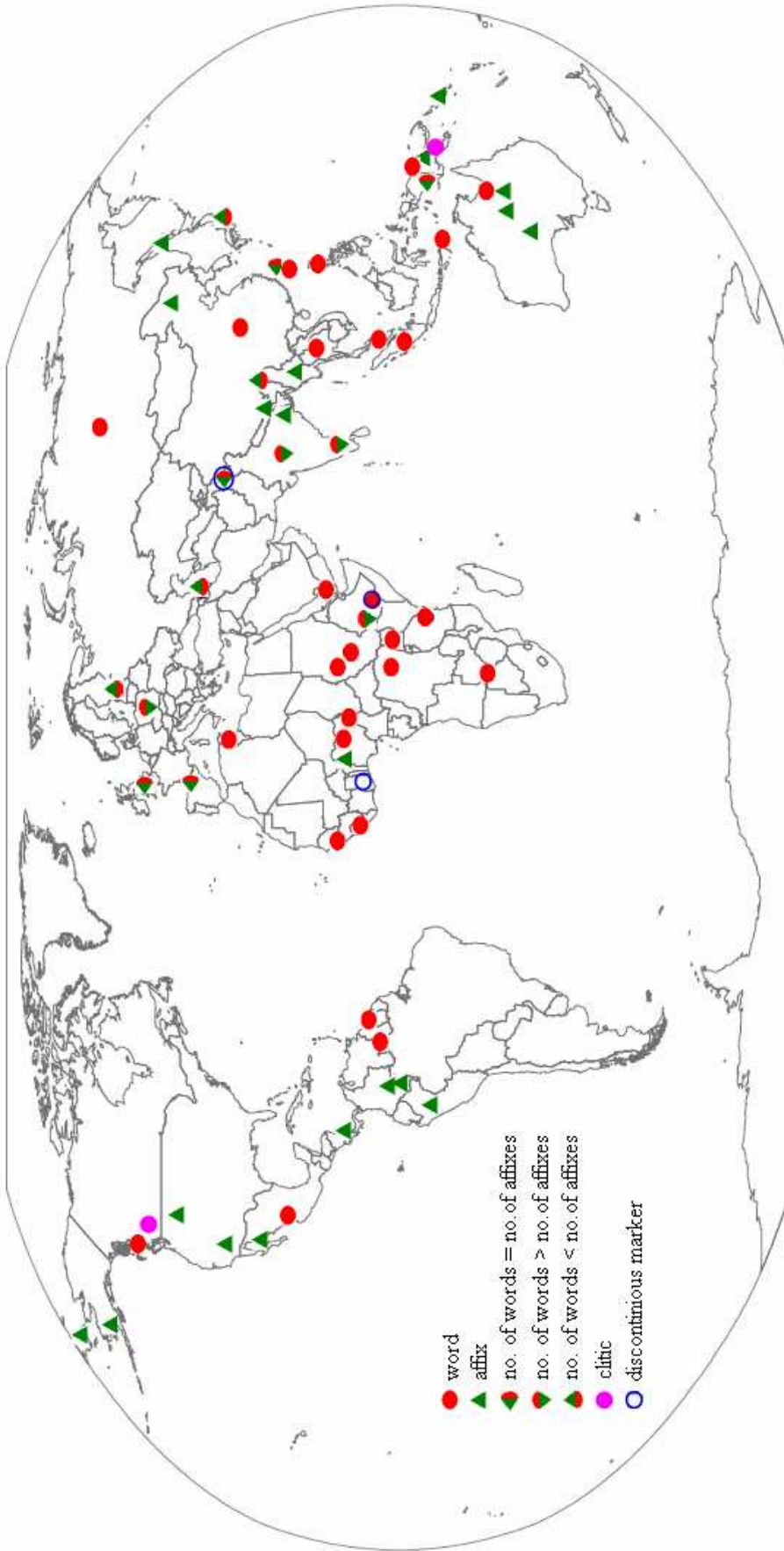
(Fig.8.3.) Distribution of forms of c-glossemes

The abbreviations stand for: W - words; A - , affixes, CLIT – clitics, DISC - discontinuous markers, COMB – combinations of words and affixes. If a particular abbreviation is put in small letters it means that the form it refers to is in a particular language used rarer than the one that has been put in capitals. If more than one abbreviation appears in capitals, it indicates the forms are of equal frequency.

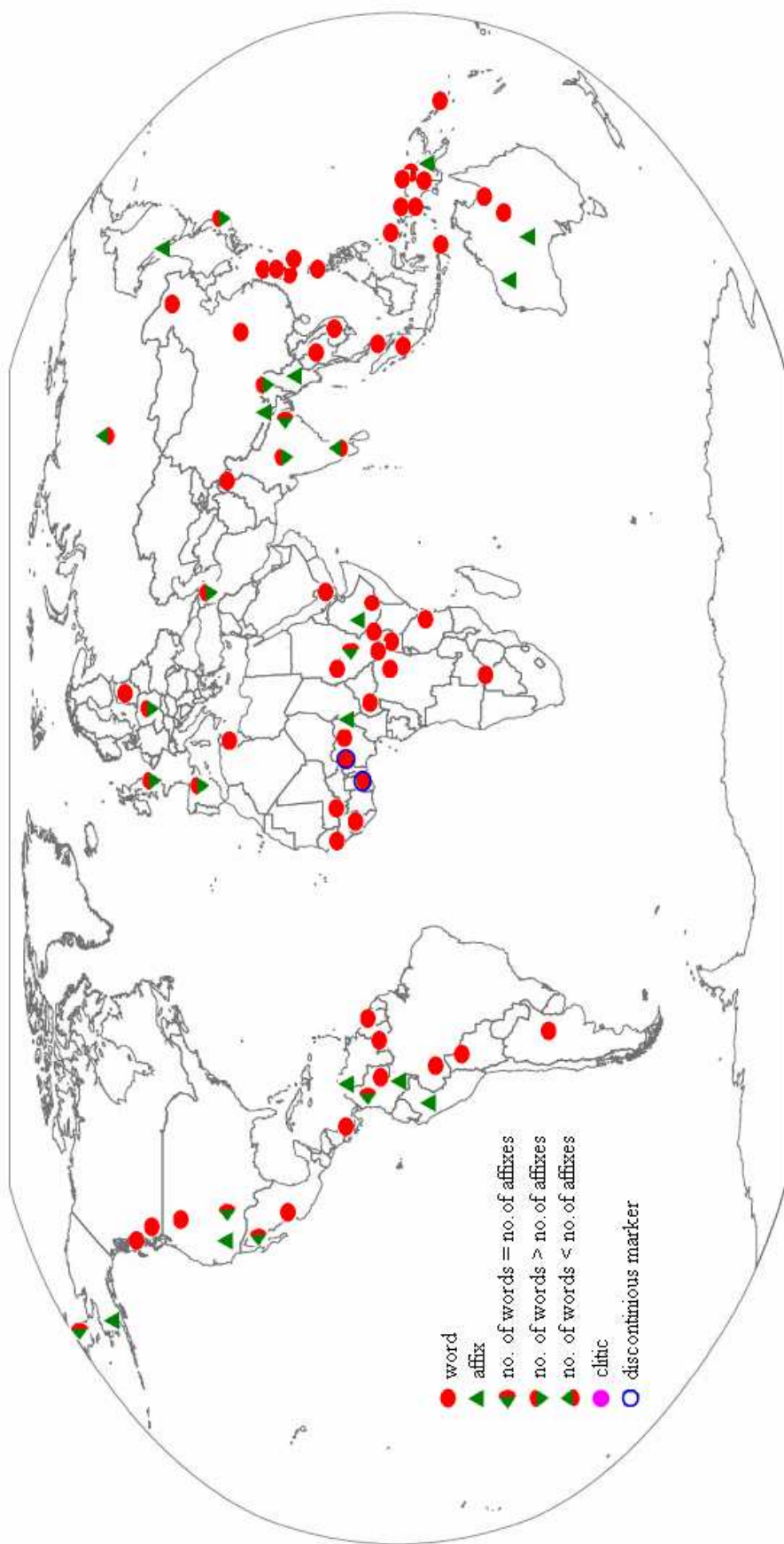
The maps presented in (Fig.8.4.-8.7.) help us to discover several other interesting facts about the distribution of the particular forms of c-glossemes in each of the four relations.¹ On each of the maps we noticed several clusters of values, the most significant of which are the aforementioned affixes in North and Central South America and Australia on the one hand, and the mixture of words and affixes in Europe, India and along the Indian border on the other. The presence of certain preferences concerning the types of clause linkers has been already noticed by Kortmann (1997) and Bisang (1998). The study conducted by the first author has revealed that the languages on the westernmost and easternmost parts of Europe use converbs as a primary method of clause combining which distinguishes them from the languages of the middle. Bisang, going outside Europe, has distinguished between three types: European, in which the number of converbs is limited but there is a richness of adverbial subordinators; Eurasian which is characterized by a wealth of converbs and relatively small number of free adverbial subordinators; and Far East type in which converbs occur only marginally and free subordinators serve clause linking functions. I do not go as far as distinguishing between subordinators and converbs but the point I am making here, is that also with respect to the more general distinctions between various types of c-glossemes such as the ones considered here, there are some geographic preferences.

Interestingly, the preferences differ across certain groups of languages in a rather random way. For instance, the number of words encoding the relation of anteriority in India and along the Indian border is lower than for any other of the three relations. In North America, on the other hand it is not only anteriority but also purpose that in many languages is encoded exclusively by affixes which puts these two relations in direct contrast with the forms of linkers of causality and conditionality in those languages. In languages such as Polish, English and Estonian the clause linkers of purpose and conditionality are exclusively lexical, while for the other two relations the languages use a mixture of words and affixes.

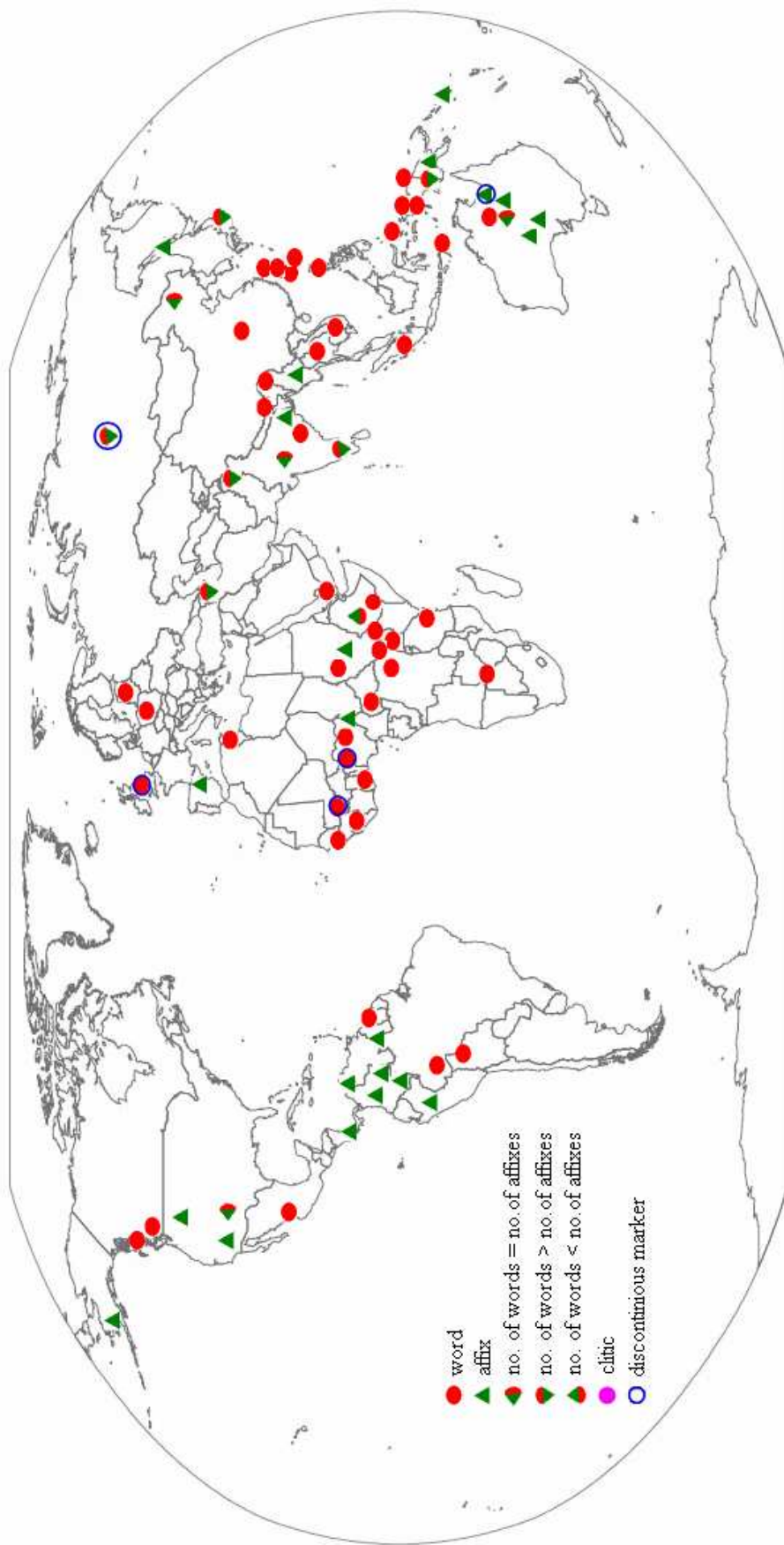
¹ Combinations of words and affixes, due to their low frequency, have been omitted from the maps.



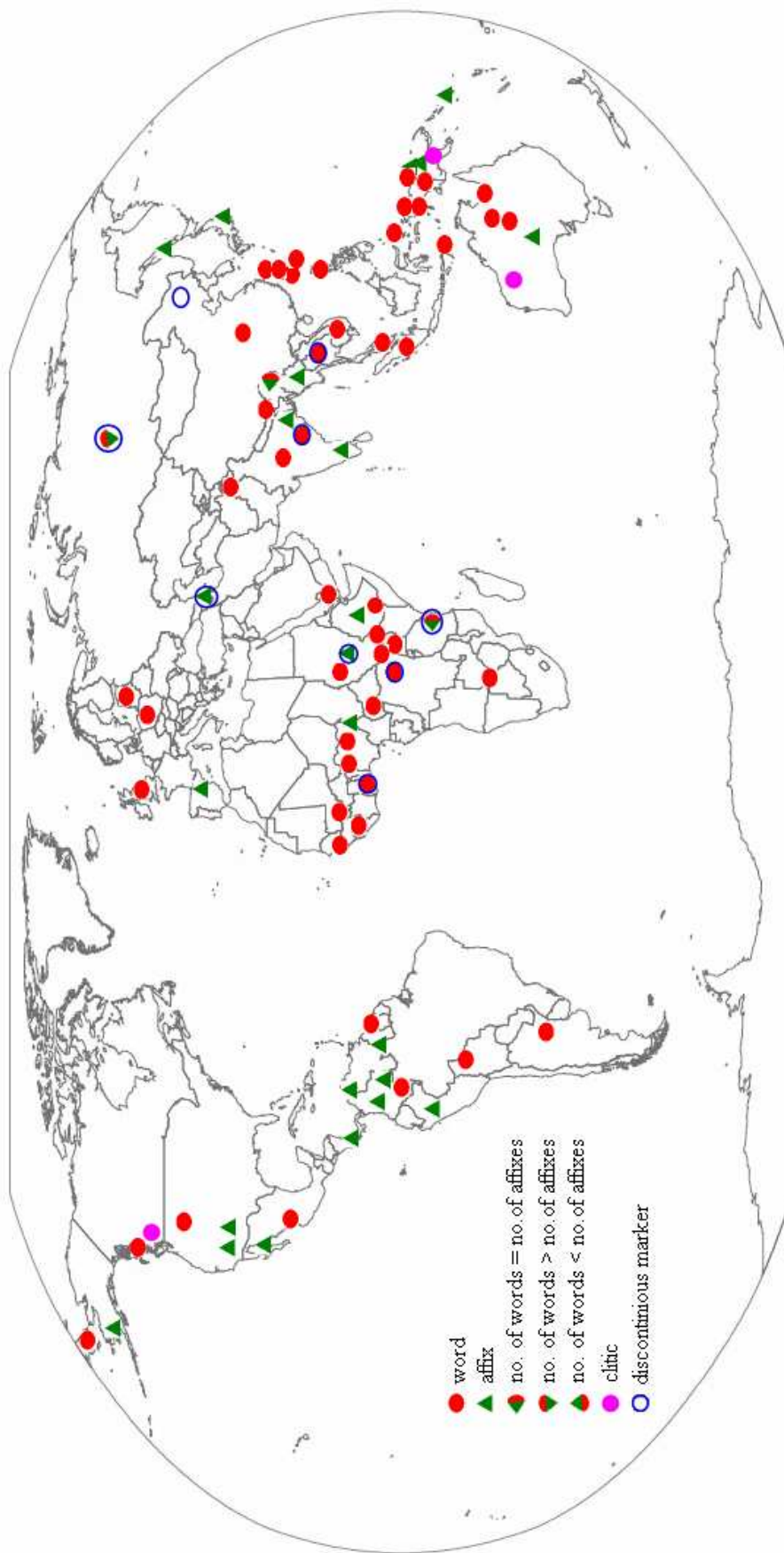
(Fig. 8.4.) Forms of anteriority c-glossemes – geographic distribution



(Fig 8.5.) Forms of causality c-glossems – geographic distribution



(Fig. 8.6.) Forms of purpose c-glossemes – geographic distribution



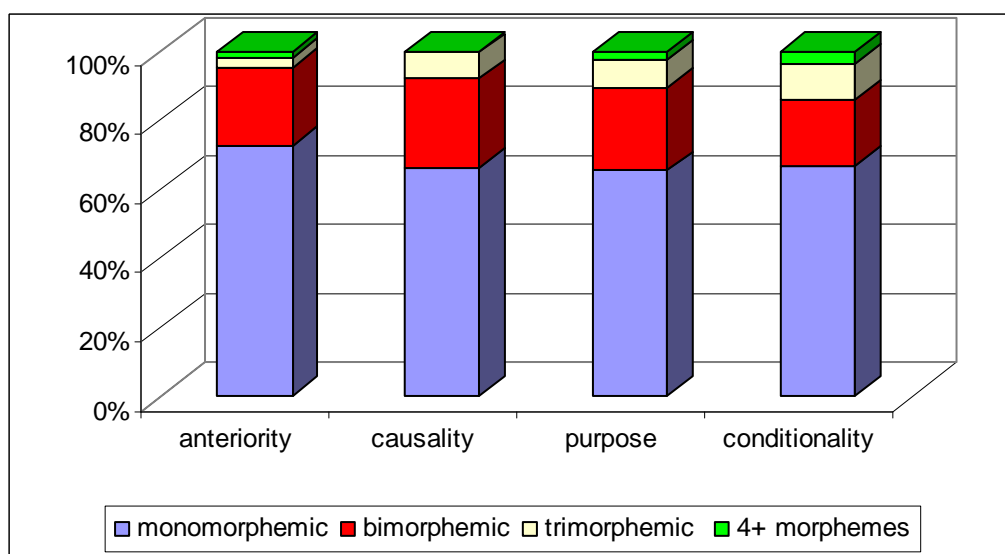
(Fig.8.7.) Forms of conditionality c-glossemes – geographic distribution

8.2. DEGREES OF MORPHOLOGICAL COMPLEXITY

Apart from distinguishing between various morphological forms of c-glossemes a distinction between synchronically monomorphemic and polymorphemic markers has also been introduced in the analytical sections. The more complex markers have been further divided into bimorphemic, trimorphemic and those consisting of 4 and more morphemes. The results of the analysis are presented jointly in table (Fig.8.8.) and diagram (Fig.8.9.) below.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
monomorphemic	105	72.41%	123	66.13%	100	65.36%	109	66.87%
bimorphemic	33	22.76%	48	25.81%	37	24.18%	31	19.02%
trimorphemic	4	2.76%	14	7.53%	12	7.84%	17	10.43%
4+ morphemes	3	2.07%	1	0.53%	4	2.62%	6	3.68%
TOTAL	145	100%	186	100%	153	100%	163	100%

(Fig.8.8.) Comparison of the internal complexity of c-glossemes



(Fig.8.9.) Variation in internal complexity of c-glossemes

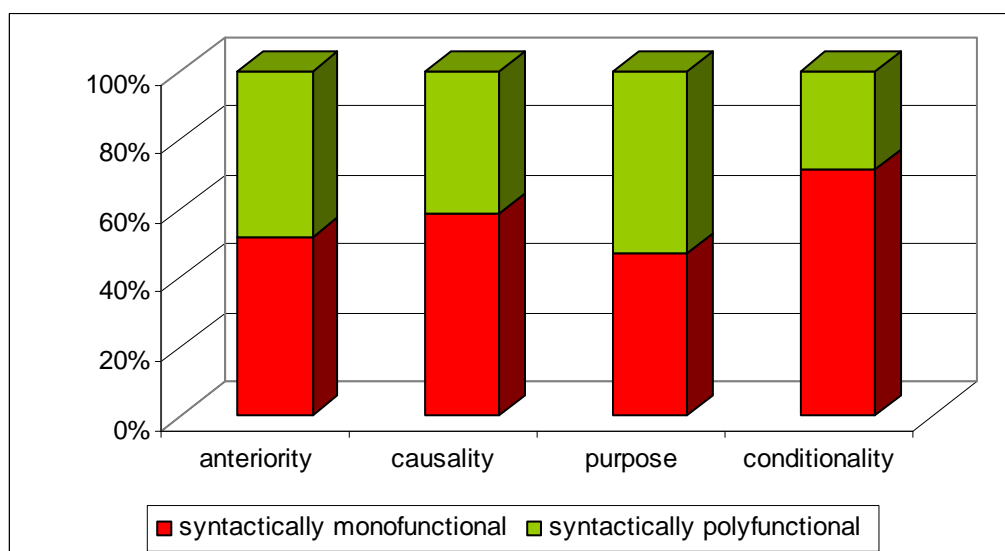
The differences in proportions, especially when it comes to monomorphemic and bimorphemic markers, are very small. The highest percentage of monomorphemic markers, over 72%, has been observed for anteriority but it is only 7 points higher than for the relation with the smallest percentage of these linkers – purpose. The gap between the highest and lowest percentage of bimorphemic markers is even smaller – showing only 6 points difference.

8.3. DEGREES OF SYNTACTIC POLYFUNCTIONALITY

In the material collected for the purpose of this thesis many examples of syntactically polyfunctional markers have been identified. As the two figures below (Fig.8.10. and 8.11.) illustrate, the percentage of polyfunctional markers is highest for purpose and anteriority, slightly lower for causality and considerably lower for conditionality.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
syntactically monofunctional	73	51.77%	108	58.70%	71	47.01%	114	71.25%
syntactically polyfunctional	86	48.23%	76	41.30%	80	52.99%	46	28.75%
TOTAL	141	100%	184	100%	151	100%	160	100%

(Fig.8.10.) Comparison of the numbers of syntactically monofunctional and syntactically polyfunctional c-glossemes



(Fig.8.11.) Distribution of syntactically monofunctional and syntactically polyfunctional c-glossemes

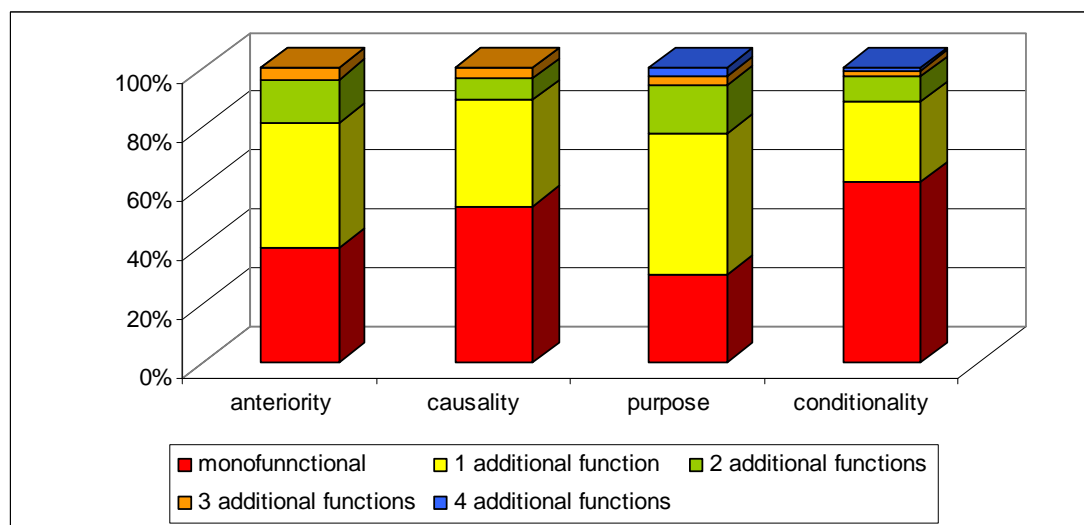
Within the group of syntactically polyfunctional markers I have introduced a distinction between degrees of polyfunctionality i.e. the number of additional syntactic functions served by a given c-glosseme. While discussing the degrees an additional distinction between monomorphemic and polymorphemic markers has also been made. I shall keep it here and start from the comparative overview of the degree of syntactic polyfunctionality in monomorphemic c-glossemes.

The numbers presented in (Fig.8.12.) leave no doubts that among the monomorphemic markers it is conditionality and causality that have the highest

proportion of monofunctional markers – 61.32% and 52.99% respectively. Interestingly, for both of these relations almost 90% of the monomorphemic markers are those with up to 1 additional function (red and yellow bars in the diagram in Fig.8.13). The same holds for about 80% of anteriority and purpose linkers. Of the four relations it is purpose that has the smallest number of monomorphemic and syntactically monofunctional markers and, at the same time, the only one for which more than 1 example of a marker with 4 additional functions has been found.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
monofunctional	39	38.61%	62	52.99%	29	29.59%	65	61.32%
1 additional function	43	42.58%	42	35.90%	47	47.96%	29	27.36%
2 additional functions	15	14.85%	9	7.69%	16	16.33%	9	8.49%
3 additional functions	4	3.96%	4	3.42%	3	3.06%	2	1.89%
4 additional functions	-	-	-	-	3	3.06%	1	0.94%
TOTAL	101	100%	117	100%	98	100%	106	100%

(Fig.8.12.) The degree of syntactic polyfunctionality among monomorphemic c-glossemes



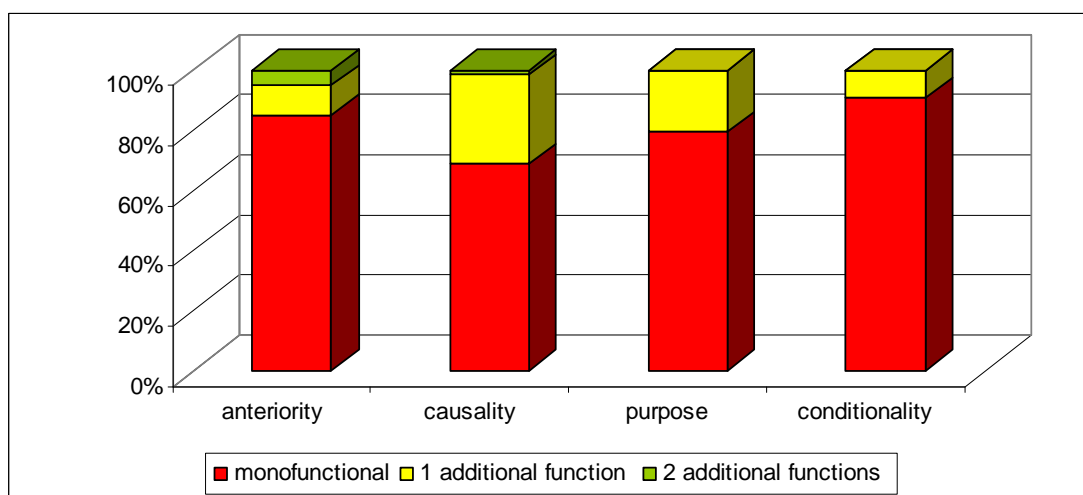
(Fig.8.13.) Distribution of degrees of syntactic polyfunctionality among monomorphemic c-glossemes

The situation among polymorphemic markers looks quite different to that presented for monomorphemic c-glossemes. If we compare the data in (Fig.8.12) with those in (Fig.8.14) we notice that the only similarity is that in both cases conditionality is the relation with the highest proportion of monofunctional markers. Among polymorphemic markers it is not anteriority, and not purpose but causality, which –

with its 20 markers that can be used in one additional syntactic functions – has the smallest number of monofunctional linkers. In fact, anteriority, with only 15% of syntactically polyfunctional markers, stands right behind conditionality.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
monofunctional	34	85.00%	46	68.66%	42	79.25%	49	90.70%
1 additional function	4	10.00%	20	29.85%	11	20.75%	5	9.30%
2 additional functions	2	5.00%	1	1.49%	-	-	-	-
TOTAL	40	100%	67	100%	53	100%	54	100%

(Fig.8.14.) The degree of syntactic polyfunctionality among polymorphemic c-glossemes



(Fig.8.15.) Distribution of degrees of syntactic polyfunctionality among polymorphemic c-glossemes

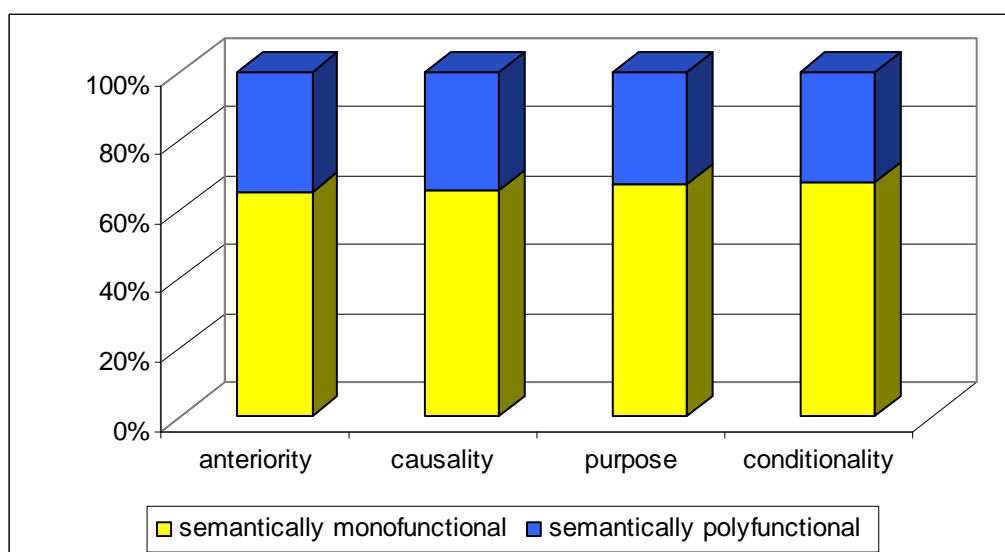
The data once again draw our attention to the complexity → (syntactic) monofunctionality tendency – the more morphologically complex markers tend to have lower degree of polyfunctionality. This tendency has been observed for all the four relations. Nonetheless, when it comes to drawing general conclusions from what has been presented above, we are entailed to conclude that conditionality reveals highest and purpose lowest proportion of monofunctional markers with anteriority and causality in the middle of the scale.

8.4. DEGREES OF SEMANTIC POLYFUNCTIONALITY

The discrepancies between the proportion of semantically monofunctional and polyfunctional markers in the four analysed (see Fig.8.16. and Fig.8.17.) are much smaller than for the syntactic polyfunctionality relations discussed in the previous section. In fact, between the relation with the lowest percentage of semantically polyfunctional linkers (conditionality) and the relation with the highest score (purpose) there is only 3% difference.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
semantically monofunctional	93	65.04%	121	65.76%	101	67.30%	111	68.10%
semantically polyfunctional	50	34.96%	63	34.24%	49	32.67%	52	31.90%
TOTAL	143	100%	184	100%	150	100%	163	100%

(Fig.8.16.) Comparison of the numbers of semantically monofunctional and semantically polyfunctional c-glossemes



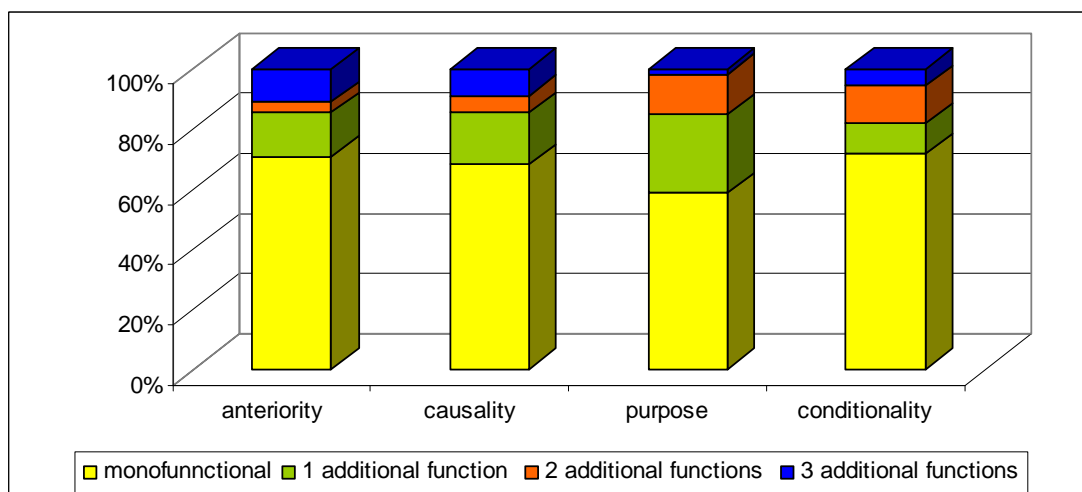
(Fig.8.17.) Distribution of semantically monofunctional and semantically polyfunctional c-glossemes

The discrepancies become slightly more salient if we exclude from the summaries the general c-glossemes. Without them among the monomorphemic markers it is purpose that has the lowest number of semantically monofunctional linkers (58.65%). It is followed by causality which has 10% more of monofunctional markers and then closely also by anteriority and conditionality. As for the degree of polyfunctionality, it turns out that anteriority has the highest and purpose the lowest

number of linkers with 3 additional semantic functions. The details are presented in (Fig.8.18.) and (Fig.8.19.).

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
monofunctional	63	70.79%	75	68.18%	61	58.5%	63	71.59%
1 additional function	13	14.61%	19	17.27%	27	25.96%	9	10.23%
2 additional functions	3	3.37%	6	5.45%	14	13.46%	11	12.50%
3 additional functions	10	11.23%	10	9.10%	2	1.93%	5	5.68%
TOTAL	89	100%	110	100%	104	100%	88	100%

(Fig.8.18.) The degree of syntactic polyfunctionality among monomorphemic c-glossemes

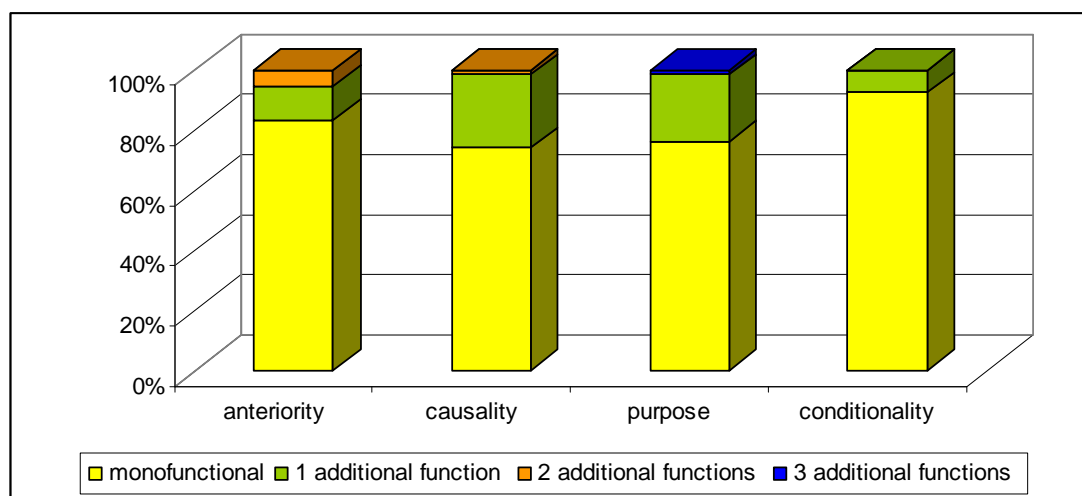


(Fig.8.19.) Distribution of degrees of semantic polyfunctionality among monomorphemic c-glossemes

The comparison of degrees of semantic polyfunctionality in polymorphemic markers (again, with exclusion of general c-glossemes) also reveals that conditionality is the relation with the lowest percentage of polyfunctional markers (see Fig.8.20. and 8.21.). Only 7.4% of the conditionality linkers serve one additional semantic functions and there are no instances of linkers with higher degree of polyfunctionality for this relation. In the case of causality and purpose linkers, on the other hand, over 20% of them serve 1 additional functions and for both single examples of linkers with more than 2 semantic functions have also been elicited. This puts them behind anteriority for which over 83% of the polymorphemic markers are monofunctional.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
monofunctional	30	83.33%	46	74.19%	48	76.19%	50	92.60%
1 additional function	4	11.11%	15	24.20%	14	22.22%	4	7.40%
2 additional functions	2	5.56%	1	1.61%	-	-	-	-
3 additional functions	-	-	-	-	1	1.59%	-	-
TOTAL	36	100%	62	100%	63	100%	54	100%

(Fig.8.20.) The degree of syntactic polyfunctionality among polymorphemic c-glossemes



(Fig.8.21.) Distribution of degrees of semantic polyfunctionality among polymorphemic c-glossemes

On the whole, we may conclude from the comparisons presented here that the clause linkers of conditionality and causality are more often lexicalized and semantically and syntactically monofunctional than the linkers of purpose and anteriority but all these four groups reveal a similar level of morphological complexity. The usefulness of the parameters compared here goes, however, far beyond such simple comparison. They are very useful tools in assessing a degree of grammaticalization. If we want to find the most highly grammaticalized item in a group of clause linkers (or in any other lexical or grammatical group of items) we should look for the one that is morphologically most simple and which has the lowest degree of semantic and syntactic polyfunctionality. If we want to assess the degree of grammaticalization of a concept in comparison to another concept in a given language we have to find out for which of the concepts a more highly grammaticalized exponent is available. And finally, if we are interested in finding out which of two (or more concepts) reveals a higher degree of grammaticalization in a cross-linguistic perspective, for each of these concepts we have to compare how many of the languages possess the most highly

grammaticalized items to encode them. Adding to these three parameters information about the morphosyntactic form of the item(s) we may, in a similar manner, assess the degree of lexicalization. It is these very questions – of cross-linguistic tendencies to grammaticalize and lexicalize certain concepts to a higher extent than others that are the most interesting ones. To answer them we need to refer to additional set of data – a set which is highly relevant for the problems considered in part three of the thesis. Thus, I postpone the discussion until chapter 9.

8.5. SOURCES OF C-GLOSSEMES

The present study, contributing to other works in the domain of grammaticalization, has yielded some interesting results that bear on our understanding of how clause linkers of anteriority, causality, purpose and conditionality came into being. In this section of the chapter I endeavour to put the results together and refer them to the more general findings concerning the sources of clause linkers reported in the linguistic literature which, recall, include adpositions, case markers, adverbs, complementizers, relativizers, verbs and nouns.

8.5.1. Evidence from the observation of patterns of polysemy

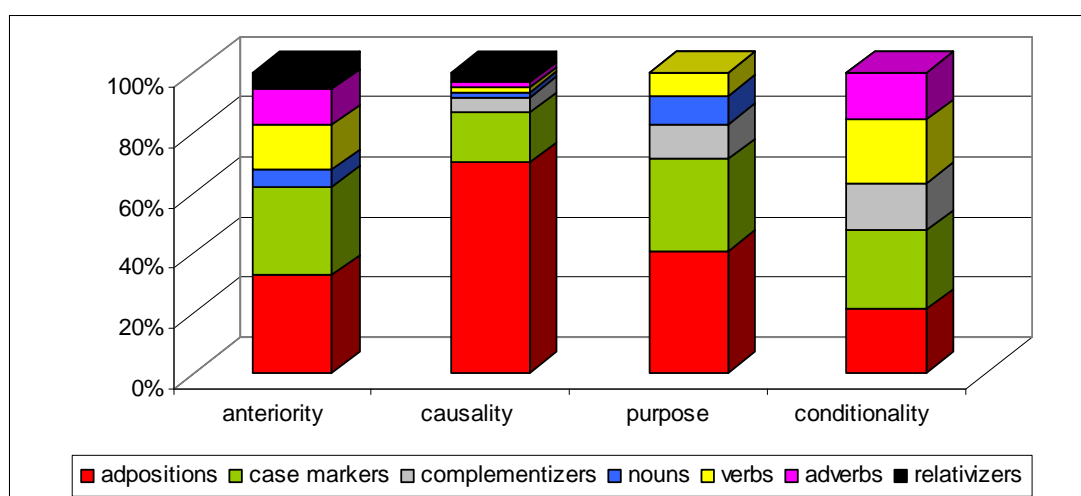
The first method of the insight into the origin of c-glossemes applied in this thesis has been the analysis of patterns of polysemy. In the tables presenting the results in each of the sections polysemes with members in just two syntactic categories (i.e. cases of binary overlaps) have been distinguished from those which have their members in three and more categories (i.e. cases of multiple overlaps). It has been assumed (cf. section 3.2.9.) that the first group of polysemes give us more reliable evidence for the origin of the markers since it eliminates confusion that the reconstruction of grammaticalization pathways in cases of multiple overlaps brings about. I intend to keep the binary overlaps separately here and I start the comparative overview from discussion on the distribution of the polysemes in the categories considered by other researchers to be the most common sources of clause linkers.

As (Fig.8.22.) and (Fig.8.23.) below illustrate, the range of sources of clause linkers, reconstructed on the basis of evidence provided by binary overlaps, confirms the important role of adpositions and case markers in the formation of clause-linking markers. It does not escape our attention, however, that there are rather striking differences between the four relations analysed. While adpositions contribute 70% of the binary overlaps for causality the same holds only for just over 20% for

conditionality. The proportion of case markers, on the other hand is much lower for causality than for any other relation. Complementizers do not occur at all in binary overlaps with anteriority but constitute 11.54% and 15.79% of polysemes of purpose and conditionality respectively. Adverbs, on the other hand, occupy a high position in the case of anteriority and conditionality but are only marginal as sources of causality, and are completely absent in the group of binary overlaps of purpose. It also becomes obvious that verbs are more likely to be sources of conditionality and anteriority than of purpose and causality. Among the markers of the latter relation we very rarely find overlaps with nouns and we do not find them at all in conditionality linkers. Nouns are not too common as binary polysemes of anteriority either but they appear to be more frequent among polysemes of purpose markers. Finally, for all of the analysed relations relativizers acting as (potential) sources of clause linkers are less frequent than any other syntactic category.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
adpositions	11	32.35%	42	70.00%	21	40.38%	4	21.05%
case markers	10	29.41%	10	16.67%	16	30.76%	5	26.32%
complementizers	-	-	3	5.00%	6	11.54%	3	15.79%
nouns	2	5.88%	1	1.67%	5	9.62%	-	-
verbs	5	14.71%	1	1.67%	4	7.70%	4	21.05%
adverbs	4	11.77%	1	1.67%	-	-	3	15.79%
relativizers	2	5.88%	2	3.33%	-	-	-	-
TOTAL	34	100%	60	100%	52	100%	19	100%

(Fig.8.22.) Comparative overview of the distribution of categories viewed as the most common sources of clause linkers – markers displaying binary overlaps

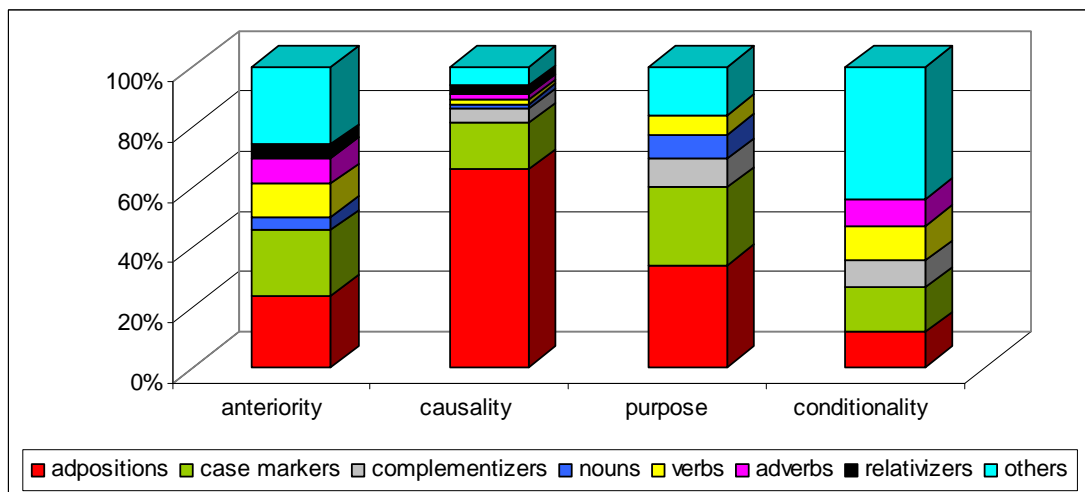


(Fig.8.23.) Graphic representation of the distribution of categories viewed as the most common sources of clause linkers – markers displaying binary overlaps

Of course, the list of the seven categories given above does not exhaust the range of homonyms/polysemes of the four relations identified in the data collected – many other relation-specific binary overlaps have been listed in the analytical chapters. Fig.8.24. and Fig.8.25. include these additional categories labelling them as ‘other’. The data indicate that the set of the seven categories remains in similar proportions as previously described (Fig.8.22. and 8.23.) only for causality. For anteriority, purpose and conditionality the bars indicating the proportion of categories outside of this hypothesized core set are significantly higher. However, as it has been discussed in section 4.2.1., in the case of anteriority, with just one exception, the remaining 13 overlaps are distributed over 12 categories. Only one of the categories – aspect markers – occurs twice as a polyseme of anteriority linkers, and so the other categories are either language-specific polysemes or homonyms. As for the relation of purpose, the 10 polysemes constituting the category ‘other’ include 2 instances of TAM markers, 2 modality suffixes/particles and 2 conjunctions (see section 6.2.1.). The 15 members of the category ‘other’ in conditionality, on the other hand, include 3 topic markers, 3 TAM and modality markers, and 2 demonstratives (see chapter 7.2.1.) For both purpose and conditionality almost all these categories occur also in multiple overlaps. Moreover, for conditionality c-glossemes all three: topic markers, demonstratives and modality markers have been claimed to be cross-linguistically important sources of this group of linkers in earlier works by Traugott (1985) and Heine and Kuteva (2007) – see section 3.2.8. In the light of these facts we can conclude from the observations of binary overlaps that the core set of sources of purpose and conditionality markers is larger than for the other two relations.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
adpositions	11	23.92%	42	65.62%	21	33.87%	4	11.77%
case markers	10	21.75%	10	15.62%	16	25.80%	5	14.70%
complementizers	-	-	3	4.69%	6	9.68%	3	8.82%
nouns	2	4.35%	1	1.57%	5	8.07%	-	-
verbs	5	10.87%	1	1.57%	4	6.45%	4	11.77%
adverbs	4	8.67%	1	1.57%	-	-	3	8.82%
relativizers	2	4.35%	2	3.12%	-	-	-	-
others	12	26.09%	4	6.24%	10	16.13%	15	44.12%
TOTAL	46	100%	64	100%	62	100%	34	100%

(Fig.8.24.) Comparative overview of the distribution of sources of clause linkers – markers displaying binary overlaps

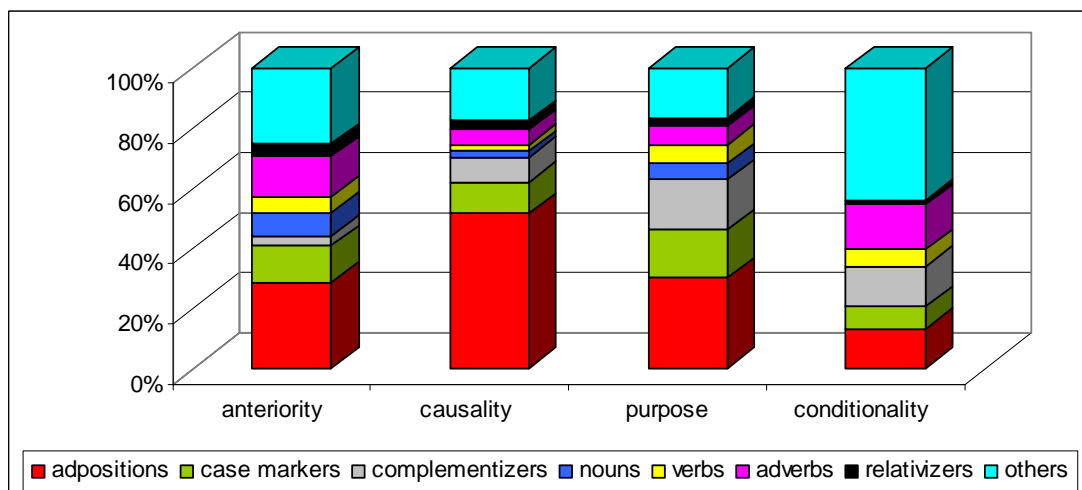


(Fig.8.25.) Graphical representation of the distribution of sources of clause linkers – markers displaying binary overlaps

Quite surprisingly, it turns out that if we look at the distribution of categories in both binary and multiple overlaps (where each category is listed separately irrespective of the other categories it overlaps with), the picture (see Fig.8.26. and Fig.8.27.) looks very similar to that depicted for binary overlaps. This strengthens the earlier observations.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
adpositions	26	28.57%	50	51.55%	34	30.08%	9	13.24%
case markers	11	12.09%	10	10.31%	18	15.93%	5	7.35%
complementizers	3	3.30%	8	8.25%	19	16.81%	9	13.24%
nouns	7	7.69%	2	2.06%	6	5.31%	-	-
verbs	5	5.50%	2	2.06%	7	6.20%	4	5.88%
adverbs	12	13.19%	5	5.15%	7	6.20%	10	14.71%
relativizer	4	4.39%	3	3.09%	3	2.66%	1	1.47%
others	23	25.27%	17	17.53%	19	16.81%	30	44.11%
TOTAL	91	100%	97	100%	113	100%	68	100%

(Fig.8.26.) Comparative overview of the distribution of sources of clause linkers – all overlaps included



(Fig.8.27.) Graphical representation of the distribution of sources of clause linkers – all overlaps included

8.5.2. Evidence from the analysis of incorporated material

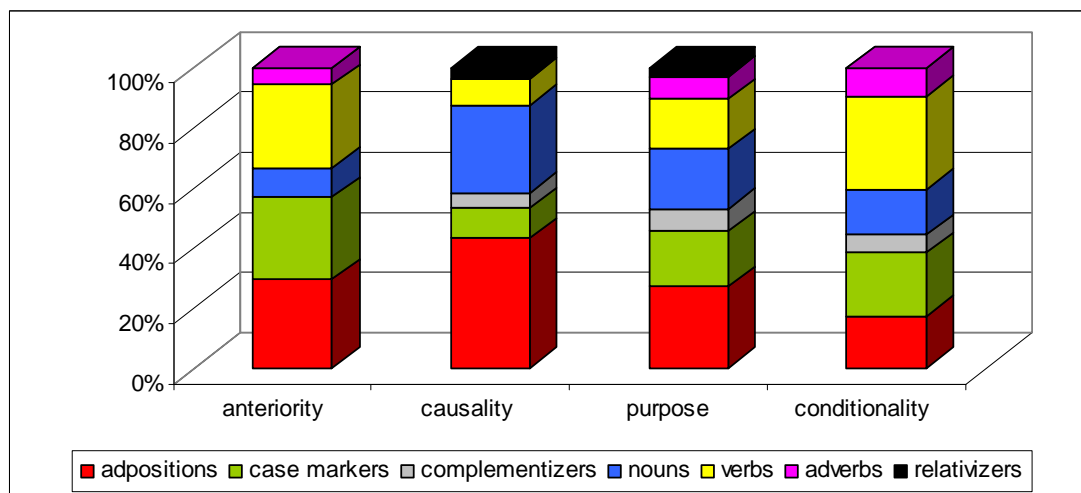
Apart from reconstructing synchronic patterns of syntactic polyfunctionality in the quest for the origin of clause linkers the internal structure of polymorphemic markers has also been analysed. The results have shown that the categories most commonly incorporated in the structure of these polymorphemic markers (ergo their sources) are, indeed, the 7 categories most commonly considered to be the sources of clause linkers.

In the table below (Fig.8.28.) the categories, together with numerical data, are listed for each of the relations. For clarity of presentation only the items belonging to one category have been listed. Hence, for instance, for anteriority the adpositions incorporated in the structure of polymorphemic markers have been included as have been adverbs, but if a particular item belongs to both categories, it has been omitted. The number of items excluded in this way is, however, not significant and so it does not affect the validity of observations.

	anteriority		causality		purpose		conditionality	
	count	%	count	%	count	%	count	%
adpositions	15	29.42%	34	43.04%	18	27.27%	9	17.31%
case markers	14	27.45%	8	10.13%	12	18.18%	11	21.15%
complementizers	-	-	4	5.06%	5	7.58%	3	5.77%
nouns	5	9.80%	23	29.11%	13	19.69%	8	15.38%
verbs	14	27.45%	7	8.86%	11	16.67%	16	30.77%
adverbs	3	5.88%	-	-	5	7.58%	5	9.62%
relativizers	-	-	3	3.80%	2	3.03%	-	-
TOTAL	51	100%	79	100%	66	100%	52	100%

(Fig.8.28.) Comparative overview of the distribution of categories incorporated in the structure of polymorphemic markers

The diagram illustrates the data graphically:



(Fig.8.29.) Graphical representation of the distribution of categories incorporated in the structure of polymorphemic markers

What is probably the most interesting conclusion here is the fact that the proportion of nouns as well as verbs has increased significantly in comparison to the results obtained in the analysis of patterns of polysemy (cf. Fig.8.23, Fig.8.25. and Fig.8.27.). This trend is noticeable for all of the relations but is especially striking for causality and conditionality. Among the material incorporated in polymorphemic causality linkers listed in the table above, almost 30% are nouns and almost 9% are verbs. In the case of conditionality nouns contribute over 15% and verbs over 30%. The presence of these two content categories is self-explanatory. The polymorphemic markers are often inflected nouns and verbs, nominal and verbal phrases or some other structures built around these two categories. They have not undergone phonological erosion and thus their internal make-up remains transparent. Monomorphemic markers which served as

the main source of data in the previous section, on the other hand, rarely reveal the full history of their origin – in majority of cases their synchronic structure allows to trace their origin not further back than to Heine and Kuteva's layer IV of grammatical development (see section 3.2.3.) i.e. the layer occupied by demonstratives, adpositions, aspect markers and negation markers. The second interesting observation that we can make here is that, in both the analysis of patterns of syntactic overlaps and incorporated material, the proportion of adpositions appears to be very similar – with the highest score among causality linkers and lowest among conditionality markers. The proportion of case markers is also similar in both aspects of the analysis.

8.5.3. Other c-glossemes as sources

Apart from the discussion on the development of clause linkers out of other syntactic categories the scenario of one clause linker giving rise to another one has also been considered. It has been emphasized that on a basis of synchronic observations only, it is often difficult to reconstruct the pathways of grammaticalization that lead to the emergence of what has been labelled *semantic overlaps*. For instance, there is no convincing evidence for the development of anteriority linkers out of some other type of circumstantial linkers. Although there is a high degree of semantic affinity between the relation of ANTE and SIOVER ('when') the directionality of the extension cannot be easily established. However, a sufficient amount of evidence exist to support the hypothesis that anteriority linkers give rise to causality linkers. The latter ones seems also to commonly develop from markers of other temporal relations – including SIOVER ('when') and TAQUO ('since'). Another important overlap – CAUSE-PURPOSE, on the other hand, in the majority of cases seems to originate from the extension of the causal meaning over purposive but this does not seem exceptionless. The analysis of further overlaps of the markers of purpose does not provide conclusive evidence as for the possible pathway of development either and I am not aware of any other studies that have looked into the issue. Conditionality linkers, on the other hand, without any doubts develop commonly from markers of the relation of simultaneity overlap (SIOVER, 'when'). The data collected suggest also that linkers encoding the relations of anteriority, simultaneity duration (SIDUR 'while') and causality may get reanalysed into conditionality c-glossemes too.

Some of the pathways presented here have been reported in other works. Others, requiring verification on the basis of diachronic evidence, or at least more detailed historical reconstructions, for the time being have to remain merely suspicions and hypotheses. On the whole we may conclude, however, that the category of clause linkers itself should be included to the core of the discussion on origin of the markers of circumstantial relations along with other syntactic categories. So far the fact of one

type of clause linker giving rise to another type has been treated rather modestly in the grammaticalization literature.

8.5.4. *The emerging picture*

The analysis has confirmed, beyond all doubts, that “grammaticalization of items, whether lexical or morphological, is constrained by the grammatical function to be expressed, and by the appropriateness of the inferences from the source items for the function in question” (Hopper and Traugott 2003:186). It has also confirmed that on the whole for the analysed groups of c-glossems, the most common categories giving rise to clause linkers are, indeed, adpositions, case markers, nouns and verbs. However, the role of other categories often viewed as important material for grammaticalization of clause linkers – i.e. complementizers, relativizers, adverbs, demonstratives and interrogatives – has turned out to be far less significant. Some of these more rare categories are without any doubts important sources of specific types of clause linkers (such as complementizers for purpose or adverbs for conditionality) and there are several more categories that emerge only at type-specific level (aspect markers in the case of anteriority or topic markers for conditionality, for instance). This concerns also clause linkers themselves, which, as has been summarized in the previous section, in the process of semantic extension may give rise to markers of other circumstantial relations.

The overall conclusion that we may draw from these observations is that we should be careful when making general claims about the origin and sources of the clause linkers – it is far too easy to oversimplify focusing only on the core set of circumstantial relations. In reality, the deeper we go into the cross-linguistic analysis of markers of a particular circumstantial relation (or any other group of items, for that matter), the more the grammaticalization pathways start to resemble a multi-node rhizome rather than a neat diagram. This becomes obvious, as this study shows, even from studying the fossilized evidence of grammaticalization processes preserved in synchronic material.

The study has looked into the issue of origin of linkers of the four relations in more details than any other previously, resorting to theoretical discussion, qualitative and quantitative analyses. Many of the observations made here support the hypothesis put forward by linguists studying grammaticalization. Some other findings cast doubts on certain earlier claims and yet others present us with new evidence which will be worth looking at in more details in future research. In all cases, I emphasize once again, the results this study yields should be viewed in the context of its purpose – to discover cross-linguistic tendencies rather than universals.

8.6. SEMANTIC AFFINITIES BETWEEN THE RELATIONS

It has been already said in chapter 4 that it is Kortmann's study (1997) that inspired me to look at the problem of semantic affinities between the circumstantial relations. My aim was not to reconstruct a complete picture of such a network – an impossible task in a study that looks at four circumstantial relations only – but to verify whether a language sample completely different and database consisting of a variety of clause linkers and not just 'ideal adverbial subordinators' would yield similar results. The null hypothesis was that it would as it had been assumed that human perception and cognitive organization (at least when it comes to the circumstantial relations) is universal and does not depend on the language we speak or the type of clause linkers we use. The results confirmed the hypothesis. All the major and strongest semantic affinities involving the four relations identified by Kortmann have been confirmed. This includes ANTE-SIOVER, ANTE-CAUSE, ANTE-SIDUR, CAUSE-SIOVER, CAUSE-PURPOSE, PURPOSE-RESULT, COND-SIOVER and COND-CAUSE. The small differences in the strength of particular semantic affinities between mine and Kortmann's study cannot be explained by conceptual differences between speakers of different languages. I would attribute them rather to the differences in the designs of these two studies.

Apart from the positive verification of the validity of Kortmann's findings in a larger and more diverse sample of both languages and structures, this study has proposed a more detailed insight into the arrangement of particular overlaps in the network of semantic affinities. While Kortmann depicted only the most salient affinities (cf. Fig.3.8.), I – taking the advantage of the smaller number of relations in my study – have presented the full picture emerging from my analysis. This kind of presentation has two main advantages: it allows us to look not only at the core but also at the margins of the conceptual space of each of the analysed relations and depicts clearly not just binary but also multiple overlaps. Although these overlaps are often limited when it comes to the number of languages they have been identified for, the reconstructed networks nonetheless offer us very close and direct insight into the impressive cognitive organization of circumstantial concepts. It is unlikely that any tool other than analysis of cross-linguistic patterns of semantic overlaps will be ever better in revealing the truth about the way these concepts are related in our minds.

PART THREE

**CROSS-LINGUISTIC VARIATION
AND ITS SOCIO-CULTURAL CORRELATES**

If we take grammaticalisation seriously, that is, if we understand that all aspects of grammar are the result of grammaticalisation, and we understand that grammaticalisation (and lexicalisation) is the conventionalisation of repeated patterns of use (using the same form to constrain the addressee's interpretation of the speaker's communicative intention in the same way over and over again), then there must by logical necessity be a connection between all conventionalised aspects of language and the culture/cognition of the speakers, otherwise the speakers would not have used those particular forms in those particular ways over and over again to constrain the interpretation of that particular semantic domain in that particular way, to the extent that the forms became conventionalized. That is, constraining the interpretation of that particular semantic domain in that way must have been important for them, important enough for them to put the extra effort into constraining the interpretation in that way.

Randy LaPolla

In the previous part of the thesis our attention has been devoted predominantly to the analysis of semantic and syntactic polyfunctionalities in particular groups of circumstantial glossems. Having presented the findings concerning the origins and synchronic ambiguity of the markers we may now turn to two more general, yet very interesting problems.

The first one concerns the issue of cross-linguistic variation in the range and types of c-glossems. To approach the topic three parameters are to be considered in chapter 9:

- a) degree of grammaticalization – the extent to which a particular circumstantial relation is grammaticalized in the languages in the sample;
- b) degree of lexicalization – availability of (fully) lexicalized c-glossems in the languages;
- c) degree of explicitness – availability of unambiguous means for expressing a particular relation.

The chapter discusses also some possible motivations for the emergence of particular types of marking. This is where the second problem: causes/explanations for linguistic change and reasons for the cross-linguistic variation in the availability of c-glossems is presented.

The development of clause linkers, or any other strategy of clause linking for that matter, is to be viewed as a part of linguistic change at the highest level of grammar – complex sentence formation. As in any type of linguistic change, also here we should accept that there are both internal and external causes, or explanations, for

changes.¹ The internal ones include physical and cognitive explanations: limitations and potentials of human speech production and perception, processing, and learning. The external causes, on the other hand, as Harris and Campbell (1995:316) explain, “involve factors that are largely outside of language *per se* (outside the human organism); they include, for example, expressive uses of language, positive and negative social evaluations (prestige, stigma), the effects of literacy, prescriptive grammar, and educational policies, political decree, language planning, and language contact”. Furthermore, as the authors aptly notice, the factors naturally interact in complex ways – overlapping and, sometimes, competing with each other, complicating the task of trying to understand linguistic change.

In a work of limited length, such as this one, it is not possible to consider all the factors and scenarios. I shall therefore focus on those elements which, in my opinion, are the most important for the emergence and functioning of c-glossems: pragmatic, cognitive and socio-cultural. Among those three, special attention is devoted to the last group which, with the development of linguistic typology and cross-linguistic research, is gaining more and more interest in the discussions of differences in structures of the world’s languages. Aiming to contribute to these discussions, and still addressing the problem of origin and functioning of clause linkers, in chapter 10 I present an analysis of correlations between the three parameters listed above: degree of grammaticalization, lexicalization and explicitness, and a range of socio-cultural factors such as: presence and length of written tradition, population size, type of society in which the language is spoken, presence of the language in school teaching and in TV and radio broadcasting. Finally, in chapter 11, I attempt to propose a unified explanation of the reasons behind the cross-linguistic variation in the domain of c-glossems and clause-linking strategies in general.

¹ For some discussion and examples, see for instance: Campbell and Ringen 1981; Jahr 1989; Hawkins 1983; Campbell, Bubenik, and Saxon 1988.

CHAPTER 9

Degrees of grammaticalization, lexicalization and explicitness

The world's languages differ as to the range and types of circumstantial clause linkers they have at their disposal. While some languages possess fully grammaticalized or lexicalized markers of particular type of circumstantial relation, others have at their disposal only ambiguous ones. The aim of this chapter is to investigate this variation. In section 9.1. I focus on the analysis of degree of grammaticalization and in section 9.2. on the degree of lexicalization of particular groups of c-glossemes. In section 9.3., devoted to an inquiry into the degree of explicitness of expressing circumstantial relations, I take into account not only clause linkers but also other strategies that languages employ when they do not have at their disposal any c-glossemes. It is also in that section where borrowing of clause linkers is considered in more detail.

9.1. DEGREE OF GRAMMATICALIZATION

The notion of grammaticalization was the subject of discussion in chapter 3. The analytical chapters that follow it have revealed that next to monomorphemic and unambiguous clause linkers many of the c-glossemes are not only morphologically complex but also used in a variety of other syntactic functions and as exponents of more than one circumstantial relation. The question I wish to address here is to what extent the languages in the sample have grammaticalized the analysed concepts of anteriority, causality, purpose and conditionality and what is the geographic distribution of the degrees of grammaticalization of these relations.

b) **Group 2 – M1** (monomorphemic c-glosseme encoding up to one additional circumstantial meaning)²:

M1M – monomorphemic c-glosseme encoding up to one additional circumstantial meaning and syntactically monofunctional;

M11 – monomorphemic, c-glosseme encoding up to one additional circumstantial meaning and homonymous/polysemous with an item in one other syntactic category;

M12 – monomorphemic, c-glosseme encoding up to one additional circumstantial meaning and homonymous/polysemous with items in two other syntactic categories.

Importantly, the categories of the second group (M1) are inclusive of the first group (MM) but not the other way round. Moreover, category MM2 is inclusive of MM1 (and so is M11 of M12) but MMM is not inclusive of MM1 (and neither is M1M of M11). Although the set of values is limited it, nonetheless, allows us to make strong conclusions about the degree to which the four circumstantial relations are grammaticalized in the languages in this sample.³ The most important indicator of the degree of grammaticalization is, of course, the presence of MMM marking.

For comparative purposes I have excluded from the summary below the languages for which data on one or more relation are missing. The summary concerns 67 of the languages with full set of data available. Since grammaticalization, although externally motivated in some cases (cf. section 3.2.7.), is in principle a language internal process, the summaries do not include borrowings.

The results of the analysis (see Fig.9.1.) show that purpose reveals the lowest degree of grammaticalization with just 9 languages possessing MMM linkers. The score for anteriority is not much higher – only 13 of the languages have at their disposal a fully grammaticalized marker of this relation. Causality, which scored highest, is fully grammaticalized in as many as 26 languages and conditionality is not far away from it with the score of 24. The differences between purpose and anteriority on the one hand and conditionality and causality on the other are quite striking. The same ranking: purpose < anteriority < conditionality < causality, holds also if we take into account markers which being monomorphemic and monofunctional semantically serve one

² When categorizing the c-glosseme into group M1 only those markers which have semantic scope over relations recognized as cognitively close to the relations analysed have been included (see sections 4.3., 5.3., 6.3, and 7.3. respectively). This was dictated by the goal of treating semantic affinity as an additional indicator of tightness between various points in grammaticalization pathways.

³ The reason why, for instance, c-glosseme with more than two additional syntactic functions are omitted from the analysis is that they contribute only a small amount of items (from 3% to 6% depending on the relation as shown in Fig.8.13. in chapter 8).

(columns MM1) or two (columns MM2) additional syntactic function, despite the fact that the discrepancies between the relations in these cases are much smaller than for the MMM group.⁴

ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
MMM	MM1	MM2	MMM	MM1	MM2	MMM	MM1	MM2	MMM	MM1	MM2
13	26	29	26	35	36	9	22	27	24	33	36

(Fig.9.1.) Degrees of grammaticalization – MM markers (67 languages included)⁵

Since among the markers of purpose there is a strong cross-linguistic tendency for specialization of marking of same- and different-subject clauses, a separate analysis has been performed in order to find whether there exist a difference in degrees of grammaticalization of these two groups of markers. All the 70 languages for which the information necessary for encoding the values are available have been taken into account. The data have not revealed any differences though. For both same-subject and different-subject clauses MMM markers are available in 9 languages (see Fig.9.2.).

PURPOSE SAME-SUBJECT			PURPOSE DIFFERENT-SUBJECT		
MMM	MM1	MM2	MMM	MM1	MM2
9	27	34	9	26	32

(Fig9.2.) Degrees of grammaticalization of MM purpose markers in same- and different-subject clauses⁶

The results obtained after extending the analysis to markers of the second group – M1 (monomorphemic c-glossemes encoding up to one additional circumstantial meaning) – are quite similar to those from the first group. As we read from (Fig.9.3.) it is, again, purpose that reveals the lowest degree of grammaticalization (9 occurrences of M1M markers). It is followed quite closely by anteriority (14 occurrences). Causality has a M1M marking in almost twice as many languages as anteriority (27) but is slightly overtaken by conditionality (30 occurrences). The difference between causality and conditionality, similarly as in the MM group is, however, very insignificant and so no strong conclusions can be made from these findings. What is obvious, however, is that these two relations stay in direct opposition to anteriority and purpose. The discrepancies in the degrees of grammaticalization are smaller for the M11 and M12

⁴ If we go back to the summaries presented in chapter 9, we can conclude that the increase in columns MM1 and MM2 for purpose and anteriority is due to the high number of binary overlaps of exponents of these two relations with adpositions and case markers (see especially Fig.8.22.).

⁵ For full list of degrees of grammaticalization of particular languages in the MM group see Appendix IV.

⁶ For full list of degrees of grammaticalization of same- and different-subject purpose markers in particular languages in the MM group see Appendix V.

markers but causality and conditionality still overtake the other two relations. There are no implicational hierarchies emerging from the analysis of the relations in either the first (MM) or the second group (M1).

ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
M1M	M11	M12	M1M	M11	M12	M1M	M11	M12	M1M	M11	M12
14	27	31	27	43	44	9	28	33	30	45	48

(Fig.9.3.) Degrees of grammaticalization – M1 markers⁷

Before we move on to conclusions that we may draw from these observations, let us first look at the geographic distribution of the degrees of grammaticalization in the analysed sample.

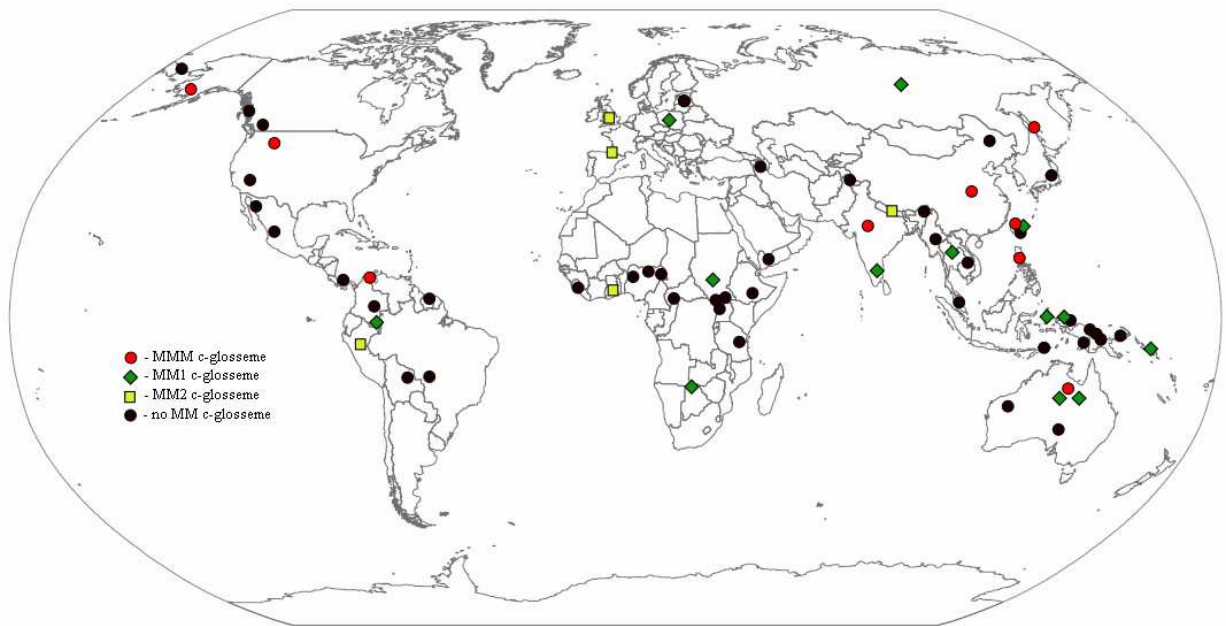
9.1.2. Geographic distribution

Although it would be very interesting to look more closely at the genetic distribution of the degrees of grammaticalization (i.e. distribution in particular language families), the size of the sample used for the purpose of this thesis does not allow for such analyses. Hence, the discussion in this section (as well as in the parallel sections devoted to discussion of degrees of lexicalization and explicitness) is limited to geographic distribution with only occasional remarks on language families. For comparative purposes the languages marked on the maps in this section are limited to the 67 included in analysis in the previous section. Markers which are monomorphemic and semantically monofunctional (MM) have been, again, treated separately from those which are monomorphemic and semantically polyfunctional (M1).

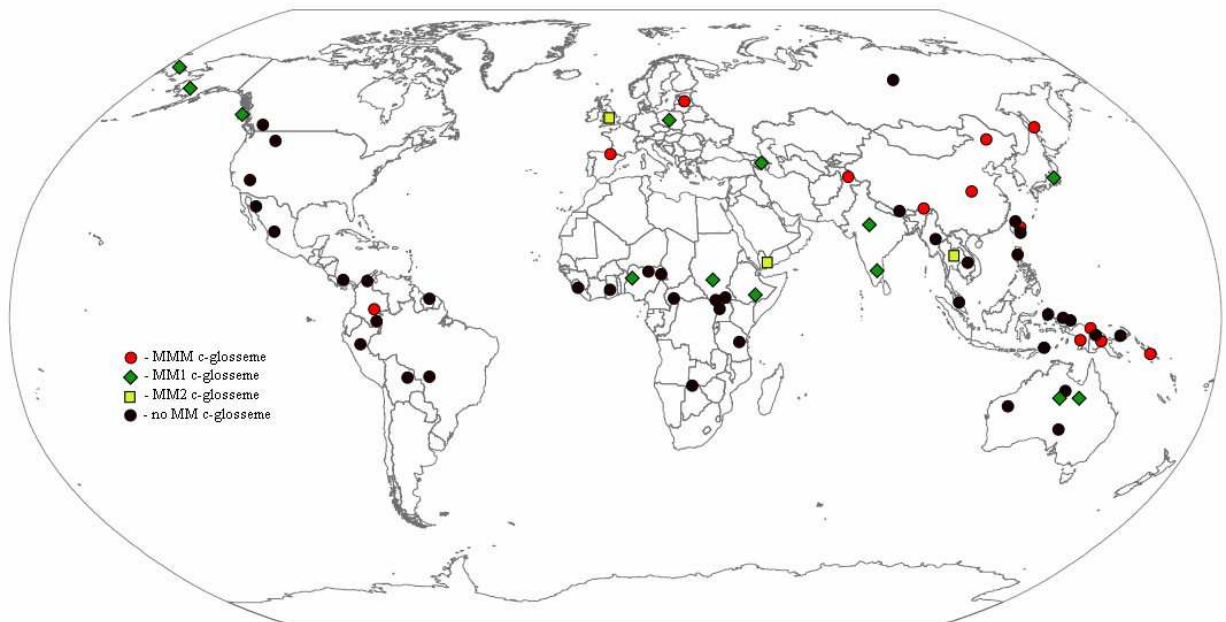
Let us begin with the comparison of the distribution of degrees of grammaticalization in the two relations that turned out to be the least grammaticalized – purpose and anteriority. For both of them, as we can see from (Fig.9.4.) and (Fig.9.5.) there are virtually no fully grammaticalized (MMM) purpose and anteriority markers in the languages of Africa and a majority of languages in this region do not have at their disposal MM1 or MM2 markers either. A similar picture is revealed in the languages of Oceania and Southeast Asia as well as Australia and Taiwan. There are only four examples of MMM marking of purpose clauses and five examples of MMM marking of anteriority clauses in this group. As for the latter, four of the 5 markers come from the languages of New Guinea. Anteriority appears to be better grammaticalized also in the languages of Europe and the languages of central and East Asia. Both purpose and anteriority are poorly grammaticalized in the Americas – there is just one case of MMM marker of anteriority (in Cubeo) and three of purpose (in Yup'ik, Nez Perce and

⁷ For full list of degrees of grammaticalization of particular languages in the M1 group see Appendix VI.

Ika). In total 10 of the 14 languages of Americas depicted on the maps do not have any MM c-glosseme of purpose and 11 do not have any MM c-glosseme of anteriority.

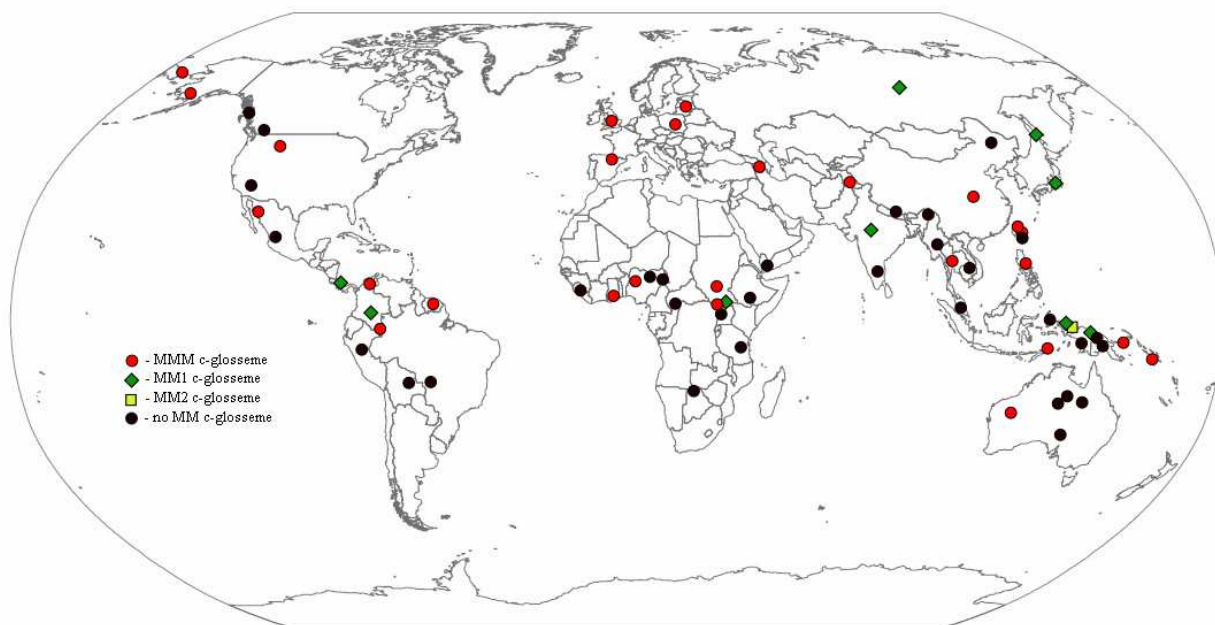


(Fig.9.4.) Geographic distribution of MM purpose markers

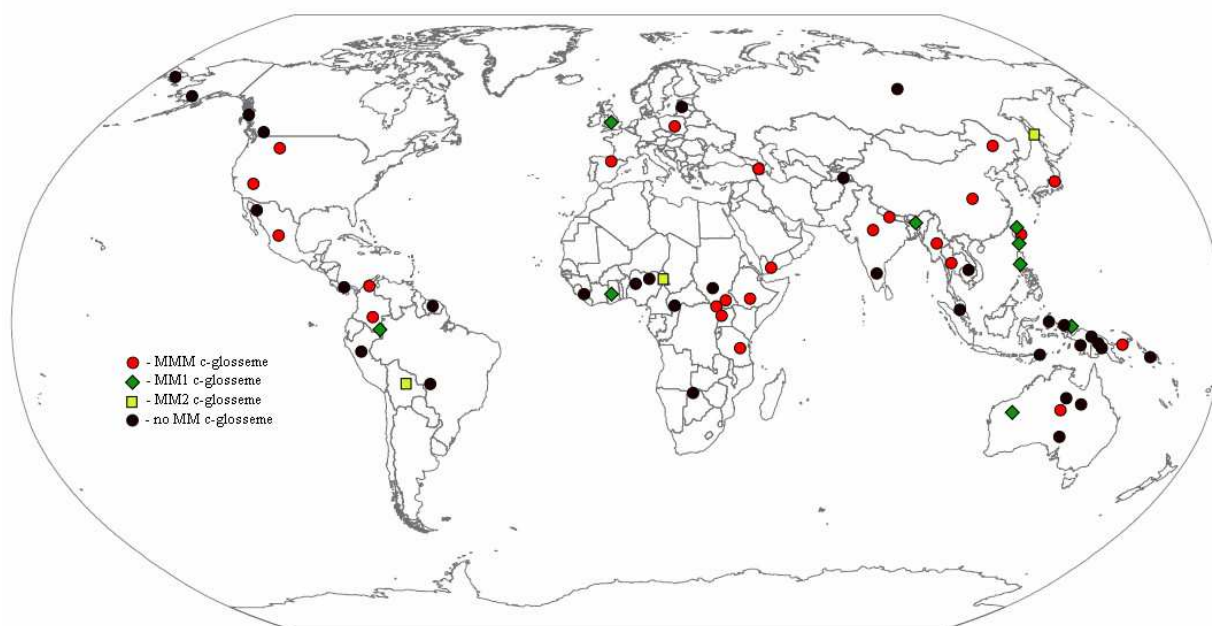


(Fig.9.5.) Geographic distribution of MM anteriority markers

The picture looks quite different for the MM marking in causality and conditionality as depicted in (Fig.9.6.) and (Fig.9.7.) respectively. Although still over a half of the African languages analysed do not have MM marking of these two relations, causality and conditionality are far better grammaticalized in this area than anteriority and purpose. Lango, Ma'di and Didinga – the three Nilo-Saharan languages lying in a close proximity from each other – all have MMM marking of conditionality. Ma'di has a MMM marker for causality as well, while Didinga resorts to a MM1 marker. Among the four languages of Europe (English, Polish, Estonian and Basque) discussed here all have MMM marking of causality but the same holds only for two of them when it comes to marking of conditionality (Polish and Basque). The latter relation, on the other hand, has more MMM marking in the languages of central and Southeast Asia. In Oceanic and Australian as well as American languages both the relations are on the whole rather modestly grammaticalized but in the American languages the degree of their grammaticalization is much higher than for anteriority and purpose.

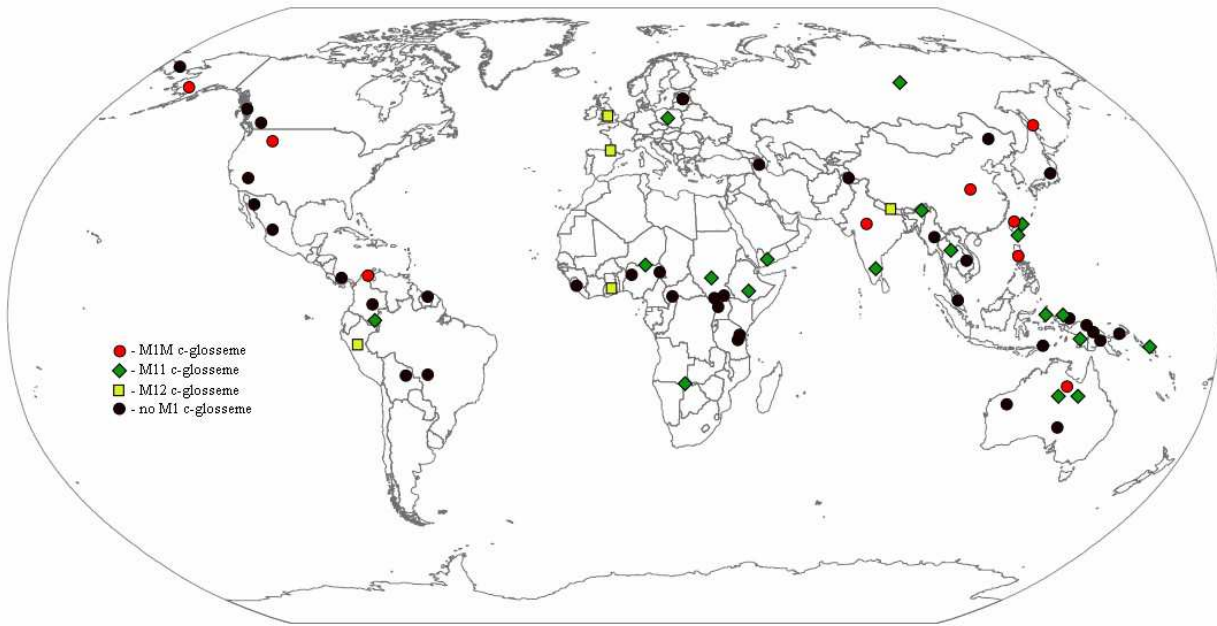


(Fig 9.6.) Geographic distribution of MM causality markers

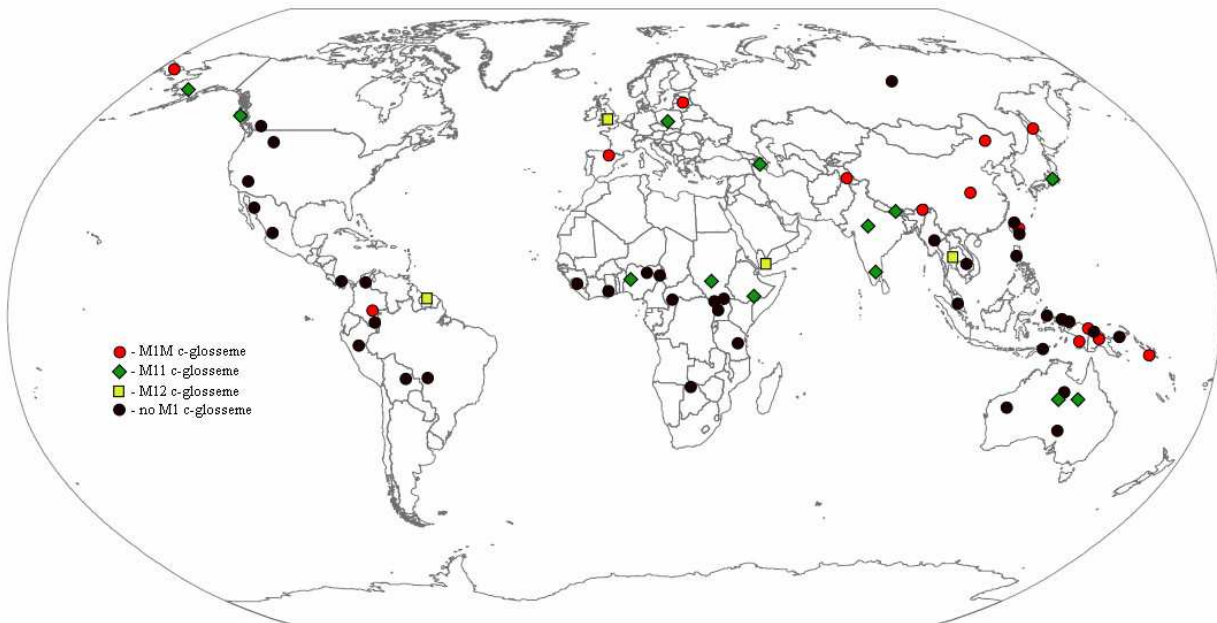


(Fig.9.7.) Geographic distribution of MM conditionality markers

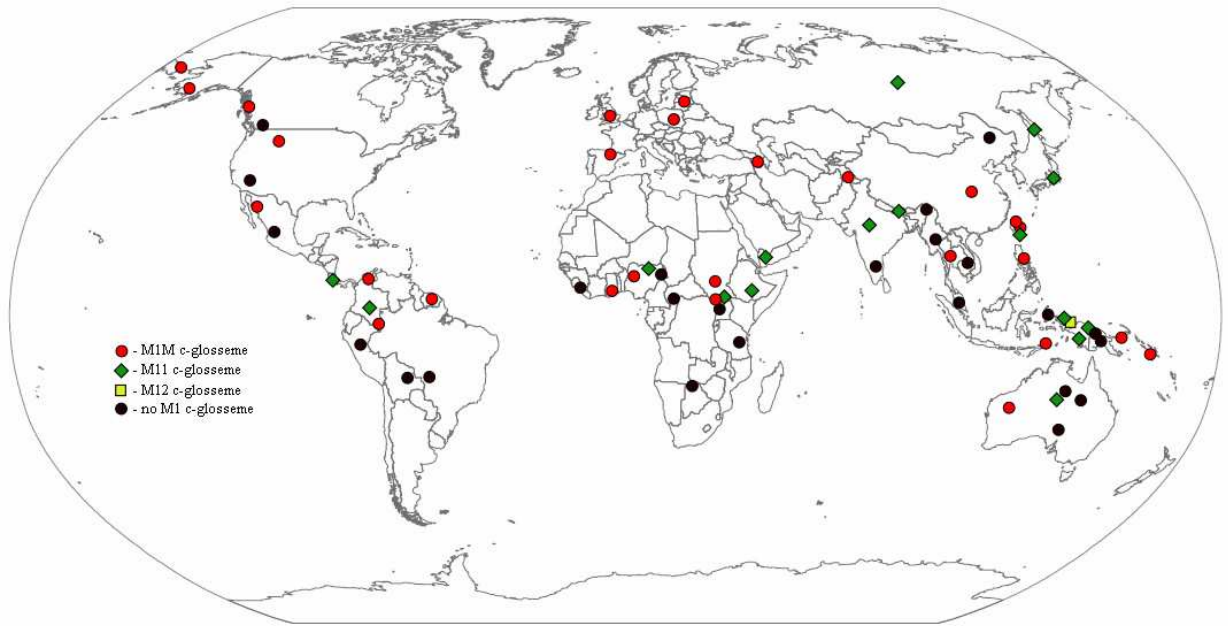
The picture emerging from the analysis of patterns of geographic distribution of M1 linkers (Fig.9.8.-9.11.) is very similar to that presented above. When comparing MMM markers with M1M markers there is no difference in grammaticalization of purpose c-glossems and only one additional c-glosseme appears on the map of anteriority (contributed by Chukchi). For anteriority the picture from (Fig.9.5.) is enriched also by two further examples of M1 c-glossems (contributed by Ndyuka and Lepcha). For purpose the map in (Fig.9.8.) is enriched in comparison to (Fig.9.4.) by 6 M11 c-glossems: from Eipo, Galo, Hausa, Konso, Arabic and Yami. This alters slightly the picture of grammaticalization of c-glossems in Africa but does not change the other conclusions made on the basis of observation of MM marking for the two relations. The same applies to the maps concerning distribution of causality and conditionality M11 markers (Fig.9.10. and Fig.9.11. respectively). On the whole it is the picture of conditionality that has changed the most – in comparison to (Fig.9.7.) (Fig.9.11.) has been enriched by contribution of M1M markers from Boko, Gola, Hausa, Jingulu, Lillooet and Yup'ik and of M11 markers from Chukchi, Krongo, Lavakuleve, Rama, Taba and Tamil. The observation of patterns of M1 marking may give us an impression that conditionality is, on the whole, better grammaticalized in the languages of Africa and Americas than causality. We need to remember, however, that the quantitative differences are quite small here and that the languages depicted on the map represent merely 1% of the languages being spoken nowadays.



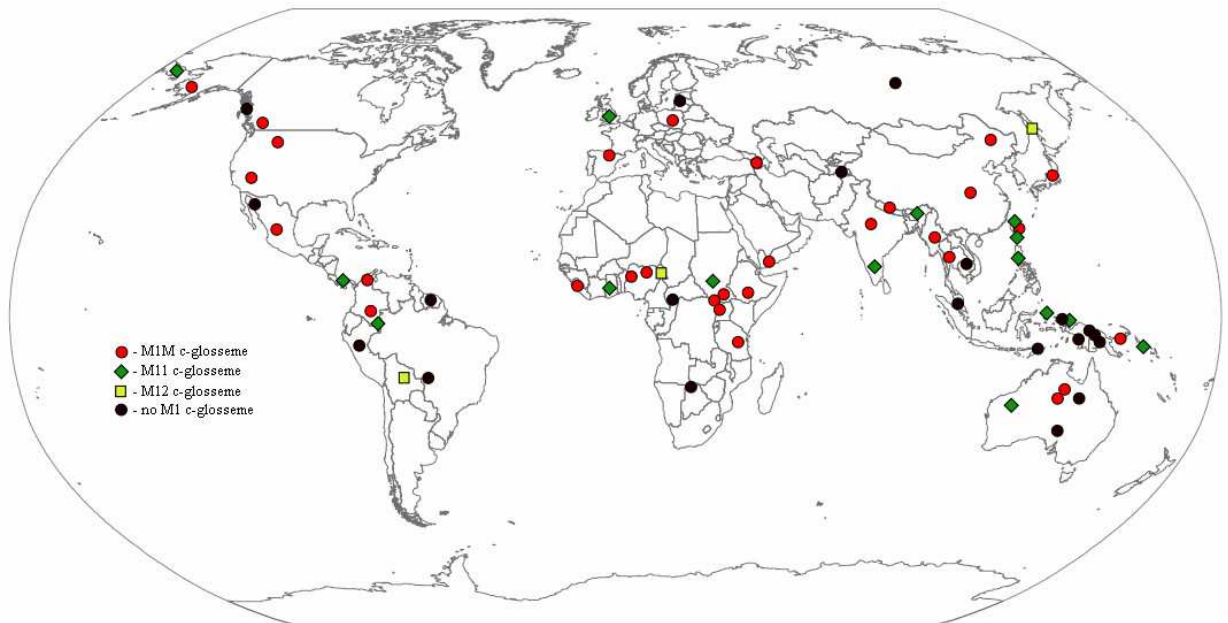
(Fig.9.8.) Geographic distribution of M1 purpose markers



(Fig.9.9.) Geographic distribution of M1 anteriority markers



(Fig.9.10.) Geographic distribution of M1 causality markers



(Fig.9.11.) Geographic distribution of M1 conditionality markers

The most important trends emerging from the observations of geographic distribution can be summarized as follows:

- Oceanic languages tend to not have highly grammaticalized markers of purpose and conditionality but do have grammaticalized anteriority and causality;

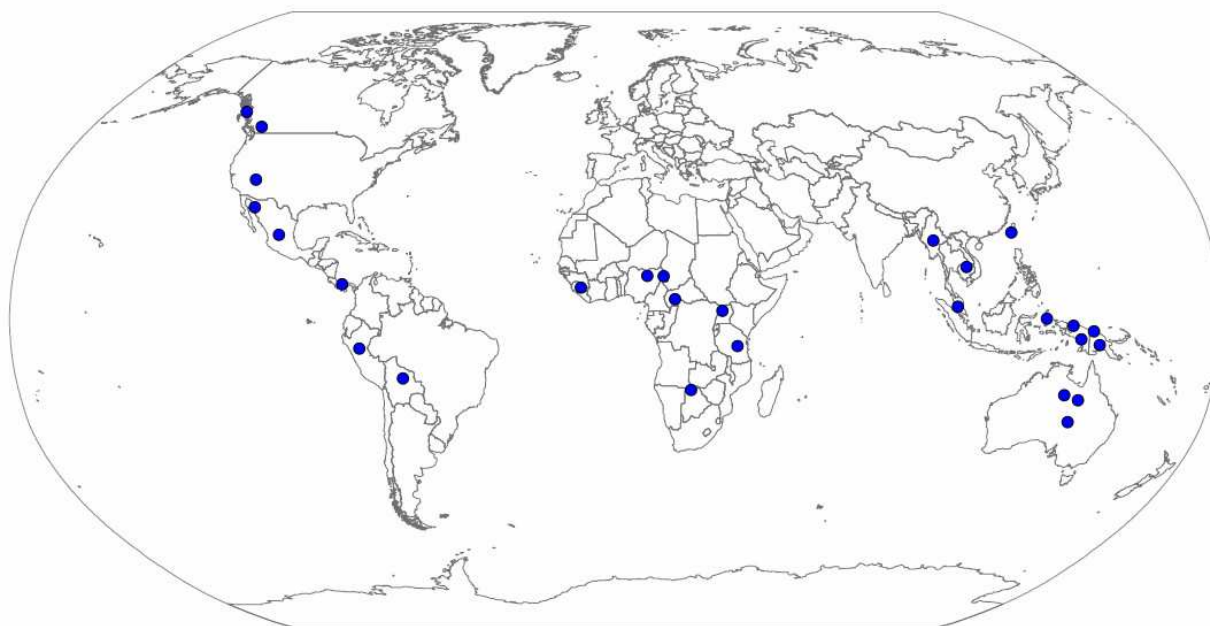
- in African and American languages the encoding of causality and conditionality is much more grammaticalized than encoding of the other two relations;
- in Asian languages purpose is the least and conditionality the most grammaticalized relation with causality and anteriority in the middle, although all the major Asian languages: Hindi, Tamil, Thai, Mandarin, Japanese have grammaticalized, almost without exceptions, all the relations using the range of markers considered here.⁸ In those languages it is conditionality that is most highly grammaticalized with purpose marking revealing the lowest degree of specialization;
- in the four languages of Europe (English, Polish, Estonian and Basque) purpose is the least grammaticalized relation; anteriority reveal slightly higher degree of grammaticalization and is closely followed by conditionality; all the four languages have at their disposal a MMM marker of causality.

Finally, it should be mentioned also that certain languages reveal much lower overall degree of grammaticalization of the analysed circumstantial relations than others. The table below (Fig.9.12.) presents the data for all the 27 languages which did not have any positive values in 4 or more of the 6 types of markers considered here (i.e. MMM, MM1, MM2, M1M, M11, M12). As can be read from the table, the languages with the lowest degree of grammaticalization of the four circumstantial relations in the analysed sample are Arabana, Isaka, Jahai, Sango and Sapuan. All the languages listed in table (Fig.9.12.) are also depicted on the map in (Fig.9.13.) from which it is clear that they come exclusively from Americas (especially from North and Central America), Africa and the Indo-Pacific region (including Southeast Asia, Taiwan, New Guinea and Australia). There are no European or central and west Asian languages among this group. On the contrary, all the European languages (with the exclusion of Estonian) have MM markers available for all four relations. The other languages revealing the same overall degree of grammaticalization are Mandarin, Thai, Hindi Nivkh and Rukai. This set is without any doubts strongly biased towards the (Indo-)European languages on the one hand and the worlds' major languages on the other. Japanese as the only one of the major languages does not have a MM marking of purpose but this concerns only marking of different-subject purpose clauses. I return to these observations in chapters 11 and 12.

⁸ The only exception is the lack of purpose markers that could fit any of the 6 categories (MMM, MM1, MM2, M1M, M11, M12) in Japanese and lack of causality marker in Tamil.

Language	Number of relations for which no MMM, MM1 or MM2 marker is available	Number of relations for which no M1M, M11 or M12 marker is available	Total number of entries with 0 value (maximum = 8)
Arabana	4	4	8
Baure	3	3	6
Eipo	3	1	4
Gola	4	3	7
Hatam	3	2	5
Hausa	4	1	5
Hualapai	3	3	6
I'saka	4	4	8
Jahai	4	4	8
Jingulu	3	2	5
Kanuri	3	3	6
Kayah Li	3	3	6
Khwe	3	3	6
Lango	3	3	6
Lillooet	4	3	7
Nisga'a	3	2	5
Quechua Huallaga	3	3	6
Rama	3	2	5
SE Tepehuan	3	3	6
Sango	4	4	8
Sapuan	4	4	8
Seri	3	3	6
Swahili	3	3	6
Taba	3	2	5
Wambaya	2	2	4
Yami	3	1	4
Yimas	3	3	6

(Fig.9.12.) Languages with the lowest degree of grammaticalization of the four circumstantial relations



(Fig.9.13.) Geographic distribution of the languages with the lowest degree of grammaticalization of the analysed groups of c-glossemes

9.1.3. Summary

The analysis presented in this section has revealed that the relations of purpose and anteriority are less often grammaticalized than the relations of causality and conditionality. The interesting question that arises is: what do these discrepancies in the degrees of grammaticalization tell us about the circumstantial relations themselves? Hopper and Traugott have remarked that:

Clause linkage markers are in their origins presumably motivated by speakers' desire to be clear and informative, particularly to give directions to hearers for interpreting clauses in terms of their linguistic environment. (2003:185)

Aikhenvald has also expressed an interesting thought related to this topic: “marginal semantic types of clause linking may not be expressed with clause linking devices at all” (2009:384). These claims draw our attention to the issues of pragmatics and, more precisely, communicative salience as a motivation for development of isomorphic marking. As Harris and Campbell notice:

A tension between the speaker's need for concise expressions and the hearer's need for redundancy and more elaborated expressions is often credited with causing change. (1995:53)

If this is the right track we could infer from the findings presented here that overall there is more pragmatic pressure on specialization of the marking of causality and conditionality than of the other two relations. The analysis of geographic distribution, however, give us some clues to consider the idea that the pragmatic pressure may differ in various geographic locations – some of these relations are better grammaticalized in certain areas than others. Are these geographic clusters accidental or are there some other factors that motivate this interesting variation? I go back to this question in chapters 10 and 11.

9.2. DEGREE OF LEXICALIZATION

9.2.1. *The notion of lexicalization*

The term *lexicalization* is broadly understood as “the adoption of a word into the lexicon of a language as a usual formation that is stored in the lexicon and can be recalled from there for use” (Bussmann 1996:276). In such sense the term “refers to the extent to which there are links between conceptual representation and syntax, and how

the nature of such links may be formalized” (Brinton and Traugott: 2006:18-20).⁹ A great majority of linguists perceive a lexicalized concept as expressible by a single word-sized unit (see, for instance, Levinson 2000, Lessau 1994, Moreno Cabrera 1998, Traugott 1994, Wischer 2000, Blank 2001). As straightforward as the basic definition may appear, there are two problems emerging from it. The first one is a definition of *word*. As Anderson aptly notices “there really is no satisfactory resolution to the problem of defining the term “word,” since it involves several mutually independent and sometimes conflicting criteria” (1985:4).¹⁰ The second problem concerns the definition of *lexicon* (or *inventory*, as it is sometimes called) which is seen as repository of content (lexical) information by some linguists and as repository of stored information (both grammatical and lexical) by others.¹¹ This, of course, makes the situation even more difficult since it requires distinguishing between lexical/content and non-lexical/grammatical/functional categories.

In order to overcome these problems, in this thesis lexicalization is understood as an adoption into the inventory of linguistic forms in a particular language of a form which, on the basis of criteria specified for that language, has been recognized as a single word regardless of whether it encodes a concept recognized as belonging to the traditionally understood content categories or functional/grammatical categories. In other words, any of the clause linking devices in my database that has been reported in the literature to be a single word is to be treated as a lexicalized item.

9.2.2. Data coding and analysis

The problem of the degree of lexicalization of clause linkers has gained some attention in the discussion on the motivations for the emergence of clause linkers of particular types. In his work on adverbial subordinators Kortmann (1997 and 1999) has put forward a hypothesis that, at least in the European languages, the more cognitively central a circumstantial relation is the more likely it is to be expressed in a variety of languages by a free-word (and more specifically, by a one-word) subordinator. In other words, the degree of lexicalization is, according to Kortmann, motivated by the cognitive salience of the relation of which the lexical item is an exponent. However, the fact that there are certain genetic/geographic tendencies for using lexicalized linkers in the world’s languages cast some doubt on the universal applicability of this assumption.

⁹ There is also a diachronic understanding of lexicalization as a process of semantic change where the original meaning can no longer be deduced from its individual elements (this includes various process of word formation, fusion and separation as discussed by Brinton and Traugott 2006:32-61). However, this understanding is not of relevance for our discussion here.

¹⁰ See also the remarks on the definition of word in section 3.1. and the references there.

¹¹ For Brinton and Traugott, for instance, “The output of lexicalization is a “lexical,” i.e., contentful item that is stored in the inventory and must be learned by speakers. The output is new or modified forms which are semantically contentful/ “lexical”, not functional/indexical/”grammatical”” (2006:98).

In section 8.1. it has been reported, for instance, that in Amerind, Indo-Pacific and Australian languages affixal c-glossemes are very common, while their frequency and significance is far smaller in African or Austric languages and they are completely absent from the two creole languages in the sample: Ndyuka and Kryiol. It has also been said that the preferences differ across certain groups of languages in a rather random way. For instance, the number of words encoding the relation of anteriority in India and along the Indian border is lower than for any other of the three relations. In North America, on the other hand it is not only anteriority but also purpose that in many languages is encoded exclusively by affixes which puts these two relations in direct contrast with the forms of linkers of causality and conditionality in those languages. These and similar observations posit the question whether, if we were to apply Kortmann's assumption, such differences suggest that in various languages (or various areas) the four circumstantial relations differ as to their cognitive salience. In order to address these problems, I propose to look at the cross-linguistic differences in the degrees of lexicalization and the geographic distribution of the degrees in a manner very similar to that presented in section 9.1. Before we move on to the analysis let me, however, first quote the entire passage in which Kortmann explains his approach:

The following three parameters will be made the basis for suggesting a layered representation of the internal structure of the domain of interclausal relations:

(i) the degree to which the individual interclausal relations are lexicalized in the languages investigated, i. e. the extent to which there exist (a) lexicalized adverbial subordinators at all, (b) one-word adverbial subordinators (including the possibility of optional complementizers), or (c) monomorphemic adverbial subordinators which express the given interclausal relation as their exclusive of primary meaning;

(ii) the degree to which the individual interclausal relations are coded, if at all, by means of adverbial subordinators across the languages in the project sample, irrespective of the distinction between primary and secondary readings;

(iii) the degree to which the individual interclausal relations are lexicalized as what has loosely been called "primary adverbial subordinators" (...), i. e. as one-word subordinators lacking membership in any other syntactic category.

Of these three, the first parameter, i. e. the availability of lexicalized adverbial subordinators for different degrees of morphological complexity in the European languages, is taken to be the most important one for judging whether a given interclausal relation belongs to the core or to the periphery of this semantic domain. The assumption underlying this parameter is that the more readily available a given interclausal relation is, i. e. the smaller the number of languages is in which it is not signaled by means of an adverbial subordinator, the closer to the core it needs to be positioned. (1997:139)

In this description parameter (i) refers then to the form of the linker, while parameters (ii) and (iii) concern what has been called by both Kortmann and me syntactic and semantic polyfunctionality respectively. Since there are only four relations in my sample, there is no need to consider all the three criteria of parameter (i) in order to reveal cross-linguistic differences in the degrees of lexicalization. I shall then focus on the third one – presence of a monomorphemic, free-word adverbial subordinator (which is, according to the author himself, the most important parameter), taking into account also the degree of semantic and syntactic polyfunctionality of the items. The

data presented here are, therefore, a subset of the data presented in section 9.1.¹² The difference lies in the fact that here the first element refers to a free-word monomorphemic marker and so the abbreviations should be understood as follows:

- **M(w)MM** – monomorphemic, free word, semantically and syntactically monofunctional c-glosseme;
- **M(w)M1** – monomorphemic, free word, semantically monofunctional c-glosseme, homonymous/polysemous with an item in one other syntactic category;
- **M(w)M2** – monomorphemic, free word, semantically monofunctional c-glosseme, homonymous/polysemous with items in two other syntactic categories;
- **M(w)1M** – monomorphemic, free word, c-glosseme encoding up to one additional circumstantial meaning and syntactically monofunctional.
- **M(w)11** – monomorphemic, free word, c-glosseme encoding up to one additional circumstantial meaning and homonymous/polysemous with an item in one other syntactic category;
- **M(w)12** – monomorphemic, free word, c-glosseme encoding up to one additional circumstantial meaning and homonymous/polysemous with a items in two other syntactic categories.

The analysis takes into account the same set of languages as in the investigation of degrees of grammaticalization and since lexicalization, similarly to grammaticalization, is viewed as a language internal process, loanwords are excluded from the analysed material.

The summaries reveal that the ranking of categories according to their degree of lexicalization in the sample is very similar to that discovered for degree of grammaticalization. Anteriority and purpose have a very similar level of degree of lexicalization, while conditionality is significantly more lexicalized than those two and is followed quite closely by causality. This order concerns all of the three sets of values: M(w)MM, M(w)M1 and M(w)M2 (see Fig.9.14.) As we might expect, there are no significant changes in the picture even if we include the items that are not only monofunctional semantically, monomorphemic free words but also those lexical and morphologically simple items which have been classified as covering two

¹² The additional, minor, criteria for distinguishing adverbial subordinators from other types of clause linkers which have been introduced in Kortmann's study (cf. section 3.2.8.) are omitted here. I have performed additional analysis considering that of Kortmann's criteria which seems to be the one which might influence the results quite significantly – the occurrence of the subordinators in finite clauses. The results obtained have not changed the picture emerging from the analyses presented in this section.

circumstantial relations (cf. Fig.9.15.). In the group of M(w)1M columns, in comparison to M(w)MM columns, we notice an increase in the number of items for conditionality which makes the distance in the degree of lexicalization between causality and conditionality smaller but still noticeable. The numbers for anteriority have not changed at all in comparison to the M(w)M group while for causality and purpose a noticeable increase has been observed only for M(w)11 and M(w)12 markers.

ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
M(w) MM	M(w) M1	M(w) M2	M(w) MM	M(w) M1	M(w) M2	M(w) MM	M(w) M1	M(w) M2	M(w) MM	M(w) M1	M(w) M2
3	10	13	22	29	30	5	13	17	16	23	25

(Fig.9.14.) Degrees of lexicalization – M(w)M markers¹³

ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
M(w) 1M	M(w) 11	M(w) 12	M(w) 1M	M(w) 11	M(w) 12	M(w) 1M	M(w) 11	M(w) 12	M(w) 1M	M(w) 11	M(w) 12
3	10	13	23	37	38	5	19	23	20	29	31

(Fig.9.15.) Degrees of lexicalization – M(w)1 marker¹⁴

9.2.3. Geographic distribution

On the maps depicting the distribution of degrees of lexicalization of anteriority and purpose (Fig.9.16. and Fig.9.17. respectively) we notice that the presence of M(w)M markers is highly restricted geographically. Interestingly, for anteriority it is the major languages in the sample (English, Polish, Hindi, Japanese, Thai, Mandarin) that contribute almost half of the free-word linkers depicted. All of these languages are classified traditionally as either isolating or flecnal and for all of them, as Bisang has noticed (1998, see section 8.1.), richness of free word adverbial subordinators is one of their defining features.

¹³ For full list of degrees of lexicalization of particular languages in the M(w)M group see Appendix VII.

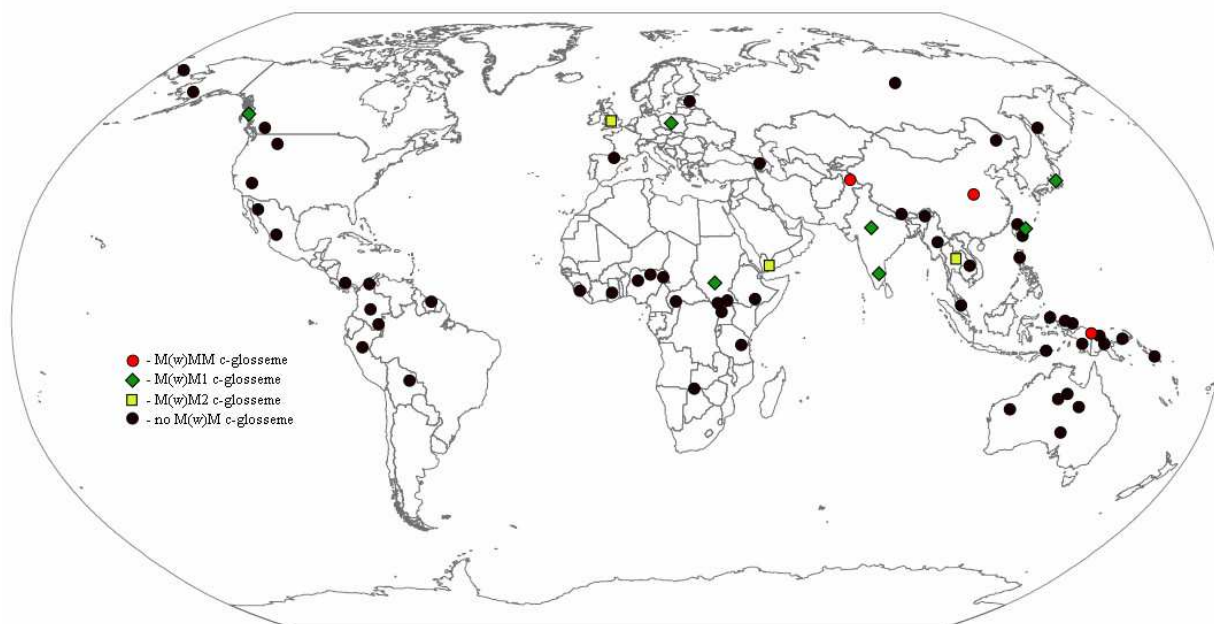
¹⁴ For full list of degrees of lexicalization of particular languages in the M(w)1 group see Appendix VIII.

Free-word monomorphemic purpose markers have a very similar distribution – with some more occurrences in Oceania and two examples in Australia. This clearly indicates a geographic phenomenon which have not emerged from the observation of degrees of grammaticalization presented in the previous section, where, recall, MM marking of anteriority and purpose was quite scattered on the map.

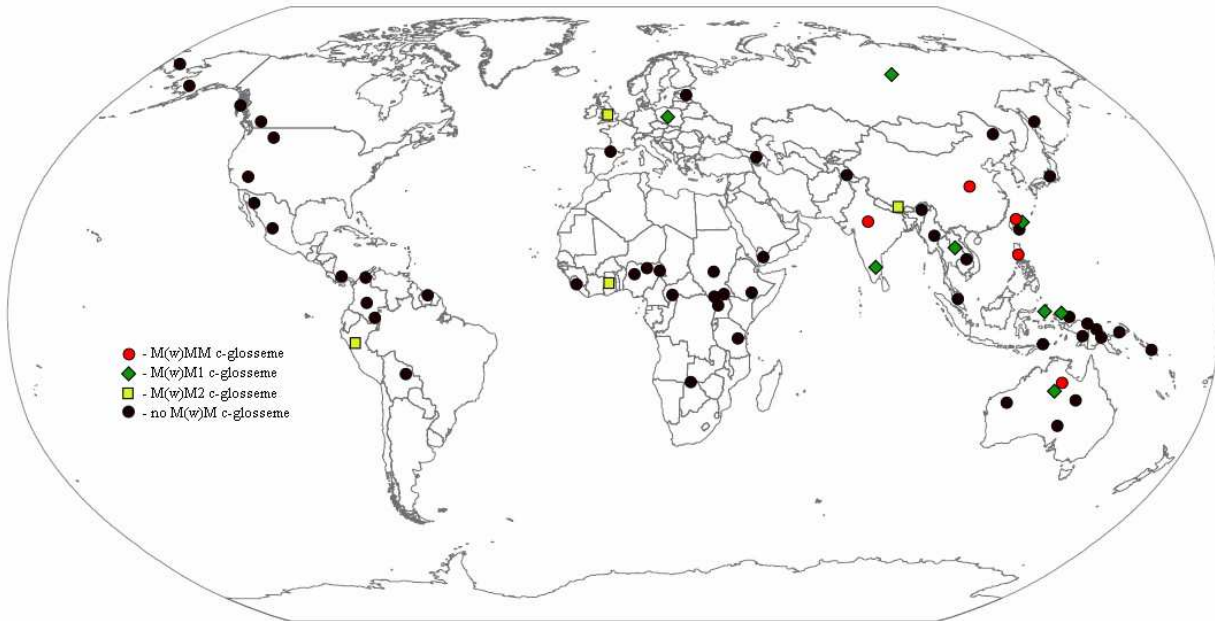
As for causality (Fig.9.18.), we notice that in all the European languages in the sample the relation reveals the highest degree of lexicalization and the aforementioned major languages also contribute either M(w)MM or M(w)M1 values. Apart from them there are four cases of M(w)MM causality markers in Africa, three in the Americas, two in Taiwan and single examples in other regions.

For the M(w)M marking of conditionality (Fig.9.19.) two clusters are noticeable – in West Africa and central and SE Asia (including, again, the major languages of Asia: Hindi, Thai, Mandarin). Among the four languages of Europe only English and Polish reveal lexicalization of conditionality at M(w)M level.

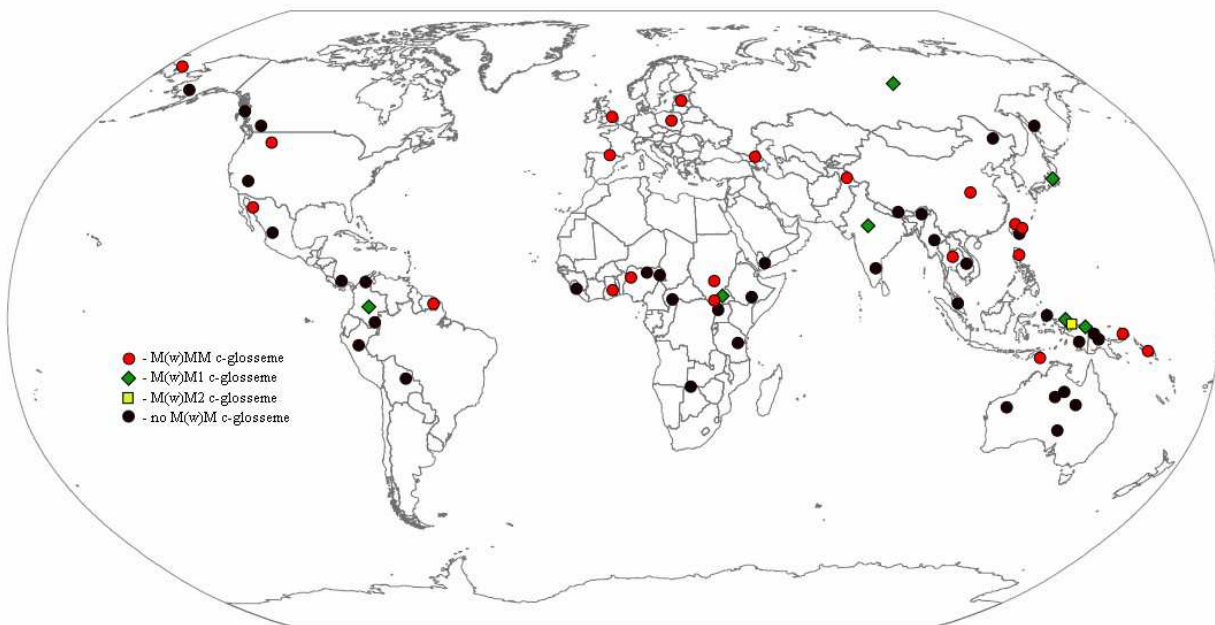
Interestingly, the only languages that have M(w)M marking available for all four relations are, with the exception of Mantauran Rukai, all major languages: English, Hindi, Polish, Thai and Mandarin.



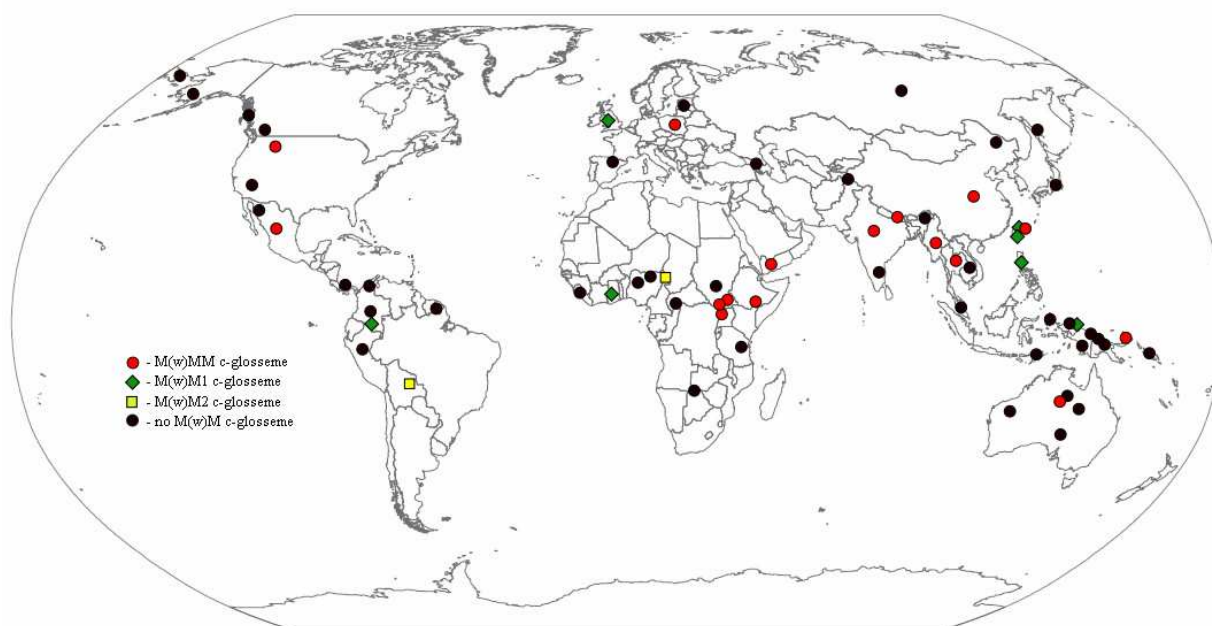
(Fig.9.16.) Geographic distribution of M(w)M anteriority markers



(Fig.9.17.) Geographic distribution of M(w)M purpose markers

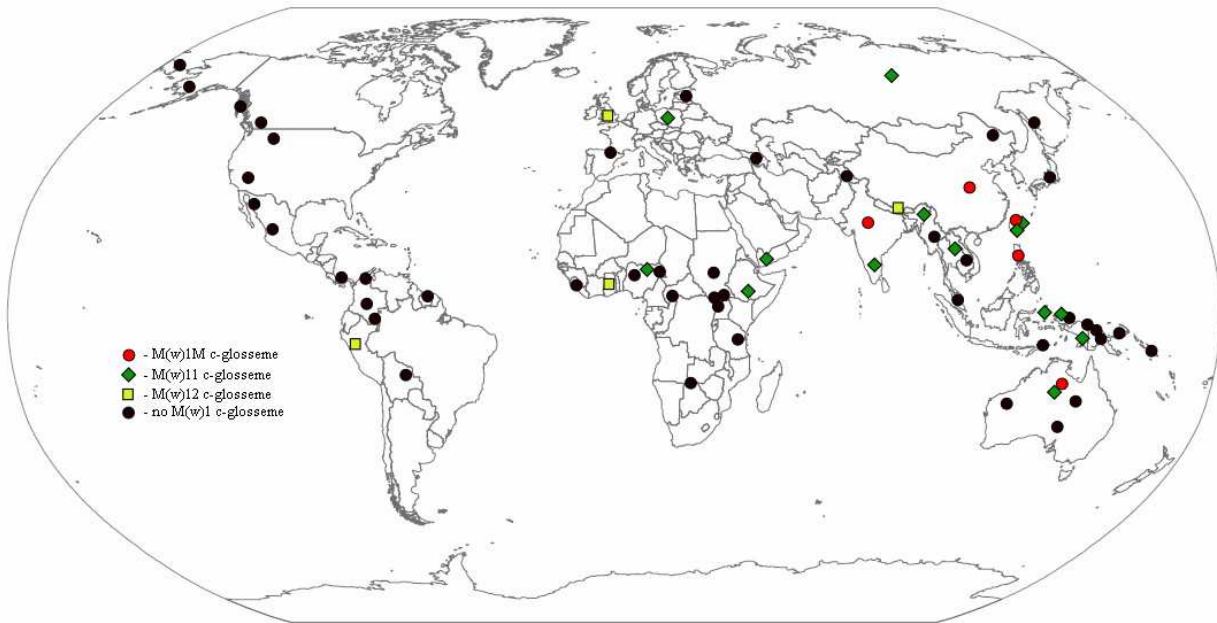


(Fig.9.18.) Geographic distribution of M(w)M causality markers

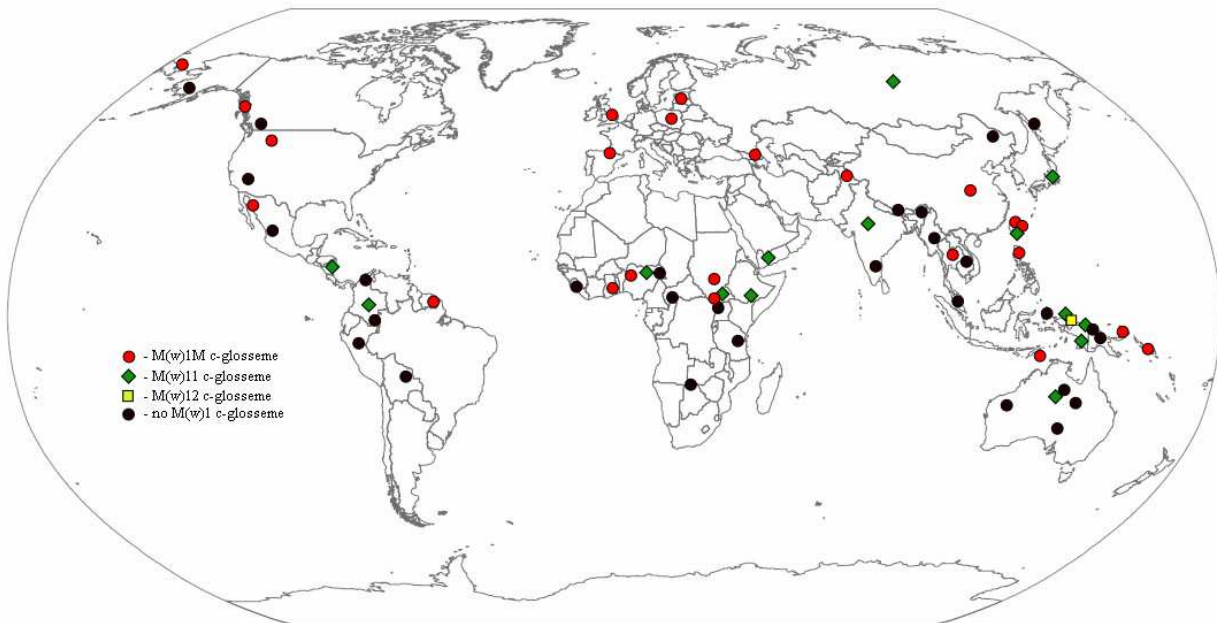


(Fig.9.19.) Geographic distribution of M(w)M conditionality markers

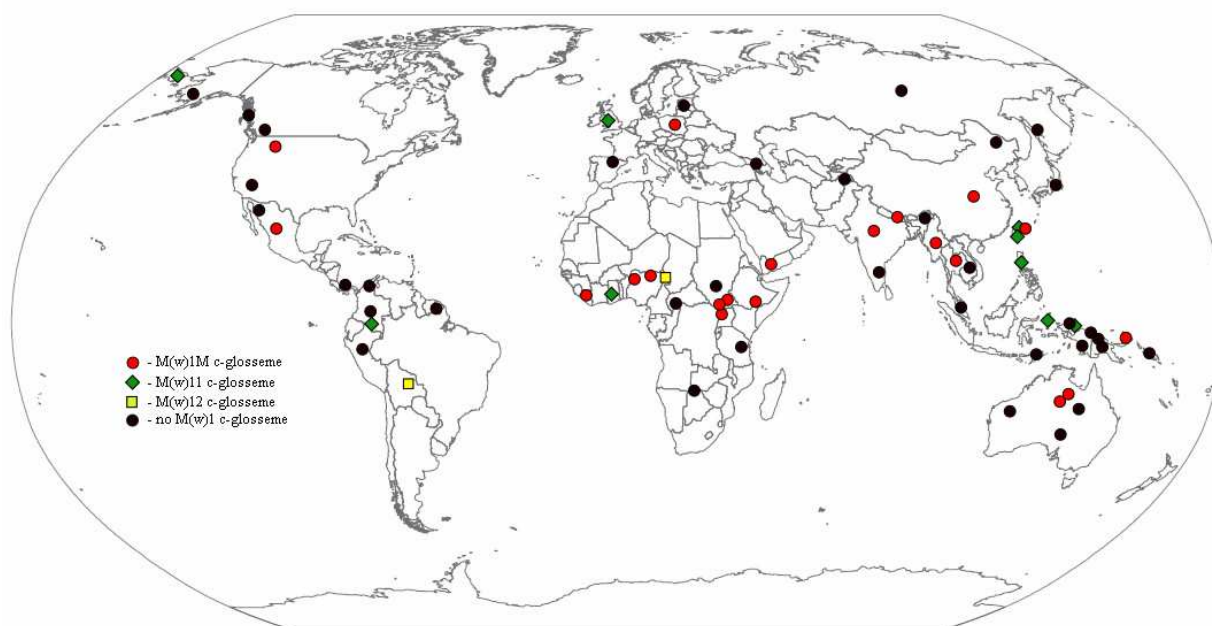
Comparing the maps in (Fig.9.16.-Fig.9.19.) with the maps presenting the datapoints for the monomorphemic free-word clause linkers which cover up to two circumstantial meanings (Fig.9.20.-Fig.9.22.) we still notice that certain geographic areas are more prone to the introduction of lexicalized exponents of the relations than others. As already said, the picture of distribution of M(w)1 markers of anteriority is exactly the same as for the M(w)M markers and so no further comments are required here. In the case of purpose, the picture is changed slightly (Fig.9.20.) by contributions from 6 languages – including two from Africa (Hausa, Konso) and three from Asia (Arabic, Yami, Galo). The datapoints for M(w)1 marking are scattered more for causality (Fig.9.21.) than for any other relation and, as already mentioned, the new contributions come almost exclusively from the M(w)11 and M(w)12 group. For M(w)1 markers of conditionality (Fig.9.22.) we observe an increase in comparison to the M(w)M marking of that relation (Fig.9.19.) especially in the languages of Africa which contribute 3 of the 6 new datapoints for M(w)11 markers (these languages are Swahili, Hausa and Gola).



(Fig.9.20.) Geographic distribution of M(w)1 purpose markers



(Fig.9.21.) Geographic distribution of M(w)1 causality markers



(Fig .9.22.) Geographic distribution of M(w)1 conditionality markers

9.2.4. Summary

We can conclude from the analysis and observations presented here that overall there are strong geographic patterns of lexicalization of clause linkers. Despite the fact that among the four relations it is causality that appears to be the most independent of geographic bias, it cannot escape our attention that the datapoints for the highest degree of lexicalization have been contributed in the majority of cases by analytic or moderately synthetic languages which mark functional categories by words rather than affixes. This supports my initial reservations about applying Kortmann's approach of treating the degree of lexicalization of clause linkers of particular relations as the indicator of cognitive salience of these relations to an investigation involving a world's sample of languages. If we were to apply his assumption, we would have to make a somewhat problematic and controversial conclusion: although overall it appears that in the internal structure of the domain of interclausal relations causality and conditionality occupy much more central position than purpose and anteriority, it seems that the cognitive salience of these relations differs across the globe. In other words, we would hypothesize that there are quite significant differences in the architecture of human cognition in various parts of the world when it comes to such basic concepts as those analysed here. Since such a scenario is very unlikely, I would not treat the degree of lexicalization as a universal parameter in the quest for cognitive salience of a particular

relations. I believe that it is the observations of degrees of grammaticalization that would suit this function better since it is free from genetic and/or geographic biases.¹⁵

9.3. DEGREE OF EXPLICITNESS

9.3.1. *The notion of explicitness*

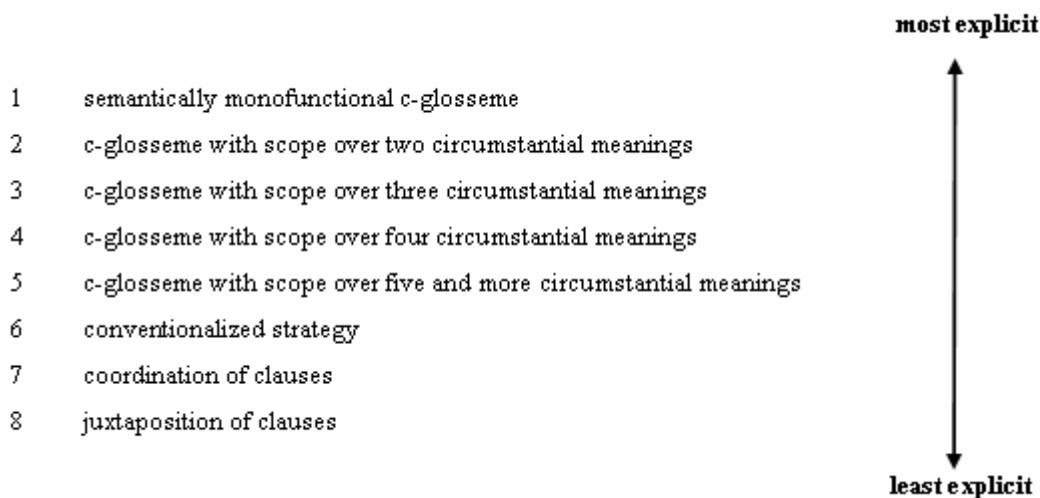
The common sense definition of *explicit* is “distinctly, unambiguously expressing all that is meant; leaving nothing merely implied or suggested; expressed”. When it comes to linguistics, the term and its antonym – *implicit* – seem to be most frequently used in the field of pragmatic theory. Grice (1975) used them in the sense of “what is said” and “what is meant” respectively. “What is said” for Grice is what is “closely related to the conventional meaning of the (...) sentence (...) uttered” and corresponds “to the elements of the sentence, their order and their syntactic character” (1989:87). “What is meant” relates both to what is meant by a speaker and what is inferred by the hearer. Explicitness is a notion commonly used also in syntax (cf. explicit subject, verbs, objects, relative pronouns) and morphology (cf. explicit case marking, tense marking, gender marking etc.) in various branches of linguistics (from historical linguistic through descriptive studies of languages, psycholinguistics, neurolinguistics, syntactic theories to computational linguistics). What is meant by *explicitness* in such cases is simply the surface presence of a marker which is contrasted with non-overt marking (absence of a marker).

However, since language is not a black-and-white universum where every element fits either one or the other category on many occasion the binary distinction explicit/expressed vs implicit/not-expressed is not sufficient and linguistic theorizing has to take into account the gradual, hierarchical nature of the phenomenon in language. This results in proposals of hierarchies of explicitness, some examples of which are: explicitness hierarchy of homophones (Simon and Wiese 2002), hierarchy of explicitness of reference forms in sign and oral language (Marschark et al. 2005) and hierarchy of explicitness in the paradigm (Cysouw 2003). In approaching the topic of explicitness of clause linkers and the broader phenomena of clause linking, a hierarchy is needed too.

¹⁵ I do not ignore the fact that both the analysis of degree of lexicalization and degree of grammaticalization reveal almost in all the elements of analysis presented so far the same ranking of relations (with purpose and anteriority clearly contrasted with conditionality and causality) but I postpone the discussion on this issue until chapter 11.

9.3.2. Data coding and analysis

The hierarchy I propose to use in order to fulfil the task I have set – assessment of the degrees of explicitness of encoding of circumstantial relations – is depicted in (Fig.9.23.) below¹⁶.



(Fig.9.23.) Hierarchy of explicitness of encoding circumstantial relations

While preparing the hierarchy, I was focused first and foremost on the question “what is said” and not “how is it said”. The forms of the marks on levels 1-5 do not play any role in assessing the degree of explicitness since it does not make a difference in a communicative situation whether a particular meaning is expressed by a free word, an affix, a clitic, a combination of free and bound morphemes or by a discontinuous marker. Similarly, it is not relevant whether a particular c-glosseme has polysemes/homonyms in other syntactic categories since these other functions do not influence the communicative power of its linking function in any way. What is relevant is the semantic scope of the c-glosseme, or in other words, its degree of semantic polyfunctionality in the space of circumstantial relations. Since a semantically monofunctional c-glosseme by definition expresses a particular meaning in an unambiguous way, it is placed at the top of the hierarchy. The lower down we go, the more ambiguous the c-glossemes are. Those capable of expressing five and more circumstantial meanings (labelled “general c-glossemes” in earlier chapters) occupy the last place in this group. Since the range of strategies of linking clauses is not limited to the use of clause linkers, as discussed in chapter 1, the hierarchy takes into account also

¹⁶ It needs to be stated here that the hierarchy I propose is concerned with phenomena only partially related to those considered in the well known continuum of explicitness of linking clauses by Lehmann (1988). The author was concerned first and foremost with properties which indicate not explicitness of expressing interclausal relations but the tightness of the link between two clauses in an asymmetrical relation (hence the importance of anaphors, connective adverbs and non-finite verb forms in his proposal).

those alternative means. Following the classification of strategies I have proposed in section 1.5.4., I include in the hierarchy also conventionalized strategies, coordination and juxtaposition of clauses. Conventionalized strategies are put at level 6, below c-glossemes and above coordination and juxtaposition. Since the reading they dictate is usually clear for the hearer, conventionalized strategies are more explicit than coordination but as the meanings are encoded by the structure itself and not by a dedicated linker, I view these conventionalized means as less explicit than c-glossemes. Group 7 of the hierarchy comprises coordinating structures which express the relations between SoAs in a very loose way nonetheless indicating, by the presence of the coordinating ‘and’ marker, certain degree of their mutual relevance. Finally, the bottom of the hierarchy is occupied by juxtaposition of clauses where neither lexical nor morphological or structural elements indicate the type of relation meant by the speaker.¹⁷

In order to assess the degree of explicitness of encoding for the four circumstantial relations this study deals with, for each of the languages for which the required data were collected, it has been established what is the highest level that its strategies occupy in the hierarchy. Importantly, two alterations have been introduced to the initial hierarchy presented in (Fig.9.23.) to adapt it fully to the purpose of this study.

The first of the alterations results from the fact that the relation of anteriority is commonly expressed cross-linguistically by the use of so called *connective adverbs* (or *discourse markers*) glossed usually as ‘then’ (sometimes in combination with a coordinating conjunction: ‘and then’), ‘afterwards’, ‘later’ etc. Such a marker, without any doubt, conveys in an explicit way the temporal relation between the two SoAs it links. However, since it is always attached to the clause encoding the temporally later SoA it does not qualify as a c-glosseme of anteriority clause (cf. section 1.4.1. and examples 9.1. and 9.2. below).

(9.1.) *He talked to Jack. **Then** he called his office.*

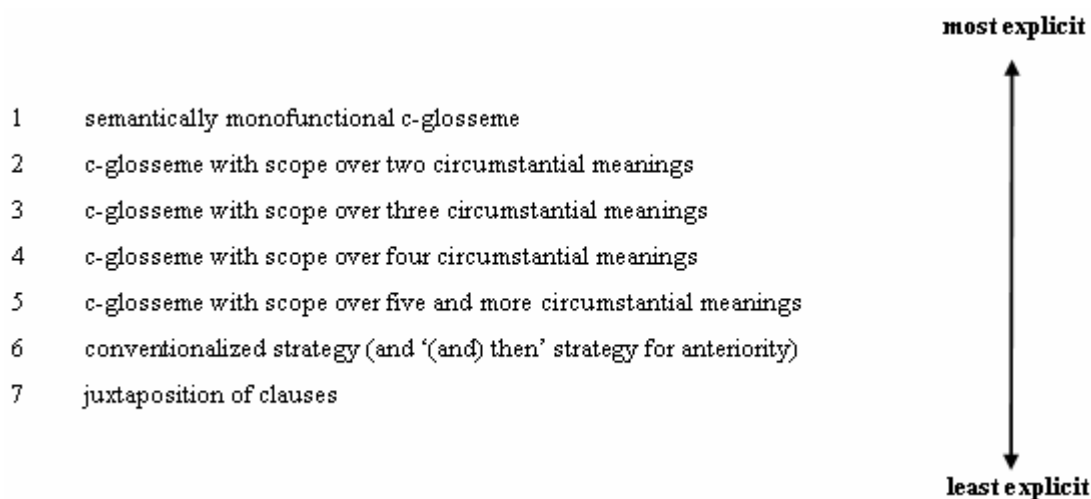
(9.2.) ***After** he talked to Jack he called his office.*

On the other hand, the presence of an ‘(and) then’ marker surely results in a higher degree of explicitness than the use of simple coordination. Hence, for anteriority I have

¹⁷ As with every theoretical proposal, the hierarchy proposed here is an idealization. It omits, for instance, the issue of intonation as a possible clue for inferring about the type of relation between the SoA meant by the speaker as well as the problem of iconicity (cf. sections 1.5.2.) and possibly also other elements of non-linguistic knowledge about principles governing the world as well as possible syntactic or contextual disambiguation of semantically polyfunctional markers.

placed the ‘then’ and alike markers together with conventionalized structures at level 6 of the hierarchy.¹⁸

The second alteration is the exclusion of coordination from the initial hierarchy. The rationale behind it is that in the analysed material none of the languages used coordination as the most explicit strategy for expressing the circumstantial relations – when no c-glosseme was available the languages used either juxtaposition or one of the conventionalized strategies. Thus, the hierarchy altered for the purpose of this thesis looks as follows:



(Fig.9.24.) Customized hierarchy of explicitness of encoding circumstantial relations

In the table below (Fig.9.25) the degrees of explicitness are listed for each of the 84 languages in the sample. Values 1-7 refer to the values listed in (Fig.9.24.). Marking of purpose has been analysed separately for same-subject and different-subject clauses and cells for missing values have been left empty.

What is of primary interest for us is the mean number of degree of explicitness for each of the relations: given at the end of the table: 2.42 for anteriority, 1.44 for causality, 1.59 for conditionality, 1.76. for same-subject purpose clauses and 1.77 for different subject purpose clauses. The relation with the lowest score – causality – is then the relation revealing the highest degree of explicitness. Conditionality follows it closely and purpose is not far away from the two. Anteriority, in turn, appears to be far less explicitly encoded. We need to remember, however, that the data for c-glossemes cover encoding on the semantically dependent clauses. If we were to treat the 14 cases of ‘then’ strategy of expressing anteriority listed in the table (cf. value 6 in Fig.9.25) as equally explicit to monofunctional semantically ‘after’ linker, the mean for the relation would be much lower – 1.53. This would put anteriority not only before purpose but also before conditionality. I return to this observation later on in this part of the thesis.

¹⁸ For the other three relations no strategies similar to the ‘(and) then’ strategy are available.

	ANTERIORITY	CAUSALITY	CONDITIONALITY	PURPOSE SAME-SUBJECT	PURPOSE DIFFERENT-SUBJECT	MEAN
Achagua	4	4	4	1	1	2.8
Akan	1	1	1			
Ama	3		3			
Apache	6	3	3	3		
Apurina	7					
Arabana	6	6	7	1	1	4.2
Arabic	1	2	1	2	2	1.6
Au	1	1	4	4		
Basque	1	1	1	1	1	1.0
Batak Karo	1	1	1	1	1	1.0
Baure	7	1	1	1	1	2.2
Boko	1	1	2	2	2	1.6
Burushaski	1	1	2	1		
Chukchi	1	1	2			
Cubeo	1	1	1	6	1	2.0
Dagur	1	3	1			
Didinga	1	1	1	1	1	1.0
Eipo	1	2	3	2	2	2.0
English	1	1	1	1	1	1.0
Estonian	1	1	1	1	1	1.0
Fur	2	1	2	1	1	1.4
Galo	1	2	1	2	2	1.6
Gola	4	1	2	1		
Hatam	6	1	1	1	1	2.0
Hausa	1	2	2	2	2	1.8
Hindi	1	1	1	1	1	1.0
Hualapai	5	5	1	5	5	4.2
Ika	6	1	1	1	1	2.0
Ilokano	1	1	1	1	1	1.0

I'saka	6	7	7	7	7	6.8
Jahai				7	7	
Japanese	1	1	1	1	1	1.0
Jingulu	7	6	2	1	1	
Kanuri	2	1	1	4	4	
Kayah Li	5	5	1	5	5	4.2
Ket	1	1	1	1	1	1.0
Khwe	1	1		1	1	
Konso	1	2	1	1	1	1.2
Krongo	1	1	1	1	2	1.2
Kryiol	1	1	1		1	
Lango	3	1	1	1		
Lavukaleve	1	1	3	1	1	1.4
Lepcha	1	2	1	1	1	1.2
Leti		1		1	1	
Lezgian	1	1	1	1	1	1.0
Lillooet	2	2	2	1	1	1.6
Maale	1	1	1			
Macushi			2			
Ma'di	6	1	1	1		
Mandarin	1	1	1	1	1	1.0
Mantauran Rukai	1	1	1	1	1	1.0
Meyah	6	1	2	1	1	2.2
Mayogo	1	1	1	1		
Mocovi	6	1	1			
Ndyuka	1	1	2	2	2	1.6
Nez Perce	1	1	1	1	1	1.0
Nisga'a	1	2	1	3	3	2.0
Nivkh	1	1	1	1	1	1.0
Paiwan	3		3		3	
Pitjatjantjara	2	2	5	1	1	2.2

Polish	1	1	1	1	1	1.0
Quechua Huallaga	4	4	5	1	1	3.0
Rama	2	1	2	1	1	1.4
Retuarã	1	1	1	1	1	1.0
SE Tepehuan	5	1	1	1	1	1.8
Sango	6	2	3	2	2	3.0
Santali	2	1	1	1		
Sapuan	7					
Seediq	6	1	1	1	1	2.0
Seri	5	1	5		6	
Shelha	1	1				
Supyire	6	1	2	1	1	2.2
Swahili	0	1	1			
Taba	6		2	1	1	
Tamil	1	1	3	1	1	1.4
Thai	1	1	1	1	1	1.0
Vitu	6	1	1	1	1	2.0
Wambaya	1	2	7	1	1	2.4
Warlpiri	1	2	1	1	1	1.2
Yami	6	1	1	1	1	2.0
Yanyuwa			3	1	1	
Yimas	1	5	5	6	6	4.6
Yindjibarndi	7	1	1	7	6	4.4
Yup'ik	1	1	1	1	1	1.0
TOTAL	191	111	124	123	115	
Number of languages with values present	79	77	78	70	65	
Mean	2.42	1.44	1.59	1.76	1.77	

(Fig.9.25.) Degree of explicitness of the project languages

The table presents also information on the mean degree of explicitness for each language for which information on the degree of explicitness of all four relations (including both types of purpose marking) was available. I wish to point out here one especially interesting fact – among the 17 languages which received the highest score of 1.0 in this ranking (put in bold in the table):

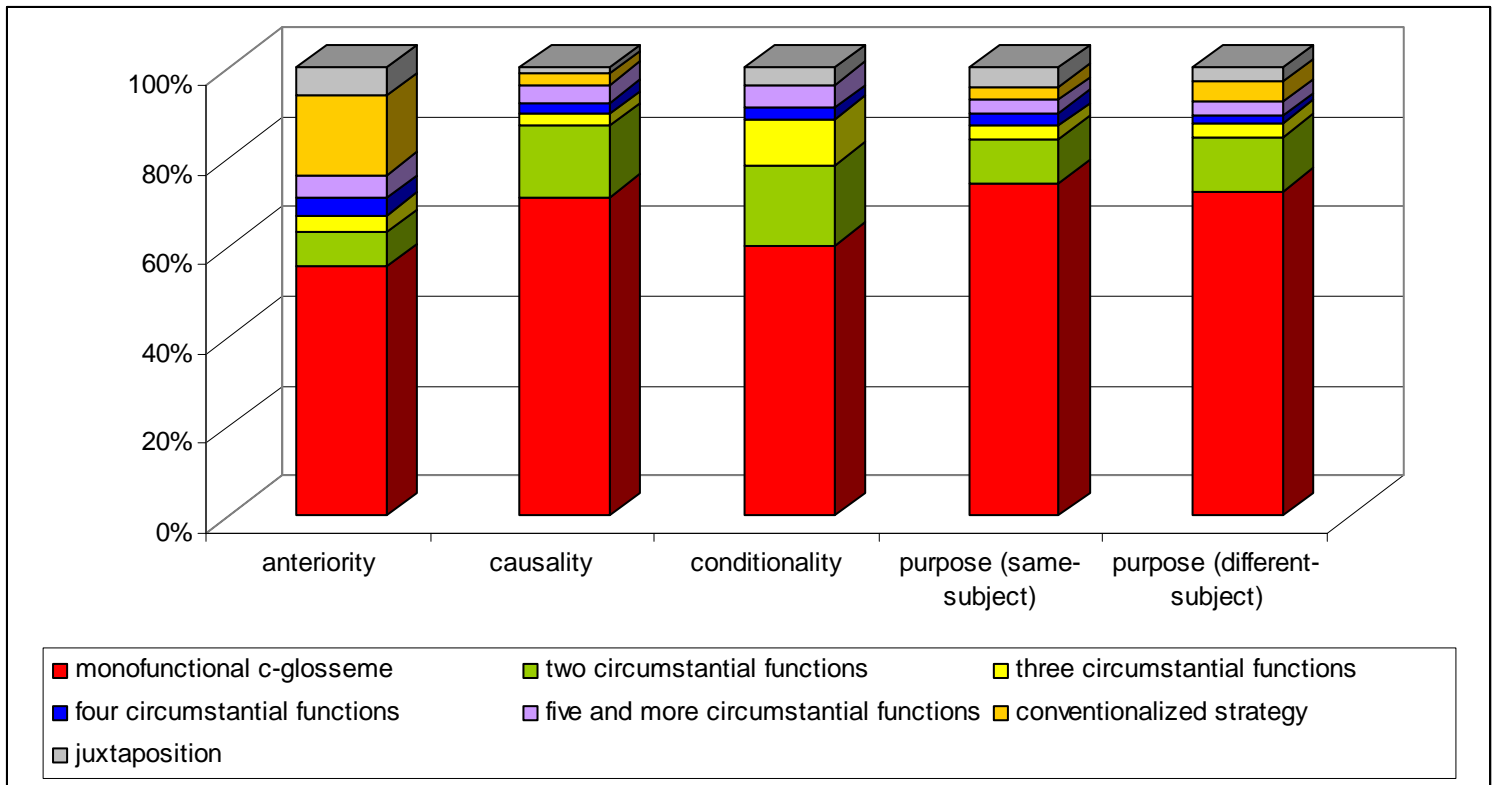
- 4 are European languages: English, Polish, Basque and Estonian (the first two being major languages);
- 5 are major Asian languages (Hindi, Japanese, Mandarin, Thai and Ilokano);
- the remaining 8 are languages which are under significant influence of world's major languages: Ket, Lezgian and Nivkh (influenced by Russian), Nez Perce and Yup'ik (under influence of English), Mantauran Rukai (influenced by Mandarin and Japanese) and Barak Karo (under influence of Malay).

I elaborate on the significance of these findings in chapter 11.

On the whole, as illustrated by the data in (Fig.9.26.) and the graph in (Fig.9.27), only over 55% of the languages encode anteriority in the most explicit way – by using a dedicated, unambiguous c-glosseme. The same is true about 60% of the languages when it comes to conditionality marking and over 70% for causality and purpose. This, of course, includes c-glossemes of various forms, internal complexity and various degree of syntactic polyfunctionality. The proportion of strategies other than c-glosseme is highest for anteriority and lowest for causality.

	ANTERIORITY		CAUSALITY		CONDITIONALITY		PURPOSE (SAME-SUBJECT)		PURPOSE (DIFFERENT-SUBJECT)	
	Valid %	Cumulative %	Valid %	Cumulative %	Valid %	Cumulative %	Valid %	Cumulative %	Valid %	Cumulative %
monofunctional c-glosseme	55.7	55.7	71.1	71.1	60.3	60.3	74.3	74.3	72.3	72.3
two circumstantial functions	7.6	63.3	15.8	86.8	17.9	78.2	10	84.3	12.3	84.6
three circumstantial functions	3.8	67.1	2.6	89.5	10.3	88.5	2.9	87.1	3.1	87.7
four circumstantial functions	3.8	70.9	2.6	92.1	2.6	91.0	2.9	90.0	1.5	89.2
five and more circumstantial functions	5.1	75.9	3.9	96.1	5.1	96.2	2.9	92.9	3.1	92.3
conventionalized strategy	17.7	93.7	2.6	98.7	0.0	96.2	2.9	95.7	4.6	96.9
juxtaposition	6.3	100.0	1.3	100.0	3.8	100.0	4.3	100.0	3.1	100.0

(Fig.9.26.) Cross-linguistic summary of the degrees of explicitness



(Fig.9.27.) Graphic summary of the degrees of explicitness

9.3.3. Geographic distribution

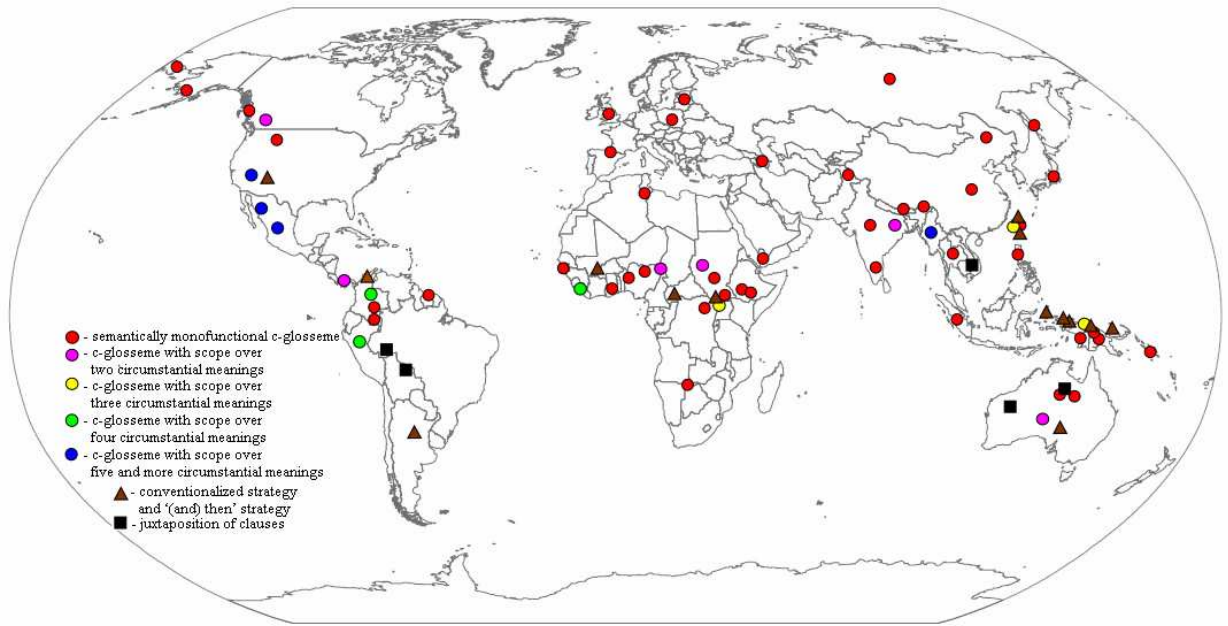
On the following pages a set of maps (Fig.9.28-9.32) depicting the distribution of the degrees of explicitness in the analysed languages is presented separately for each of the four relations.

As we can see from (Fig.9.28.) the regions which reveal the lowest degree of explicitness of anteriority are New Guinea, Oceania, Taiwan, South America and Australia. Quite low scores are to be found also in North America, where SE Tepehuan, Seri and Hulapai have at their disposal merely general c-glossems and Apache Jicarilla uses a connective adverb. The adverb is also the strategy of highest explicitness for three African languages: Ma'di, Supyire and Sango (cf. Fig.9.25.).

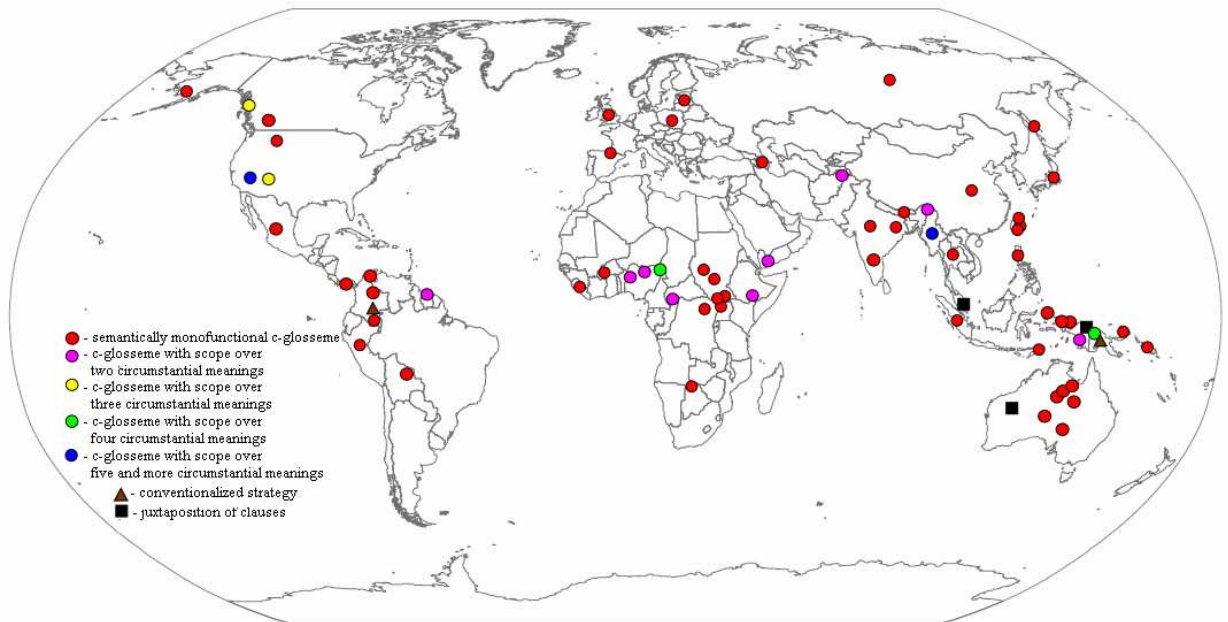
For both same- and different-subject purpose marking (Fig.9.29. and Fig.9.30.) the regions with the lowest degrees of explicitness are not so well defined but we can say, again, that the proportion of semantically monofunctional linkers (i.e. those occupying level 1. in the hierarchy of explicitness) is smallest for New Guinea, Africa and North America. South America and Taiwan, by contrast to the relation of anteriority, reveal quite high degrees of explicitness of same-subject purpose marking.

Causality, as already said, is overall very explicitly marked (see Fig.9.31.). There is just one example of juxtaposition (in Isaka), two of conventionalized strategies (Jingulu and Arabana) and three cases of general c-glossems (in Yimas, Kayah Li and Hualapai). The regions with the lowest proportion of semantically monofunctional markers are North America, New Guinea, Australia and Africa although it has to be emphasized that in all these regions the mixture of semantically monofunctional c-glossems and c-glossems with one additional function dominates over other strategies.

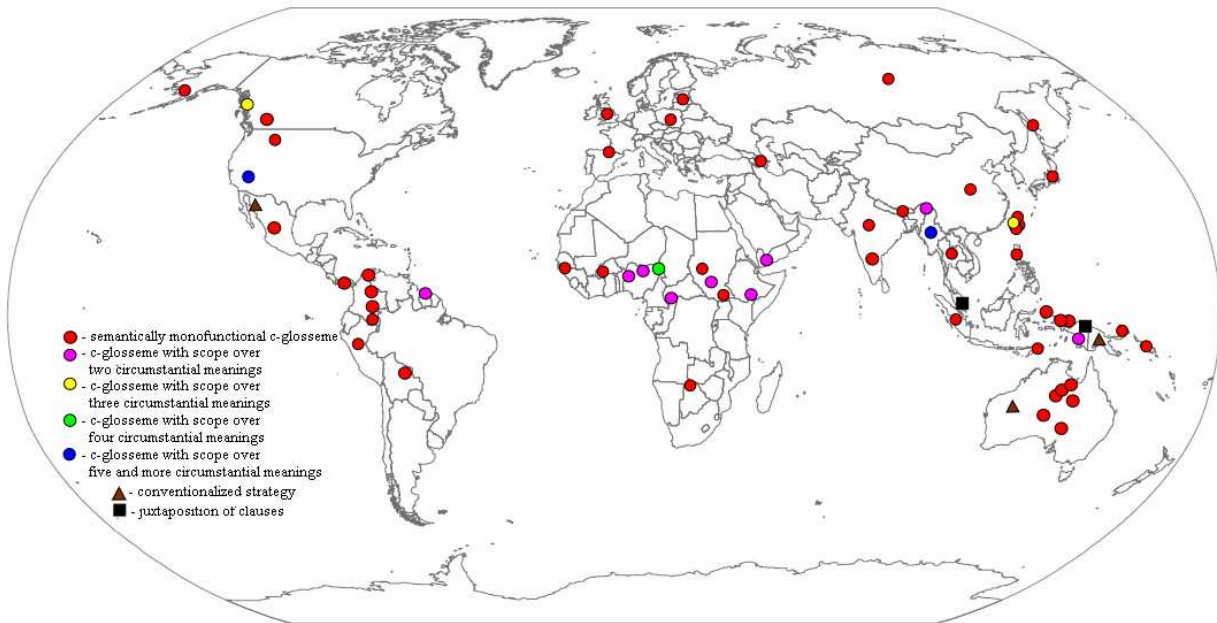
For conditionality (Fig.9.32.) we observe a very interesting mixture of degrees of explicitness in New Guinea and Australia – from juxtaposition in Isaka, Arabana and Wambaya, through general c-glossems in Yimas and Pitjantjatjara, quadri- and trifunctional markers in Au, Ama, Eipo and Yanyuwa (as well as Lavukaleve) to bifunctional markers in Meyah and Jingulu and monofunctional c-glossems in Hatam, Vitu, Yindjibarndi and Warlpiri. These two regions reveal the lowest degree of explicitness. In the Americas and Africa, with only some exceptions, the picture is dominated by a mixture of languages of level 1 and 2 of explicitness.



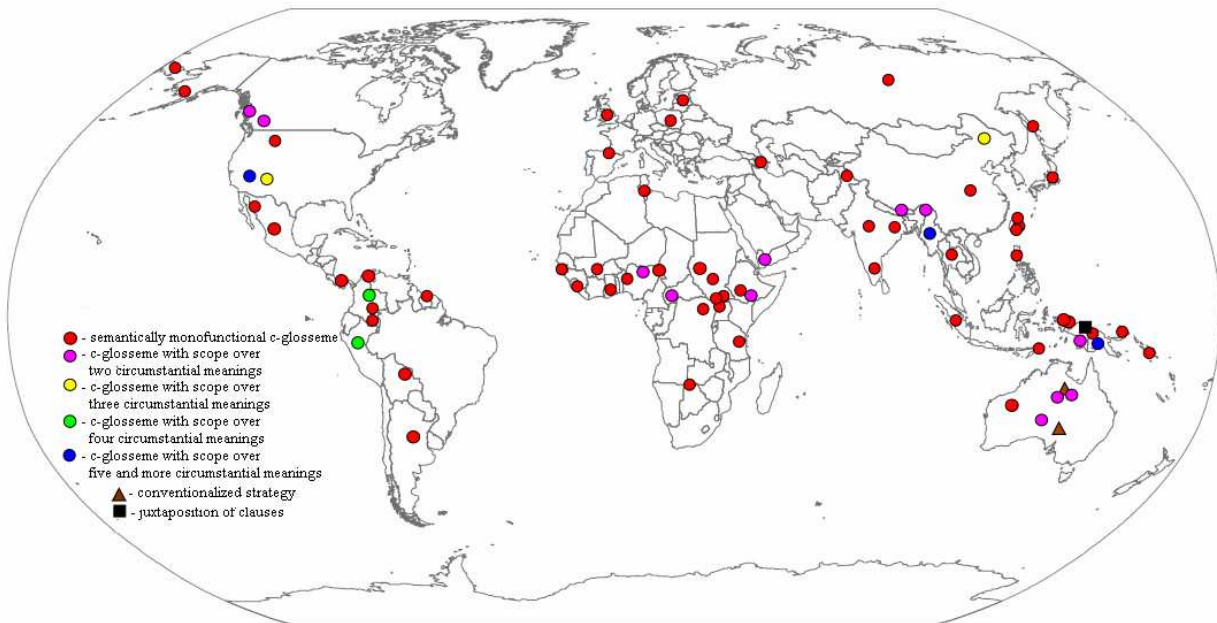
(Fig.9.28.) Geographic distribution of degrees of explicitness in encoding of anteriority



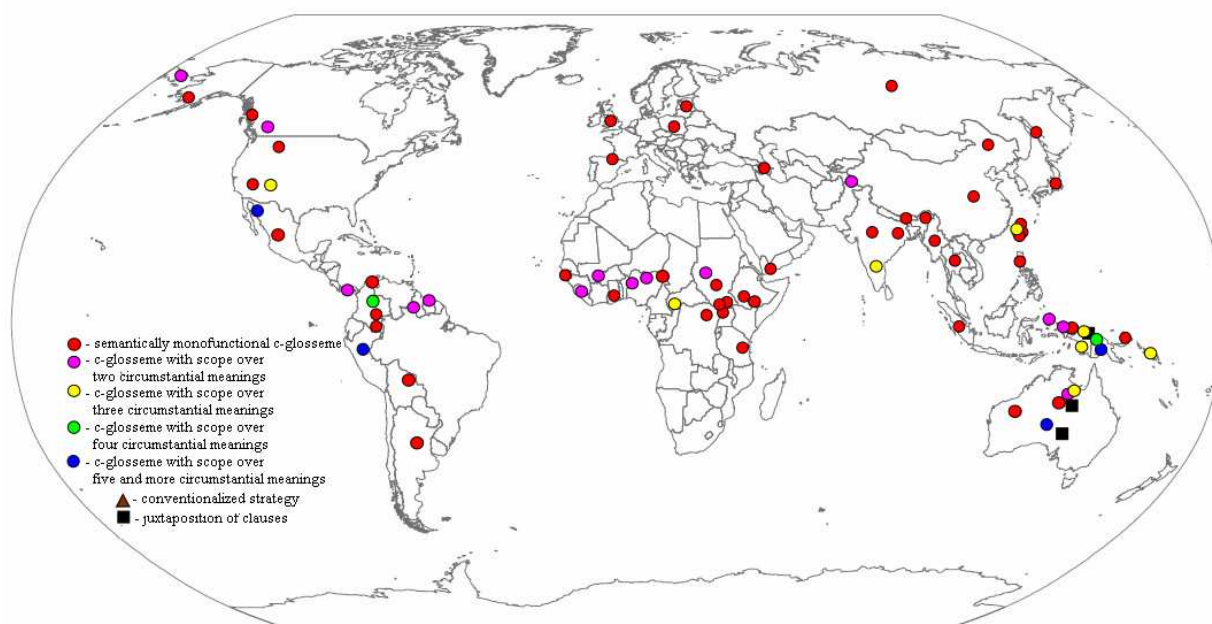
(Fig.9.29.) Geographic distribution of degrees of explicitness in encoding of purpose in same-subject clauses



(Fig.9.30.) Geographic distribution of degrees of explicitness in encoding of purpose in different-subject clauses



(Fig.9.31.) Geographic distribution of degrees of explicitness in encoding of causality



(Fig.9.32.) Geographic distribution of degrees of explicitness in encoding of conditionality

9.3.4. Explicitness and borrowing

One additional element that has not been discussed here in more detail so far and which is surely worth our attention, especially in the light of discussion on explicitness of encoding of circumstantial relations, is the issue of borrowing.

The presence of borrowed items among the c-glossemes in the sample (including borrowed c-glossemes incorporated in the structure of polymorphic markers) has been mentioned on several occasions in chapters of part II. It has also been said (see section 3.2.7.) that a recent study by Matras (2007) showed that coordinating and subordinating conjunctions are by far most susceptible to borrowing among the variety of grammatical categories looked at in his study and are, in fact, overtaken when it comes to loanwords only by nouns.

As for the motivations behind borrowing there are at least three important hypotheses. Hock argues that

The motivation for borrowing which perhaps most readily comes to mind is **need** [emphasis – A.M.]: if the speakers of a given language take over new technical, religious, etc., concepts, or references to foreign locations, fauna, flora, etc., there obviously is a need for vocabulary to ‘house’ these concepts or references. (1986:408).

He adds also that the spheres of vocabulary other than nouns referring to technology, names, artifacts are less commonly borrowed and

(...) seem to require more special motivations in order to be adopted from another language. The most important special motivation for this kind of borrowing is the notion of **prestige** [emphasis– A.M.]. (1986:384-5)

Harris and Campbell draw our attention to another type of motivation, referring specifically to clause linkers – **gap-filling**:

Some languages borrow precisely because they lack otherwise useful syntactic categories or constructions which they encounter in other languages with which they come into contact. More precisely, it has been claimed for several languages that they borrowed conjunctions and/or various subordinating devices only after and because they came into contact with other languages already possessing these things, seen as “gaps” in the grammars of the borrowers, thus explaining why they set upon acquiring the new material so rapidly when the notions became familiar to them from contact languages. (...) Needless to say, the notion of filling structural gaps is controversial and not supported by all scholars. (1995:129)¹⁹

In my sample, the highest number of borrowings has been recorded for conditionality – in 11 languages, and causality – in 10 languages. Purpose markers have been borrowed in 7 languages and linkers of anteriority – only in 3 (interestingly, the order of the relations listed here is the same as just discussed for the degree of explicitness). In total 17 languages have been found to borrow c-glossems from other languages: Basque, Burushaski, Gola, Hindi, Jahai, Ket, Khwe, Leti, Lezgian, Mocovi, Sango, Sapuan, SE Tepehuan, Shelha, Supyire, Swahili and Taba. These several scenarios of borrowing that emerge from this group are briefly discussed below.

The borrowed items are often introduced to languages in which no native marker for a particular type of relation was available. One of such examples is an Indonesian language Jahai which did not have a native conditionality marker and so acquired it from Malay (see example 9.3.) together with markers for other types of clauses – all three motivations listed above may be used as very likely explanations for borrowing in this and similar cases.

(9.3.) Jahai (Burenhult 2005: 136)

Kalɔw jeʔ crɔʔ jarej
{COND} 1SG be.hungry IRR.eat

‘I will eat if I’m hungry.’

In the African language Sango, the French *si* ‘if’ became to be used on its own or together with the semantically polyfunctional marker *tongana* (expressing the meanings of simultaneity duration, comparison and conditionality) resulting in a

¹⁹ For some examples illustrating such scenario and more on gap-filling see, for instance: Hale 1971; Heath 1978:115-16; Mithun 1980; Campbell and Mithun 1980; Hill and Hill 1981; Harris, Campbell 1995:129; Thompson, Longacre and Hwang 2007:207-209.

semantically monofunctional *si tongana* marker which is understood unambiguously as a marker of conditionality (Christina Thornell, personal communication).

Cases where borrowed items are used along with the native unambiguous markers have also been reported in the analysed literature. An example comes, for instance, from SE Tepehuan where the Spanish *porque* accompanies the native semantically monofunctional marker *na guß* apparently to add prestige:

(9.4.) SE Tepehuan (Willett 1987:51)

porque nagu' gu ja'oc vañdyagua', vañjuguia'
 {CAUSE} {CAUSE} the devil FUT.grab.me FUT.eat.me

‘(...) because the devil will grab me and eat me.’

Finally, in some languages the loanwords can be used as an alternative to the native, equally explicit markers – such a situation has been observed by Carlson for Supyire where the younger generation often uses a French borrowing in place of the native complex, yet unambiguous, causality marker *jàhá ná yε* (9.5.). Also here prestige seems to be the most likely explanation for the introduction of the French linker.

(9.5.) Supyire (Carlson 1994:581)

pàsige ηwɔhɔyi ba mÉ
 {CAUSE} fables it.is.not NEG

‘(...) because they are not (just) fables.’

As is already evident from some of the examples above, the introduction of a borrowed item does not always lead to the increase in explicitness of encoding of a particular relation. In fact, the presence of borrowed items alters the degree of explicitness of the analysed relations only in 9 languages: Burushaski, Jahai, Khwe, Leti, Sango, Santali, Sapuan, Swahili and Taba. Hence the overall results of the inclusion of these loanwords to the summary of degrees of explicitness are not very significant. As depicted in (Fig.9.31.) the mean degrees of explicitness after inclusion of borrowings are 2.37 for anteriority, 1.74 for purpose in same-subject and 1.74 in different-subject clauses, 1.41 for causality and 1.54 for conditionality. The increase is then, in comparison to the data excluding borrowings presented in the previous section, at 0.05 level for anteriority, conditionality and purpose in different-subject clauses, 0.03 for causality and 0.02 for purpose in same-subject clauses. More importantly, the borrowings do not change the explicitness of the relations strong enough to alter the anteriority < purpose < conditionality < causality order reconstructed from the data earlier. They add, however, 4 languages (Burushaski, Khwe, Leti and Swahili) to the group of the languages with the highest degree of explicitness listed in the previous section.

It seems that, at least in the analysed sample, the introduction of borrowed c-glossemes is as often motivated by prestige (or gap-filling) as by the need of adding to the inventory items which would express a given relation in an unambiguous way.

	ANTERIORITY	CAUSALITY	CONDITIONALITY	PURPOSE SAME- SUBJECT	PURPOSE DIFFERENT- SUBJECT
TOTAL SCORE	194	113	125	125	119
Number of languages with values present	82	80	81	72	72
Mean	2.37	1.41	1.54	1.74	1.74

(Fig.9.31.) Degree of explicitness of the languages in the sample with borrowings included²⁰

9.3.5. Summary

The analysis of degrees of explicitness presented in this section has shown that among the four circumstantial relations it is causality that is cross-linguistically most explicitly encoded. The relation of conditionality and purpose display slightly lower degrees of explicitness and it leaves no doubts that anteriority is the one least commonly explicitly marked on the semantically dependent clause.

Summarizing the observations of geographic distribution of the degrees of explicitness, as far as we may conclude from a study limited to the analysis of less than 2% of the world's languages and only 4 circumstantial relations, the languages with the lowest degree of explicitness are spoken in New Guinea and Australia. They are followed by the native languages of South and North America. African languages, are – on the whole – far more explicit but they are nowhere close the high level of explicitness displayed by languages of Europe and Mainland Asia. In the last two regions the languages depicted on the maps only occasionally show explicitness below the highest level. Finally, of all the four relations the distribution of the degrees of explicitness is most random for causality, while geographic clustering is most evident for anteriority and conditionality.

The presence of loanwords has also been considered as a factor that may influence the picture of explicitness. It has been found that, indeed, for some languages the introduction of borrowed clause linkers significantly increases the level of explicitness but at the same time inclusion of the borrowings to the analysis did not alter the overall picture of levels of explicitness of the analysed relations.

²⁰ A detailed list including all the languages is to be found in Appendix IX.

9.4. CONCLUSIONS

The overview of the three parameters presented in this chapter: degree of grammaticalization, degree of lexicalization and degree of explicitness has revealed an interesting picture. Causality has turned out to be the one cross-linguistically most highly grammaticalized, lexicalized and explicit. Together with conditionality it has to be placed in direct opposition to purpose and anteriority. The two latter relations only on some occasions reveal more significant differences between each other and in such cases it is usually purpose that overtakes anteriority (cf. levels of lexicalization for the M(w)1 group and levels of explicitness). It has to be remembered, however, that for encoding of anteriority languages often employ another, highly explicit strategy – use of connective adverb ‘(and) then’ (or alike) which acts in an anaphoric way – encoding the relation of anteriority being a part of the clause expressing the temporally later event. This type of encoding is not a subject of this thesis but, nonetheless, cannot be ignored.

It is worthwhile emphasizing that the markers of the two most highly grammaticalized, lexicalized and most explicit relations – causality and conditionality – are also more often borrowed than markers of the other two relations. A precise answer to the question whether and when the borrowing of these items is motivated by need, gap-filling or prestige would, however, requires a more detailed study.

While discussing the findings two important motivations for the introduction (development) of c-glossems to language systems have been mentioned – pragmatic pressure for the degree of grammaticalization (in section 9.1.) and cognitive salience for the degree of lexicalization (in section 9.2.). In both cases I have expressed my reservations about drawing strong conclusions merely on the basis of the numerical results. The major reason for that being the fact that for both these parameters (as well as for the degree of explicitness) quite strong patterns in geographic distribution of particular levels of these parameters have been found. Almost in every aspect of the analysis the regions with the lowest scores have been Oceania, Australia and Indonesia. The Americas have revealed slightly higher scores but were left behind Africa. Europe and Mainland Asia turned out to be the regions with the highest levels for all three parameters. The geographic patterns have surfaced most strongly for the degree of lexicalization which has not come as a surprise since, as it was already discussed in chapter 8, there are clear genetic/geographic preferences as for forms of encoding grammatical/functional concepts, including clause linkers. This has cast doubts on using the parameter of degree of lexicalization of clause linkers as a universal measurement of cognitive salience of the relations they encode. If this was the case, we would expect much more random distribution patterns. The link between degree of grammaticalization (and possibly also explicitness) and pragmatic pressure seems to be

a more convincing scenario to me since I am less hesitant to accept that the specifics of communicative situations may differ in various regions of the world than to accept that there are significant differences in cognitive architecture of human minds in various parts of the world. In order to extend the investigation of the differences in pragmatic pressure, it is necessary to move away from the discussion on the internal motivations for language change and to look at some external factors. If nothing else would argue for such direction of analysis, the fact that for all the parameters considered here it is the languages of the culturally quite similar Europe and Mainland Asia that obtained the highest scores, would be a good enough reason to investigate that territory. This is what I endeavour to do in the next chapter.

The final, brief, remark I wish to add here concerns the significance of the reported findings for the theory of semantic and lexical universals as presented in Wierzbicka's works (cf. especially 1996). Two of the concepts analysed in this thesis: causality ('because') and conditionality ('if') are viewed by Wierzbicka, Goddard and the adherents of the theory of semantic universals.²¹ The element of the theory which I am interested in here is the so called "strong lexicalization hypothesis". It states that "every semantic primitive meaning can be expressed through a distinct word, morpheme or fixed phrase in every language" (Goddard 1994:13). However, as shown in (Fig.9.25.), there are languages in my sample that have been reported to express conditionality and causality exclusively by juxtaposition of clauses. Since I do not have any reasons to not trust those who provided the descriptions of those languages and my language consultants, I believe these findings undermine the lexicalization hypothesis.

²¹ While discussing the concept of conditionality the author treats both real conditionals ('if') and hypothetical/counterfactual conditionals ('if...would') as two separate semantic and lexical universals. My definition of conditionality (cf. section 1.4.4.) falls under the first type. The concept of 'after' is considered by Wierzbicka in its general temporal meaning, including also adpositional use and so I shall not refer to it here. The concept of purpose is not viewed as semantically primitive at all.

CHAPTER 10

Influence of socio-cultural factors

It does not seem likely . . . that there is any direct relation between the culture of a tribe and the language they speak, except in so far as the form of the language will be moulded by the state of the culture, but not in so far as a certain state of the culture is conditioned by the morphological traits of the language.

Frans Boas

The connection between culture and grammar, unavoidably related also to cognition, has been a subject of interest for numerous linguists and anthropologists. The majority of the research in this field can be viewed as a part of Ethnolinguistics – a field of linguistic anthropology studying the relationship between language and culture and perception of the world in particular ethnic groups.

Ethnolinguistics encompasses the so called Sapir-Whorf hypothesis according to which the language of a specific group reflects that group's perception of the world: "users of markedly different grammars are pointed by the grammars toward different types of observations' arriving at 'somewhat different views of the world" (Whorf 1956:221). It also encompasses Ethnosemantics – a program outlined by Fraake in 1960's (see Fraake 1969 and Mathiot 1979) which is concerned with the referential meanings of linguistic expressions across cultures and languages, and Ethnosyntax, which, in the understanding coined by Wierzbicka (1979) and adapted and extended by Enfield (2002), is a study of relations between a language's morphosyntactic resources and the cultural knowledge, attitudes and practices of its speakers. Finally Ethnolinguistics embraces Ethnogrammar in the sense adopted recently by Everett who aims to investigate the potential impact that culture has on grammar i.e. the constraints that the first one puts on the second one (in all the domains: syntax, morphology, phonology, phonetics, and semantics).

In this chapter it is, however, not the link between the cultural traits of a specific ethnographic group and its language that is of interest, but a possible influence that

more general socio-cultural facts, such as population size, presence of written form, contact between groups of speakers and alike, have on the structure of language.

I begin in section 10.1. with a brief overview of selected theoretical and empirical works relevant to the topic of influence of general extra-linguistic factors on language structure. In section 10.2. I discuss the socio-cultural parameters of analysis used in this study. Section 10.3. reports on the correlations between the extra-linguistic parameters and the three parameters presented in chapter 9 aiming to answer the following questions:

- a) to what extent do socio-cultural factors influence grammaticalization and lexicalization processes of clause linkers in a language?
- b) how do they affect explicitness of a language in the domain of clause combining?
- a) is the impact of these factors the same on all of the circumstantial relations in a given language?

In the final section (10.4.) I discuss the results collectively drawing some general conclusions.

10.1. Hypotheses and previous studies

The most extensive work discussing the issue of influence of cultural traits on grammar is Perkins' monograph on deixis (1992). The author divided the analysed domain into a set of categories including: person, dual, inclusive/exclusive (of the addressee), demonstratives (signalling spatial coordinates relative to the speaker), and tense (specifying the relation between utterance and event time). He correlated the complexity of these systems in 49 languages (world-wide choice) with complexity of culture which, in turn, he assessed on the basis of 9 variables such as: the type and intensity of agriculture, regional organizations, craft specializations, class stratifications, size of the cities, inheritance of moveable property, population size etc. Perkins has found that in the investigated languages deictic markings on nouns and verbs are inversely correlated with the cultural complexity of the societies which they are spoken in. In other words – the less complex the culture, the more deictic distinctions are grammaticalized.

Since the publication of Perkins' work many other authors looked into the issue of correlation between culture and grammar. Many of them have aimed to analyse the relation from the point of view of linguistic complexity which, after decades of being absent in the mainstream discussion, has recently flourished in a number of

publications (see, for instance, Dahl 2004, Karlsson, Miestamo, and Sinnemäki 2008, Sampson, Gil, and Trudgill 2009 and references there)¹.

This section provides an overview of the works and findings related to the investigation of correlations between extra-linguistic factors and language structure which are most relevant for this study.

10.1.1. Society structure

In his work from 1976, focusing on deixis systems, Kay suggested that

In small, homogeneous speech communities there is a maximum of shared background between speakers, which is the stuff on which deixis depends. As society evolves toward complexity and the speech community becomes less homogeneous, speakers share less background information, and so one would need to build more of the message into what was actually said. (1976:119)

Extending the hypothesis over language in general, Trudgill in his numerous works (1989, 1996, 2004a,b) proposed three potential factors that may influence linguistic complexity: language contact, community size and network structure. According to the author:

Small, isolated, low-contact communities with tight social network structures are more likely to be able to maintain linguistic norms and ensure the transmission of linguistic complexity from one generation to another. Such communities are thus likely to be more linguistically conservative, i.e., to show a slower rate of linguistic change, and more likely to demonstrate complexities and irregularities. (...) Small, isolated, low-contact communities with tight social network structures will have large amounts of shared information in common and will therefore be able to tolerate lower degrees of linguistic redundancy of certain types. (2004a:306)

Similar claims have been repeated by Nettle (1999) and by Wray and Grace (2007). As the latter authors have hypothesized:

Languages that are used predominantly for esoteric (intra-group) communication tend to have features that are semantically and grammatically 'complex', while those used also (or even exclusively) for exoteric (intergroup) communication become 'simplified' towards rule-based regularity and semantic transparency. (2007:543)

Wray and Grace (2007) have also argued, importantly for this study, that the content of messages in languages spoken by coherent social groups (due to the fact that the contexts of communication is shared in such groups) tends to be implicitly expressed.

So far all the cross-linguistic studies that have attempted to verify the correlation between language structure and society structure have focused on just one element:

¹ The main reason for the absence of the topic of linguistic complexity has been the alleged implication of subordinate status of the less complex languages drawn from the discussions. This has been linked in a straight line with racism and political incorrectness. It was repeated by numerous researchers since then that neither an investigation into the correlations between culture, cognition and grammar nor the results obtained in studies presuppose or imply any inequalities between the speakers of various languages.

population size. Pericliev (2004) was the first to test Trudgill's (2004a) hypothesis on the influence of society structure on phoneme inventory in a sample of 417 languages. He has concluded that "there is no correlation of the kind [i.e. negative – A.M.] suggested by Trudgill between the size of a community speaking a language and the size of the consonantal inventory of that language" (2004:382). Hay and Bauer (2007) have also investigated the same relationship in a balanced sample of 216 languages and found that there is, in fact, a positive correlation on a statistically significant level. Defending his claims in response to Pericliev's paper, Trudgill (2004b) has argued that in his opinion effects of population size, network structure and language-contact situation need to be considered together and that there is no reason to expect a simple correlation between number of speakers and the number of phonemes in a language (2004b:386). However, as Sinnemäki (2009) aptly notices, it will not be easy to approach the proposal methodologically as speech communities consist of smaller groups and there are no methods of determining how to classify them.² In his own study, having tested a sample of 50 languages for a correlation between number of speakers and core argument marking complexity, Sinnemäki has obtained negative and statistically significant results for a number of different threshold sizes. Nichols (2009) also mentioned that in her cross-linguistic study on variation in grammatical complexity she correlated her results with population size and discovered, indeed, that a smaller population size favours higher complexity. However, as she has added, this can be a purely geographical phenomenon.³ In fact, the geographic effects, as well as the choice of thresholds for population sizes may be responsible for the lack of consistency in the correlations found by various researchers. Not mentioning the possibility that population size may be (accidentally or not) correlated with only some of the phenomena the researchers have looked at.

No study, to my knowledge, has looked at other aspects of influence of society structure on language structure. This concerns *inter alia* the phenomena of explicitness hypothesized by Kay, Trudgill and Wray and Grace.

10.1.2. Written form and other modes of displaced communication

The issue of influence of literacy on language structure has received a considerable amount of linguists' attention due to the easily noticeable differences between spoken and written (as well as formal and informal) varieties of particular

² C.f. the language situation in Russia where Russian is the official language but there are also dozens of other languages spoken in the country, with speakers often knowing more than one of them and therefore, being potential candidates for members of various speech communities.

³ Nichols adds also that although population sizes plummeted in the Americas and Australia on European contact and consequent smallpox epidemics, pre-contact protolanguages, as the comparative method shows, had similar grammars and lexicons to those that we find in these languages nowadays. This allows her to conclude that today's levels of complexity do not depend on today's population sizes and have in fact been stable despite drastic population size fluctuations.

languages. Miller and Weinert (1998) have even claimed that the terms “spoken language” and “written language” do not refer merely to different media but to partially different systems of morphology, syntax, vocabulary, and the organization of texts. Chafe (1994:49) has gone even further, proposing to divide the uses of language into three physically distinct types: language in thinking, speaking, and writing. One of the foremost researchers who indicated such discrepancies was, however, Kloss (1967) who argued that spoken language undergoes a process of reshaping in order to become a standardized tool of literary expression. The noticeable differences between natural and re-shaped forms lead him to the distinction between *Abstrandsprache* and *Ausbausprache*. The term *Ausbausprache* was defined as “language by development” since it has been shaped or reshaped, molded or remolded out of *Abstrandsprache* (“distance language”). A number of similar claims appeared in the linguistic literature since then – Hirsch (1977) has argued that writing establishes what he called “context-free language” and what Olson (1977) called “autonomous discourse”. A similar idea emerged also in sociology where Bernstein (1974) distinguished between “restricted linguistic code” and the “elaborated linguistic code” examining the middle and upper-class dialects in Britain. The restricted linguistic code, according to the author, can be at least as expressive and precise as the elaborated code in contexts which are familiar to and shared by the speaker and hearer but it has a formula-like quality and puts strings of thoughts together not in careful subordination but “like beads on a frame” (1974:134). Importantly, the elaborated code, is formed with a necessary aid of writing, and, for full elaboration, of print. Hence Bernstein’s “restricted” and “elaborated linguistic codes” have been often relabelled “oral-based” and “text-based codes” respectively.

Among all the domains of language structure in which the remodelling and elaboration on the way from spoken to written form has been most visible, is the domain of clause-combining. There is, for instance, a significant body of works investigating the influence writing has on the presence and frequency of dependent/independent clauses. This includes the structural distinction between finite and non-finite clauses with the former ones being viewed as much more typical for written form and sometimes even as conditioned by the presence of writing. For English much higher percentages of subordinate clauses in written texts have been reported *inter alia* by Poole and Field (1976), Chafe (1982) and Miller and Weinert (1998). Interestingly, Kroll (1977) has found not only that written narratives contained more subordinate constructions than spoken ones (35% and 14% respectively) but also that the opposite holds for coordinate constructions (25% in written and 40% in spoken

narrative).⁴ In response to the question of whether languages actually develop syntactic subordination as a result of the introduction of writing, Kalmar (1985) gives the example of Inuktitut where grammatical subordination emerges, apparently, as a result of literacy and translation from English. In a similar fashion, Givón (1979), too, has concluded that “certain types of languages – those which have only coordination (‘clause chaining’) but no subordination – are found only in preliterate ‘societies of intimates’”(1979:306).⁵

A subject closely related to the question of role of writing in the development and spread of clausal subordination is, of course, that of presence of clause linkers. The first one who pointed a relation between presence of writing and presence of clause linkers was Lichtenberk:

The development from implicit to explicit clause connections is...due to extralinguistic factors. Especially important in this respect is the introduction of writing – since writing is typically used for communication at a distance and since it lacks all the extra-linguistic clues, it requires high degree of explicitness and one way to achieve this is by means of specialized conjunctions. (1979:84)

Chafe (1987) has also pointed out that since written form is devoid of the degrees of bond between clauses which can be easily conveyed in oral narrative by rhythm and intonation, modulation of volume, pitch and speed, the writing system is forced to introduce some other means which may reflect these bindings such as clause-linking connectors. These claims keep with Olson’s earlier observations that “there is a transition from utterance to [written – A.M.] text both culturally and developmentally and that this transition can be described as one of increasing explicitness with language increasingly able to stand as an unambiguous and autonomous representation of meaning” (1977:258).

The fact that the transition is, as Olson has noticed, also developmental has been confirmed by numerous studies in first language acquisition which we cannot overview in detail here.⁶ What many of these works have pointed at, and what we should also emphasize, is that the especially important factor in the acquisition of structures typical for written form is the amount of exposure to text, most often in the form of formal schooling. The exposure is also important in broader perspective of stabilization of certain new patterns in speech. As Miller has noticed: “Speakers with long exposure to written text produce complex language in unplanned speech. And the more experience

⁴ Similar differences have been reported for some non-European languages. See, for instance, Maw (1974) for a study on Swahili and Mithun (1984) for an overview of proportions of dependent clauses in three native American languages: Mohawk, Gunwinguu and Kathlamet.

⁵ For more extensive discussion on writing motivating development of subordination see also Harris and Campbell 1995:308-312.

⁶ To name just some of them: Johnson and Chapman (1980), Bloom et al. (1980), Wing and Kofsky Scholnic (1981), Ansfield (1984), Bloom and Capadites (1987), McShane (1991), Golinkoff and Hirsh-Pasek (1995), Bloom (2006).

speakers have of unplanned speaking in formal situation, the more likely they are to produce complex language” (2006:481).⁷

The more elaborated forms, typical of written form, may, of course, spread also through other media of displaced communication – such as radio or TV. Their presence seems to be important not only for stabilization of certain patterns but also for the very process of remodelling spoken language. What I mean by that is that with the introduction of new media of communication in general – including writing, telephones, radio and TV – the content of the message is no longer supported by visual contact between speaker and hearer (cf. Greenberg 1971:87). This can be linked with Miller and Weinert’s (1998) finding concerning the proportions of adverbial clauses vary within a spoken form depending on whether there is or there is no eye contact. Since the authors have found that lack of eye contact results in a higher proportion of adverbial clauses it is not groundless to consider the possibility that forms of displaced communication other than writing, may also have more direct effect on the reshaping of language structure.

Despite the impressive amount of research in the field of differences between written and spoken language, so far the only cross-linguistic study known to me that has remarked (very briefly) on the influence of written form on language structure is Kortmann (1997) where we find one paragraph commenting on the correlation between the diversity of adverbial subordinators and not even written, but literary tradition:

The languages with a long literary tradition are also those with the most elaborate, semantically most differentiated inventories of adverbial subordinators, while those largely lacking a literary tradition, even when we include bound subordinators of subordinators operating exclusively over nonfinite constructions, have considerably smaller sets of subordinators for the specification of interclausal relations. (1997:256)

10.1.3. Contact

The final extra-linguistic factor (or, actually, a factor from the border of extra- and intra-linguistic factors) which is very often considered to be the ultimate reason for alterations in a language’s lexicon and its grammar is language contact. The effects of contact situations may be of triple nature: loss of features, addition of features or replacement of features.

The first and most important contact-derived mechanism leading to change in language structure is **linguistic borrowing** understood as the incorporation of foreign features into a group’s native language by speakers of that language (Thompson and Kaufmann 1988:37). In recent years a lot has been said and written on this topic. The findings most important for the subject of this thesis have already been mentioned in section 3.2.7. It would be beyond the limits of this overview to summarize the entire

⁷ For a similar point see also Chafe 1984.

body of literature on borrowing. Instead, I refer to the work by Curnow (2001) in which the author offers a comprehensive overview of our current state of knowledge on borrowing and its effects. Curnow classifies the categories that may be borrowed into: phonetic, phonological, lexical, grammatical, morphological, syntactic and discursive. The lexical group includes loanwords (i.e. lexical-form-and-meaning borrowings), borrowing of lexical forms only, borrowing of elements of structure of the lexicon – (c.f. loan homonyms and synonyms) as well as interjections and discourse markers and expressive word forms (such *as and so on*). The grammatical forms that may be borrowed are both free and bound grammatical form-and-meaning units (such as pronouns, complementizers, case affixes, classifiers etc.) as well as grammatical categories themselves (without lexical items – by reanalysis of existing forms or reorganization of the system).⁸ The remaining categories include borrowed positions in morphology, syntactic frames, order of constituents, clause linkage-strategies as well as genre types and organization of presentation of discourse. This, according to the author, gives us a full picture of the incredible impact that borrowing may have on the structures of languages.

While a typical borrowing is created by native speakers who import an element from another language, the new **elements** (including structural characteristics) may be also **imposed on a language**. The latter happens when non-native speakers of the receiving language have learned it imperfectly and incorporate their learners' errors into their own version of it. McWhorter (2001b and 2008) claims that in the cases of large-scale adult second language acquisition this very mechanism significantly affects language complexity. The scholar puts the idea explicitly:

I propose that heavy non-native acquisition is not merely one factor that can make a grammar drift into radical simplification, but that it is the sole factor ... Wherever complexity is radically abbreviated overall rather than in scattered, local fashion, this is not just sometimes, but always caused by a sociohistorical situation in which non-native acquisition of the language was widespread enough that grammar was transmitted to new generations in a significantly simplified form. (2008: 169)

This scenario is, of course, related to the claims about influence of population structure on linguistic forms discussed in 10.1.1. Trudgill himself (2004a), although not as strongly as McWhorter, argues that communities involved in large amounts of language contact (especially between those who are beyond the critical threshold for language acquisition) are likely to demonstrate simplification as a result of imperfect learning. A similar opinion has been expressed by DeGraff (2001) and Dahl (2004). The scale of contact, difficult to measure as it is (see next section), seems to be, indeed, a very

⁸ Similar distinction has been proposed in Matras and Sakel (2008) where the terms matter (MAT) and pattern (PAT) borrowing have been introduced. The former one refers to the borrowing of morphological material and its shape, the latter to cases where only patterns (organization, distribution) are replicated while the form itself is not borrowed. Borrowing of grammatical patterns is also called “diffusion” (c.f. Aikhenvald and Dixon 2001a).

important phenomena here. A small-scale second language acquisition is not expected to cause simplification. A good example of this is, as McWhorter (2008) has noticed, the contact between agglutinative Altaic languages and isolating Mandarin which did not lead to an analytic form of Altaic languages but to highly agglutinative hybrid dialects of Mandarin and Altaic. The author makes also an interesting point concerning the effects of literacy in cases where features of one language are imposed on another language due to imperfect second language acquisition. He argues that simplification by contact refers to those languages only which have not been standardized yet and in which literacy is not widespread. Those which have been taught in school through the medium of writing are not expected to simplify in contact situations:

Prescriptive tendencies exert a conservative influence on the written language regardless of how the language is actually spoken casually. Thus, for example, Russian's widespread usage across the former Soviet Union has had no simplificatory effect on the written language and very well may never do so. (2008:169)

It seems plausible, indeed, that literacy, standardization and prescriptivism may prevent languages from changes they would undergo had they existed in spoken form only (for this point see also McWhorter 2001a). Nevertheless, this idea would require proper examination. Importantly, the issue of speakers' attitudes seems to be not without significance since, as we know, they can be either barriers to change or promoters of change. Although in the majority of cases the more dominant (economically, politically, culturally) group exerts pressure on the more vulnerable community and its language, it is the case sometimes that the speakers simply refuse to acquire the new tongue and/or its elements (c.f. resistance of Montana Salish to English or Pirahã to Portuguese).⁹

Finally, we may also recall the phenomenon of **externally-motivated grammaticalization** discussed in section 3.2.7. as the third mechanism by which contact between languages affects their structures.

Having discussed the most important views and findings concerning influence of general elements of socio-cultural reality on language structure we may now re-focus on the domain of clause linkers and clause-linking addressing the research questions set in the introduction to this chapter.

⁹ In addition to borrowing characterized as an import of new elements to the recipient's language by its native-speakers and linguistic-shift where outsiders impose new features on their second language, there exists a borderline case where the very distinction between imported and imposed vanishes. It is known in linguistic literature as *bilingual first language acquisition*. It is reasonable to suspect, knowing the mechanisms of L1 acquisition, that in bilingual situations it may have a significant effect on complexification of languages. There has been, however, virtually no systematic research done in this area.

10.2. Parameters and design of the analysis

Having considered the difficulties of analyzing the potential correlations between the variety of cultural traits (such as those taken into account by Perkins) and likelihood of finding direct links between those traits and degrees of grammaticalization, lexicalization and explicitness of clause linkers, I have decided to restrict the scope of the study to the following main set of factors:

- level of written form development;
- presence and characteristics of the language in school teaching;
- presence and characteristics of radio broadcasting;
- presence and characteristics of TV broadcasting;
- number of speakers.

Since the factors are more often than not auto-correlated (e.g. languages with highest number of speakers are those with longest written tradition and vivid radio and TV broadcasting, as well as formal schooling), on numerous occasions more than one of them turns out to be correlated with the analysed parameters. Thus, in order to reveal more general tendencies I introduced also three additional parameters (discussed in detail below) which I label:

- type of society
- indicator of cultural complexity 1
- indicator of cultural complexity 2

The socio-cultural information necessary for encoding the differences between the analysed languages have been elicited from a variety of sources, the most important of which have been the introductions to grammars and grammar sketches and the questionnaires sent to the consultants. The other important resources I have consulted include the recent edition of *Ethnologue* (Lewis 2009), the *Dictionary of Languages* by Dalby (2006[1998]), *The World's Major Languages* by Comrie (1987) and the series *The Major Languages* (Comrie 1990). In addition to these, a variety of web pages dedicated to particular languages and speech communities have been consulted (see references for the list of these electronic resources). The type and quality of the information collected has allowed me to introduce only a limited amount of distinctions within each of the parameters. More detailed classifications would require far more detailed anthropological surveying – a task that could not be undertaken within the limits of this study. The types and characteristics of distinctions made for each of the proposed parameters are discussed below.¹⁰

¹⁰ For full list of values filled in for each of the languages see Appendix X.

Level of written form development

The (hypothesized) importance of written tradition for the development and elaboration of explicit clause linking strategies has been discussed in section 10.1.2. above.

For the purpose of this study I have classified the sample languages with respect to their level of written form development into six classes controlling for both the length and vitality of the written tradition since these two factors seem to be the most likely determinants of the potential influence of this parameter on language structure. The six classes are specified as follows:

1. no or negligible written tradition;
2. modest use of writing;
3. written language fully developed only after 20th century;
4. written language fully developed after 19th century;
5. rich written tradition before 19th century, fully developed after 19th century;
6. written language fully developed before 19th century.

According to this scale, a language for which orthography has been developed and for which primers and other reading materials have been published but which is not used in personal correspondence between speakers qualifies as a language of level 1. When the native speakers produce a limited amount of printed materials which are made available to be read by other speakers, such a language qualifies to group 2.

The four following groups (3-6) involve written communication qualified as “fully developed”. What is meant by that is the presence of not only books but, first of all, newspapers since it is newspapers that, being more readily accessible (both when it comes to amount and price), have more immediate impact on the dissemination of certain patterns of use and language structures. For a language to be classified as fully developed after 19th century it had to have its own newspaper printed before 1901. Group 5. applies to languages for which a newspaper was available before that time and in which written form was used commonly for all – personal and formal – communicative purposes by 1901. In the last group we find languages for which there were newspapers available prior to 1801 and which, too, used writing for all communicative purposes by that time.

Presence and characteristics of the language in school teaching

Presence of a particular language in school teaching has been included in the analysis due to its potentially significant role in the development and spread of new syntactic patterns including those which involve clause linkers. Without any doubt schooling in native languages is directly linked with the exposure to written form of those languages and with unplanned speaking in formal situation. Both these factors have been said to be important for the increase in production of complex language (see section 10.1.2). It

seems reasonable to assume then that the more a given language is used in school teaching and the more elaborate subjects it is used to communicate about, the greater the effect on complex sentence formation. Taking these assumptions into account for coding of this parameter four values have been used:

1. no school teaching in the language;
2. language taught only as a foreign language or as a language of instruction only in some schools (c.f. Nivkh);
3. language present as a language of instruction in first (and possibly also other) grade;
4. language fully present at all stages of education, including higher education.

Presence and characteristics of radio and TV broadcasting:

I use the presence and amount of radio and TV broadcasting, along with the level of written form development as indicators of the amount of displaced communication that a speaker of a given language is exposed to. The hypothesized importance of such exposure comes from the aforementioned finding of Miller and Weinert (1998) who discovered a correlation between lack of eye-contact and increase in the frequency of use of explicit adverbial subordinators in English. Again, considering the type of information on the characteristics of radio and TV broadcasting available for the analysed languages for the two parameters the following set of values has been used:

1. no broadcasts;
2. modest amount of broadcasting (occasional programs);
3. broadcasting fully present (stations broadcast in majority of the time in the native language).

Number of speakers (population size)

Aiming at verifying the hypotheses on the influence of society structure on the structure and explicitness of the language the society speaks (cf. section 10.1.1.), I propose to use the parameter of the number of speakers as giving a good estimation of the profile of society also when it comes to the other important elements that Trudgill and others pointed at: amount of contact and tightness of social network. I believe it is sound to assume that the higher the number of speakers the looser the network structure and the more contact with speakers of other languages. Of course, a practical issue of the problems in encoding the information required here is not without importance either. As mentioned before, the number of speakers is much easier to determine than either the amount of contact or the specifics of social network structure and so the former parameter is a natural candidate for being taken into account in the first place.

In order to limit the risk of the results being biased by a choice of a particular set of values for the coding of the number of speakers two threshold sizes have been applied and so two different parameters have been distinguished and labelled NoS(9) and NoS(5) respectively:

NoS(9) – designed using an logarithmic order of magnitude with 10 as a base

1. 1-10 speakers;
2. 11-100 speakers;
3. 101-1,000 speakers;
4. 1,001-10,000 speakers;
5. 10,001 – 100,000 speakers;
6. 100,001 – 1,000,000 speakers;
7. 1,000,001 – 10,000,000 speakers;
8. 10,000,001 – 100,000,000 speakers;
9. over 100,000,001 speakers.

NoS(5) – designed using an logarithmic order of magnitude with 100 as a base

1. 1-100 speakers;
2. 101-10,000 speakers;
3. 10,001-1,000,000 speakers;
4. 1,000,001-100,000,000 speakers;
5. over 100,000,001 speakers.

The values used in the analyses refer to the numbers for the total number of population of all countries in which the language is spoken. If no information has been provided by the grammars or consultants the numbers have been taken in the first place from the *Ethnologue* (2009) and in such cases the numbers referring to the population of speakers and not to the ethnic population size have been adapted.

Type of society

In anthropological studies, researchers use a number of traits that help to distinguish between various types of societies. Perkins (1992) applied in his investigation a variety of variables adapted from Hays (1978) but in my work, due to the lack of up-to-date information on a number of speech communities whose languages constitute this sample, I have decided to use three very general values for this very complex parameter:

1. predominantly non-urban society;
2. mixed society;
3. predominantly urban society.

By “mixed society” I understand the speech communities where a significant number of speakers do not have access to media and modern life style (this applies, for instance to speakers of Hindi and speakers of various native American, Australian and African languages)¹¹. Importantly, however, each of these three variables may be viewed as a good estimations of the general level of cultural complexity. For instance, a language spoken in a predominantly non-urban society is likely to be used by a relatively small number of speakers, not taught in schools or taught only to a limited extent, with rather poor literary tradition and hardly used in other forms of displaced communication. Since all these factors have been assumed to have some influence on the development of c-glossems and since, as said before, it is not possible to separate these factors as they are autocorrelated, the set of values considered here may be viewed as referring to more general socio-cultural profile. It is hoped that using the parameter of ‘type of society’ in the correlation analysis allows us to get closer to answering the question whether extra-linguistic factors in general have some impact on language structure. Of course, one has to remember that the set of values proposed for this parameter is very coarse-grained. This is why two other parameters, as discussed below, are used to support the more general observations.

Indicator of cultural complexity 1 (or shortly: Indicator 1)

This parameter (as well as the next one described below) has been designed specifically to give an estimation of the significance of socio-cultural factors influencing a particular language. It is a combined measurement of all the 6 traits described above: level of written form development, number of speakers (NoS9), presence of the language in school teaching, presence in radio and TV broadcast and type of society. For each language the numbers for the values for each of these traits have been added and the score has been divided by the maximum score that can be obtained from these values: 28. The highest value that may be obtained in this classification is 1 – for a language which displays highest values for each of the components of this measurement. The parameter can be then treated as a parameter normalizing, to a certain degree, the encoding of the socio-cultural profiles. There are two aspects of this normalization. Firstly, by considering a group of parameters rather than just one of them we obtain a general picture of the influence of extra-linguistic factors on language structure. Secondly, the risk of the results of analyses being biased by the set of values chosen for encoding of the component-parameters is reduced since it is the mean score that is taken into account.

¹¹ The phrase “significant number of speakers” is, yet again, a very general one.

Indicator of cultural complexity 2 (or shortly: Indicator 2)

Designed with a similar aim in mind and following the same principles as Indicator 1, but with a smaller set of components including: level of written form development, number of speakers NoS(9) and type of society. Again, the reason behind the choice of the component-parameters is the reduction of the possible bias resulting from the choice of the set of parameters and their values. In contrast to Indicator 1, for Indicator 2 I have only considered those parameters which, according to the hypotheses presented in section 10.1., are most likely to directly influence the structure of languages.

One of the factors that did not make it to the set of parameters discussed here is language contact (recall, however, that the influence of borrowing on explicitness of clause-linking has been discussed in section 10.3.4.). The main reason for its absence is the difficulty in proposing a classification of its levels. Thomason and Kaufman (1988), for instance, have proposed a five-point scale of intensity of contact: casual contact, slightly more intense contact, more intense contact, strong cultural pressure and very strong cultural pressure. It is however, not clear on what basis one assesses the intensity and from where to source the necessary information from. Initially, I have attempted to introduce a set of values which would take into account the types of languages in contact. I assumed that the effects of the influence of world's major languages, due to their elaborateness in clause-combining, may be more severe than those of the other languages and I distinguished therefore between three contact situations:

- a) monoglossic situations without severe influence of other languages;
- b) polyglossic situations involving neighbouring languages which are not the major world's languages;
- c) polyglossic situations involving major world's languages.

After categorizing the languages according to these distinctions it turned out, however, that the sample was so skewed towards languages in group c) that no reliable effects of the factors on the analysed parameters could be obtained and so language contact has been excluded from the analysis.

10.3. Results of the analysis

In this section of the chapter results of the statistical analysis of correlations between the variables listed in the previous section and the three parameters described in chapter 9: degree of grammaticalization (section 10.4.1.), degree of lexicalization (section 10.4.2.) and degree of explicitness (10.4.3.) are presented.

The correlations have been measured using non-parametric Spearman's correlation coefficient (Spearman's rho). It goes without saying that the results obtained, as with any statistical results, should be viewed as giving clues about causation but should not be overinterpreted as actual evidence for cause-effect scenarios in the analysed cases.¹² The tables included in the following sections show, for clarity of presentation, only the significant correlations.

10.3.1. Correlations with degree of grammaticalization

In the table below (Fig.10.1.) results of correlation analyses for the sample of 67 languages described in section 9.1. have been listed for three groups of monomorphemic markers: MM, M1 and M2.¹³ The abbreviations MM and M1 have been explained in section 10.1.1. The third one, which is included here to broaden the analysis even further, refers to monomorphemic c-glossems encoding up to two additional circumstantial meanings and within this group, again, three subgroups are distinguished:

- M2M – monomorphemic c-glossems encoding up to two additional circumstantial meaning and syntactically monofunctional;
- M21 – monomorphemic c-glossems encoding up to two additional circumstantial meaning, homonymous/polysemous with an item in one other syntactic category;
- M22 – monomorphemic c-glossems encoding up to two additional circumstantial meaning, homonymous/polysemous with an item in one other syntactic category.

It has been found that in the 9x9 matrix of correlations (9 socio-cultural factors and 9 grammaticalization groups) only six of the 81 cells are filled in for causality and only three for purpose. With one exception all these correlations concern the type of society.

¹² Full lists of correlations for each of the aspects of analysis are presented in Appendices XI – XIII.

¹³ The smaller sample has been used for the same reasons of normalization as explained in section 9.1. The results of analyses for all the relations including all languages without missing data have not showed, however, any significant differences from the data presented here.

For anteriority, by contrast, numerous correlations have been found – including correlations with all three modes of displaced communication: written form, radio and TV as well as with schooling. No correlations have been found, however, at the MMM level (and neither at the M1M or M2M level). As for the number of speakers – only one case has emerged but at a low level of significance. At the same time several correlations with type of society and indicator 1 and 2 suggest strongly that the socio-cultural factors have an effect on the degree of grammaticalization of anteriority.

The degree of grammaticalization of conditionality reveals, unequivocally, the strongest correlations with the 9 factors – this includes even the MMM level. Especially highly significant results have been obtained for the level of written form development and number of speakers (both parameters) which is also reflected in the scores for type of society and both indicators of cultural complexity.

.

Relation	Degree of gram.	Socio-cultural factors								
		Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2
ANTERIORITY	MMM									
	MM1	.313		.285						
		.010		.019						
	MM2	.346	.267	.355	.313	.241		.279	.303	.266
		.004	.029	.003	.010	.049		.022	.013	.029
	M1M									
	M11	.323	.269	.282					.25	
		.008	.028	.021					.037	
	M12	.336	.299	.374	.257			.327	.336	.275
		.005	.014	.002	.036			.007	.005	.024
	M2M									
M21	.302	.241	.304					.268		
	.013	.050	.012					.028		
M22	.336	.299	.374	.257			.327	.336	.275	
	.005	.014	.002	.036			.007	.005	.024	
CAUSALITY	MMM							.325	.266	
								.007	.030	
	MM1							.265		
								.030		
	MM2							.245		
								.046		
	M1M							.302		
								.013		

	Degree of gram.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
	M11									
	M12									
	M2M							.302		
								.013		
	M21									
	M22									
PURPOSE	MMM							.241		
								.050		
	MM1									
	MM2									
	M1M							.241		
								.050		
	M11									
	M12									
	M2M							.241		
								.050		
M21										
M22										

	Degree of gram.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
CONDITIONALITY	MMM	.352	.334	.267	.263	.412	.339	.352	.446	.441
		.003	.006	.029	.031	.001	.005	.003	.000	.000
	MM1	.293	.304	.301	.276	.390	.373	.341	.416	.403
		.016	.012	.013	.024	.001	.002	.005	.000	.001
	MM2	.301	.280	.276	.279	.346	.336	.281	.364	.357
		.013	.022	.024	.022	.004	.005	.021	.002	.003
	M1M	.339	.332	.256	.276	.410	.326	.384	.459	.452
		.005	.006	.037	.024	.001	.007	.001	.000	.000
	M11	.277	.273	.260		.366	.333	.317	.410	.403
		.023	.025	.034		.002	.006	.009	.001	.001
	M12	.294	.256	.242	.246	.329	.304	.261	.367	.366
		.016	.037	.049	.045	.006	.013	.033	.002	.002
	M2M	.319	<i>.304</i>		.257	.375	.298	.363	.420	.413
		.009	<i>.012</i>		.036	.002	.014	.003	.000	.001
	M21	.320	.282	.291	.281	.384	.353	.340	.422	.417
		.008	.021	.017	.021	.001	.003	.005	.000	.000
	M22	.326	.307	.306	.280	.361	.346	.267	.426	.408
		.007	.012	.012	.022	.003	.004	.029	.000	.001

(Fig.10.1.) Correlations between socio-cultural factors and degree of grammaticalization

For each of the degrees of grammaticalization (MMM, M1M etc.) the higher number refers to the value of Spearman's rho and the lower one to the statistical significance. The correlations which are statistically significant at 0.01 level have been put in bold. Those which are significant at 0.05 level have been left in regular font. Whenever a correlation is repeated and the reason for that is lack of change in the number of languages between certain groups of grammaticalization (see for instance the correlations for purpose and the number of MMM and M1M markers of purpose in Fig.9.1. and 9.3.) the results are given in italics.

10.3.2. Correlations with degree of lexicalization

In the analysis of correlations between socio-cultural factors and degrees of lexicalization conducted on the same sample of 67 languages (see Fig.10.2.) anteriority turned out to be the relation with the highest number of significant results in the M(w)M group.¹⁴ Since there has been no quantitative difference between the number of items in the subgroups of M(w)M markers and M(w)M1 group (cf. section 10.2.) the results are mirrored also in the summaries presented here. The lexicalization of purpose too, although to a smaller degree, has been found to be correlated with a number of factors the strongest of which is the number of speakers.

For causality some correlations have been found for NoS(9), the type of society and the two indicators – the majority of them at 0.01 level. The picture of conditionality is akin to it – with a number of significant results for both NoS(9) and NoS(5) as well as for both the indicators of cultural complexity.

We could have expected that high degree of lexicalization of clause linkers, in principle, is an effect of a long written tradition which contributes to the development of new means of expression, but the results presented here imply that the influence of writing is, in fact, quite modest in this domain. It is interesting, however, to see that displaced communication in general influences to a certain degree the encoding of the two relations which display the lowest overall degree of lexicalization – purpose and anteriority. Yet, at the same time the three parameters considered here as indicators of displaced communication do not seem to have any effect on lexicalization of causality and conditionality.

¹⁴ For explanations of the abbreviations see section 9.2.2.

Relation	Degree of lexical.	Socio-cultural factors								
		Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2
ANTERIORITY	M(w)MM									
	M(w)M1	.247	.265		.265	.281	.303	.270	.288	.265
		.044	.030		.030	.021	.013	.027	.018	.030
	M(w)M2	.345	.378	.360	.427	.414	.438	.429	.412	.390
		.004	.002	.003	.000	.000	.000	.000	.001	.001
	M(w)1M									
	M(w)11	.247	.265		.265	.281	.303	.270	.288	.265
		.044	.030		.030	.021	.013	.027	.018	.030
	M(w)12	.345	.378	.360	.427	.414	.438	.429	.412	.390
		.004	.002	.003	.000	.000	.000	.000	.001	.001
CAUSALITY	M(w)MM					.279		.365	.283	.249
						.022		.002	.020	.042
	M(w)M1					.311	.260	.335	.303	.283
						.010	.034	.006	.013	.020
	M(w)M2					.314	.271	.313	.291	.276
						.010	.027	.010	.017	.024
	M(w)1M							.340	.254	
								.005	.038	
	M(w)11					.268		.281	.250	
						.028		.021	.042	
	M(w)12					.274		.262		
						.025		.032		

	Degree of lexical.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2	
PURPOSE	M(w)MM										
	M(w)M1				.241						
	M(w)M2		.255		.242	.285	.301	.325	.272	.266	.270
			.037		.048	.019	.013	.007	.026	.030	.027
	M(w)1M										
	M(w)11							.241			
								.049			
M(w)12				.296	.285	.296	.350				
				.015	.020	.015	.004				
CONDITIONALITY	M(w)MM					.364	.329		.304	.304	
						.003	.007		.013	.013	
	M(w)M1						.379	.346		.314	.321
							.002	.004		.010	.008
	M(w)M2						.360	.329		.293	.306
							.003	.007		.016	.012
	M(w)1M						.371	.306		.268	.282
						.002	.012		.028	.021	
M(w)11						.396	.331		.305	.328	
						.001	.006		.012	.007	
M(w)12						.382	.318		.289	.317	
						.001	.009		.018	.009	

(Fig.10.2.) Correlations between socio-cultural factors and degree of lexicalization

The conventions of data presentation are the same as described for (Fig.10.1).

10.3.3. Correlations with degree of explicitness

The analyses of correlation between socio-cultural factors and degree of explicitness presented here have been, similarly as in section 9.3., conducted for all the languages for which the data were available and strategies of encoding of same-subject purpose have been distinguished from different-subject purpose marking.

The interpretation of the obtained results (see Fig.10.3.) is rather straightforward.¹⁵ The outcomes clearly suggest that the two relations most highly affected by socio-cultural factors when it comes to the level of explicitness are anteriority and conditionality. For causality the significance of the influence is lower but still evident from the collective measurements – indicator 1 and indicator 2. For all these three relations very significant correlations have been found with the number of speakers. This would support, to a certain degree and in the domain looked at in this study, Wray and Grace's (2007) hypothesis on the interdependency between population size and explicitness of a language (see section 10.1.). Although the authors have considered the inter-group communication an especially important factor increasing language explicitness, we are justified in extending the observation taking into account the data for the number of speakers of a particular group. The rationale behind this is the fact that the higher number of native speakers a particular language has the more prestige it gets from the neighbouring languages and the more likely it is to be learned by outsiders. In other words, more often than not the number of second-language speakers of a language is directly correlated with the number of native speakers of that language.

On the other hand, it should not escape our attention that not a single significant correlation has been found for the degree of explicitness of either same- or different-subject purpose clause marking. It has been confirmed by the results presented in section 9.3. that purpose, overall, has the mean level of explicit marking comparable to that of conditionality and causality and so the fact that only for this relation no correlations with socio-cultural factors have been found is important and calls for more detailed consideration – I investigate this issue in the next chapter.

¹⁵ In order to prevent confusion, the signs of the correlations have been changed from – to + in the table. The levels of explicitness have been encoded on a downgrading scale (1 marking the highest level) while the socio-cultural factors on an upgrading scale (the language with the highest level of written form development encoded as 6) and so the negative results of correlations between the two had to be converted to positive ones to give correct picture.

Relation	SOCIO-CULTURAL FACTORS								
	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2
ANTERIORITY (79 languages)	.354	.239	.366	.362	.294	.318	.364	.356	.312
	.001	.037	.001	.001	.009	.004	.001	.001	.005
CAUSALITY (77 languages)					.333	.277		.294	.292
					.003	.015		.010	.011
CONDITIONALITY (78 languages)	.297		.296	.306	.370	.372	.295	.361	.336
	.008		.009	.007	.001	.001	.009	.001	.003
PURPOSE same-subject (70 languages)									
PURPOSE different-subject (65 languages)									

(Fig.10.3.) Correlations between socio-cultural factors and degree of explicitness

The conventions of data presentation are the same as described for (Fig.10.1).

10.4. Summary

In this chapter, following a summary of the most important works focused on the influence of general elements of socio-cultural reality on language structure presented in section 10.1., the overview of parameters chosen for this study has been presented in section 10.2. and the results arrived at have been discussed in sections 10.3.1. – 10.3.3. In the table below (Fig.10.4.) all the findings have been gathered to illustrate the prominence of influence of the analysed factors on particular relations.

The analyses have revealed that the encoding of causality – the relation that in the previous chapter has been found to be the most strongly grammaticalized, lexicalized and most often explicitly encoded, is biased by the socio-cultural factors to a lesser degree than the other three relations. This is most visible for the correlations with degree of grammaticalization and lexicalization. By contrast, the evidence gathered for anteriority and conditionality suggest that encoding of these two relations is very prone to the influence of socio-cultural factors. For anteriority especially significant correlations have been found for the degree of lexicalization and for conditionality – for the degree of grammaticalization.

It is, however, the results for purpose that turned out to be the most intriguing ones – with a couple of significant correlations discovered only for degree of lexicalization and one (repeated) correlation for the degree of grammaticalization – the encoding of the relation seems to be very similar to the encoding of causality in terms of susceptibility to the influence of extra-linguistic factors. It has to be remembered, however, that purpose has been found to be significantly less grammaticalized and lexicalized cross-linguistically than causality and, thus, the motivations for the state that I have reconstructed in this and previous chapter have to be quite different for the two relations. I elaborate on this issue in the next chapter gathering all the findings reported in this part of the thesis.

It has been emphasized that in the search for external motivations which may influence language structure it is not possible to separate the particular factors since culture (and society) is, by definition, a network of interrelated traits. And hence, it was not expected that the results would reveal a prominence of one particular factor on the encoding of the analysed relations. Nonetheless, it should be noted that the three factors that concern displaced communication – written form development, presence of radio and TV broadcasts – are, as far as determined with the applied set of categories, clearly an important group of correlates. This concerns anteriority in particular and, to a lesser degree, also conditionality. Number of speakers (for both threshold sizes) is clearly a more important correlate for degree of lexicalization and explicitness than for degree of grammaticalization. Finally, the results of correlations with the two indicators of

cultural complexity (as well as with the parameter labelled “type of society”) strongly support the claim that overall the presence, form and explicitness of c-glossems is not immune to the influence of factors external to language and mind. And hence, adding a footnote to Newmeyer’s claim that “there is no hope of correlating a language’s gross grammatical properties with socio-cultural facts about its speakers” (2002:361) I dare to claim that after the deictic systems analysed by Perkins, the domain of clause linking has been cross-linguistically proven to be influenced by socio-cultural traits.

Relation	SOCIO-CULTURAL FACTORS																										
	Level of written form development			Presence in school teaching			Presence of radio broadcasts			Presence of TV broadcasts			NoS(9)			NoS(5)			Type of society			Indicator 1			Indicator 2		
	G	L	E	G	L	E	G	L	E	G	L	E	G	L	E	G	L	E	G	L	E	G	L	E	G	L	E
ANTERIORITY	+	+	+	+	+	+	●	●	+	●	-	+	●	+	+	-	+	+	●	+	+	●	+	+	●	+	+
CAUSALITY	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	+	+	-	-	+	+	-	+	+
CONDITIONALITY	+	-	+	+	-	-	+	-	+	-	+	+	-	+	+	-	+	+	+	-	+	+	+	+	+	+	+
PURPOSE	-	●	-	-	-	-	-	+	-	-	+	-	-	●	-	-	●	-	●	●	-	-	●	-	-	●	-

(Fig.10.4.) Influence of socio-cultural factors on encoding of circumstantial relations

The letters G, L and E in the headings refer to degree of grammaticalization, lexicalization and explicitness respectively. The + symbol refers to the presence of statistically significant correlations and – to their lack. For the degree of grammaticalization and explicitness (where 9 and 6 subgroups of c-glossemes were correlated respectively) the ● symbol refers to cases where modest number of significant correlations have been discovered.

CHAPTER 11

Conclusions

The findings presented in this part of thesis the have shown that there is a significant degree of variation between languages when it comes to the degree of grammaticalization, lexicalization and explicitness of c-glossemes they have at their disposal. The question that has been accompanying these findings from the beginning is: what are the reasons for this variation, or – in other words – what are the general motivations that lead to the development of particular semantic types of linkers. As in any type of linguistic change, it is clear that the motivations may be of various kinds (cf. the opening word for this part of the thesis), sometimes even competing with each other. In the course of the discussion here the motivations have been sought in three main types of factors: cognitive, pragmatic and socio-cultural and in addition to them certain geographic/areal tendencies have also been considered. The current chapter aims to complement the discussion by putting the observations together. However, before we do that, we need to add some more background on the cognitive and pragmatic characteristics of the analysed relations.

11.1. Cognitive salience of the four relations

Three components of mind are traditionally distinguished in cognitive psychology: cognition, affect, and conation (see, for instance, Hilgard 1980, Huitt 1996, Tallon 1997). In this trio, as Huitt and Cain explain:

Cognition refers to the process of coming to know and understand; of encoding, perceiving, storing, processing, and retrieving information. It is generally associated with the question of “what” (e.g., what happened, what is going on now, what is the meaning of that information.)

Affect refers to the emotional interpretation of perceptions, information, or knowledge. It is generally associated with one’s attachment (positive or negative) to people, objects, ideas, etc. and is associated with the question “How do I feel about this knowledge or information?”

Conation refers to the connection of knowledge and affect to behavior and is associated with the issue of “why.” It is the personal, intentional, planful, deliberate, goal-oriented, or striving component of motivation, the proactive (as opposed to reactive or habitual) aspect of behavior. (2005:1)

Three of the four relations that are the subject of this thesis: anteriority, causality and conditionality could be then classified as parts of the cognitive component. Among them anteriority is purely perceptual – based in the physical reality. Causality is more complex in that it involves not only perception (observation of regularities in the physical world the most important of which is the temporal order of SoAs) but also computation (inferring about cause-effect relation between the observed SoAs.) The third concept – conditionality involves a significant amount of processing, too. In the case of real conditions it may be based on perception (cf. *When/If it rains the grass gets wet*) but more often than not conditionality involves a more or less high degree of hypothesizing, transferring the cognitive operations to an utterly abstract domain (this concerns many cases of real conditions such as *If you go I will go with you* and all the instances of hypothetical and counterfactual conditionals). The difference in complexity between causality and conditionality boils down to the fact that causality presupposes that the propositions in both linked clauses are true while conditionality presupposes most often either disbelief or doubt about the propositions or that the speaker is uncertain about the truth of both clauses.

On the cognitive hierarchy the most basic of the three relation is then anteriority, followed by causality and conditionality. This order finds confirmation in developmental studies on utterance production – a child chains an utterance to a nonlinguistic event that was either something that she did or saw in the context before she use it in more abstract way.¹ Hence, the order of acquisition of the three concepts considered here is anteriority > causality > conditionality (cf. for instance Bloom et al. 1980, Wing and Kofsky Scholnick 1981, Diessel and Tomasello 2001)².

Purpose (in the psychological literature often labelled simply *volition*), unlike the other three concepts, comes from the domain of conation – it does not refer to perception or intelligence in the first place but to actions based on them. In broader perspective, it defines goals, choices, action plans, needs, aspirations, visions and dreams. It is also one of the main elements of self-reflection and according to some researchers (e.g. Bandura 1997; Donagan 1987; Hershberger 1988) human behaviour cannot be explained fully without it. For all these reasons, despite the fact that it is significantly different to the other circumstantial concepts considered here, it is beyond any doubt central to human life. Since it was observed that people are much more inclined to establish causal connections between events and construct intentional relations between actions within reasoning than to draw truth-functionally related conclusions from premises (Cummings 2005:91) we may assume that on the ladder of cognitive significance not only causality but also

¹ Note also that this phenomenon is parallel with the principle observed in grammaticalization studies where less abstract meanings develop into more abstract ones.

² Some other psychological studies have proved a grasp of the concept of cause in children as young as 9 months (Schlottmann and Surian 1999).

purpose is more important than conditionality.³ This claim is, however, at variance with some of the conclusions that have been drawn from observation of languages.

According to Kortmann (1997), of the four concepts considered here, two: causality and conditionality belong to the core of the semantic space of interclausal relations, and – ultimately – are the most cognitively basic of the concepts. Purpose and anteriority are listed by the author as belonging to the extended core with the first of the concepts lying closer to the core than the other one. I believe that what lead Kortmann astray is the assumption that cognitive centrality is directly mirrored by the degree of lexicalization. As described in section 9.2., Kortmann has argued that, at least for the European languages, it is a rule that the better a given circumstantial relation is lexicalized the closer to the cognitive core it is. As my analysis has shown, however, lexicalization of clause linkers is more often than not a geographic/genetic phenomenon correlating, additionally, with a variety of socio-cultural factors. For this reason I do not consider lexicalization to be a valid clue and a universal tool in the quest for cognitive salience of circumstantial relations.

The relation of clause linkers to cognition is also the subject of investigation within the already mentioned theory of semantic (and lexical) primitives proposed by Wierzbicka in early 70's and developed by herself and Goddard in the next three decades. In the centre of this theory is the claim that there exists a finite set of undecomposable meanings which can be used to explicate all the other meanings and which, therefore, can be viewed as cognitively central. As for the items being of interest to my research Wierzbicka (1996) lists only three interclausal linkers as semantic primitives: BECAUSE, IF and IF...WOULD. AFTER – as an exponent of a general temporal concept – is not treated as a *sensu stricto* clause linker (cf. also section 9.4.). The concept of purpose is not considered a semantic primitive but, commenting on BECAUSE being a semantic primitive and PURPOSE being not Wierzbicka admits that there is a specially interesting fact about the latter one:

A purposive clause can be seen as a special type of a BECAUSE-clause, and a relatively complex one. From an abstract logical point of view, one might expect that simpler, non-purposive, types of BECAUSE-clauses will be more widespread in languages of the world than purposive clauses. Empirical evidence, however, suggests that this is not the case, and that purposive clauses are much more widespread than simpler types of BECAUSE-clauses. (...) Generally speaking, one could expect that simpler configurations of semantic primes have a greater chance of being grammatically encoded in many languages than more complex ones, but of course this is not the only relevant factor. Some relatively complex configurations, such as the prototypical transitive scenario, appear to be grammatically encoded in most, if not all, languages of the world—presumably, because of the important role they play in human life (...) The purposive scenario may be another case in point. (1998:184-185)

³ The works in developmental linguistics mentioned before also place markers of purpose before markers of conditionality in the hierarchy of acquisition.

Hence, although the author does not treat purpose as a basic concept, she emphasizes its importance. Needless to say, the discussion presented here does not answer the question whether PURPOSE is a semantic primitive or not. However, Wierzbicka's observation that the concept is quite special in comparison to other similar concepts is consistent with the conclusions drawn from psychological discussion. Interestingly, the special character of purpose has also been noticed regarding some other semantic and structural characteristics of the relations between SoA involving that concept. Most recently in her typological study of purpose clauses Schmidtke-Bode (2009) has emphasized that

Purpose clauses constitute a very distinctive construction type that deviates in important ways from the archetypal characteristics shared by many adverbial clauses. The conceptual structure of purpose predetermines a number of important semantic ingredients of purpose, and is crucially defined by a mental-state relation between the purposive event and the agent of the main clause. Despite the high conceptual integration of the purposive situation into the matrix event frame, however, purpose clauses have their own information-structural value and resist the typical figure-ground conceptualization characteristic for a number of adverbial relations. (2009:202)

A similar observation concerning the non-typical position of purpose in relation to the other circumstantial relations which fit the Figure-Ground distinction has been made also by Croft (2001:326) and Kazenin (1994:93).

Although we do not have a direct insight into the cognitive architecture of our minds the observations discussed here, despite the inconsistencies between them, prove that the four concepts: anteriority, causality, conditionality and purpose are very important to us – be it on a purely cognitive or cognitive/behavioral level. We may then expect that there will be a natural tendency for them to have specialized encoding in languages across the world. This brings our discussion into the domain of pragmatics.

11.2. Pragmatic factors

The role of pragmatic factors (i.e. factors that are based in usage and speaker-hearer purposes or, in other words, in the relationship between language and context) has been considered in discussion on language change under many labels, the most frequent of which are: economy, efficiency, clarity, expressivity, and routinization. The factors are, understandably, most commonly considered in the context of grammaticalization (see, for instance, Langacker 1977, Birnbaum 1984, Harris and Campbell 1995, Hopper and Traugott 2003). Geurts summarizes the vast body of research in this domain depicting grammaticalization as

resulting from the interaction between two opposite forces: effectiveness and efficiency (also known as clarity vs. economy, force of diversification vs. force of unification, hearer's economy vs. speaker's economy, Q-principle vs. I-principle, and so on; this is a terminological free-for-all, apparently). On the one hand, speakers seek to make themselves understood and therefore strive for maximally effective messages, but on the other hand, there is a general tendency not to expend more energy than is strictly necessary and therefore to prefer economical forms to more elaborate ones. (2000:783)

Such view is, clearly, closely related to the Relevance Theory and Darwinian orientation in functionalist literature where language structures are viewed as adaptations to complex environment with communicative pressures influencing language use by hearers motivating speakers' intent to be informative and clear (cf. also the functional explanations for language change discussed in section 2.1.).

The same communicative desire – to be clear and informative – has been considered by Hopper and Traugott (2003:185) to be the motivation for the development of clause linkers (see also section 10.1.). We may expect that in the case of the four relations analysed here that are, as argued in the previous section, clearly very important concepts for humans, the communicative pressure is even higher than for other, less relevant relations. On the other hand, we have to remember that, as discussed in section 1.5.2. and remarked in section 9.3., apart from the broad category of c-glossems there are other, less specialized and less explicit strategies of expressing the relations. We may assume, therefore, that in cases where the communicative pressure is especially high, the development of clause linkers is especially favoured. The increase of communicative pressure, in turn, can be directly linked with the type of communicative situations that both speaker and hearer come across – this boils down to the extra-linguistic characteristic of the environment that the language is spoken in. I assume, in line with many of the researchers referred to in chapter 10, that the communicative pressure is lowest in small groups of speakers that share background knowledge about their day-to-day activities and can rely on context in their everyday communication. Once the number of speakers increases, once new topics arise and once new media of communication are introduced, the pressure changes. To confirm this tendency we may refer to the fact that in informal, everyday communication the least explicit strategies such as coordination and juxtaposition are still used even in languages with elaborated sets of clause linkers (including English). As remarked in chapter 1, many of the languages spoken by smaller groups have been, in fact, reported to favour juxtaposition over explicit marking.

11.3. Towards an explanation of variation

I believe that it is in the complex interaction between humans equipped with an incredible cognitive apparatus and the environment that they live in, in the coexistence of cognitive, pragmatic and socio-cultural factors, in the ultimate function of language – communication, that we may find an answer to the question why languages differ. This concerns also the domain of clause linkers.

Of the four relations analysed in this thesis **causality** has been found to be the one most highly grammaticalized, lexicalized and of the highest level of explicitness. I believe that the main reason for that is the simple fact that the concept of causality, as psychologists and philosophers have argued, is of very special importance in our cognitive architecture. It is the cognitive characteristics that translate directly onto the pressure to encode the concept in an unambiguous and explicit way using specialized marking. Being encoded in such a specialized way, regardless of the extra-linguistic factors operating over languages, it does not come as surprise that causality does reveal only occasional correlations with the general elements of socio-cultural reality discussed in chapter 11. This explains also the random geographic distribution of the languages with the highest degrees of grammaticalization and explicitness described in chapter 9.

Anteriority, too, has been viewed as cognitively important and, due to its perceptual nature, even more basic than the other relations. In contrast to causality, however, it has not only been found to be much less grammaticalized, lexicalized and explicit but also most strongly influenced by socio-cultural factors in all three aspects investigated in this thesis. Does this undermine the explanation proposed above for causality? I believe not.

The low scores for degree of grammaticalization, lexicalization and explicitness recorded for anteriority are a result of the fact that marking of the relation on the clause encoding the temporally earlier SoA is just one way of expressing it. As depicted in (Fig.9.26.) 17.7% of languages in the sample encode anteriority using not linkers that are counterparts of English ‘after’ but conventionalized narrative strategies – i.e. connective adverbs which are equivalents of ‘(and) then’ (or ‘afterwards’, ‘later’ etc.). If we add these alternative markers to the semantically monofunctional anteriority c-glossems, the level of explicitness of anteriority would be even higher than for causality, conditionality and different-subject purpose.⁴

⁴ Interestingly, studies in first language acquisition have revealed that when English-speaking children start producing utterances involving the concept of anteriority they employ ‘and then’ and ‘then’ rather than ‘after’ to encode it (see Bloom et al. 1980).

The three parameters investigated in this part of the thesis: degree of grammaticalization, lexicalization and explicitness should be thus viewed as referring to one particular way of encoding anteriority. This gives us also a clue why such a high number of correlations with socio-cultural factors has been found for the relation. The encoding by using an ‘after’ element is clearly an enrichment of a language system and more elaborated way than expression by connective adverb (discourse marker). I assume that the development of marking of anteriority on the semantically dependent clause is more often than not motivated by the new quality of expression brought about by written form and since, as mentioned in section 10.2., the socio-cultural factors are autocorrelated (e.g. languages with longer written tradition are usually spoken by higher number of speakers and in technically advanced societies with easier access to other media of communication, schooling etc.) the results show overall high correlation with extra-linguistic factors. This explains also why the distribution of the languages with the highest levels of grammaticalization, lexicalization and explicitness of anteriority is geographically limited almost exclusively to languages of Europe and Asia – it is these regions that the languages with the highest scores for socio-cultural parameters are spoken in. Finally, the high results of correlations between extra-linguistic factors and degree of explicitness revealed for this relation are partially to be accounted for by the high number of languages where juxtaposition is used as a primary method of expressing the relation. The strategy of putting clauses together without any marker of linkage between them is especially suitable for expressing a temporal relation of this kind since it is relatively straightforward for the hearer to infer about the type of relation meant by speaker on the basis of iconic order of the SoAs encoded in each of the clauses (cf. section 1.5.2. and 2.1.2. for discussion of the iconicity principle).

The encoding of **purpose** is an even more complex topic. The analyses presented in chapter 9 have revealed that although cross-linguistically the relation is grammaticalized and lexicalized to a lesser degree than causality and conditionality it is highly explicitly expressed throughout the world. Over 70% of languages have at their disposal a monofunctional purpose c-glosseme and almost 85% a c-glosseme expressing purpose and one more circumstantial relation (the most common overlap being, as we know from chapter 6., causality). The findings concerning the degree of explicitness is in line with the claims about the importance of purpose in human life referred to in section 11.1. As it is such a vital concept it is likely that in order to express it speakers will seek to make themselves clearly, unambiguously understood. Apparently this translates into explicit encoding of this relation in a number of languages but not into high levels of specialization in the sense of high degree of grammaticalization. There seems to be no easy explanation for this fact. Maybe it would be worth looking for an answer in the frequency of expressions of purpose in

relation to the other circumstantial concepts – after all it is well known (cf. Zipf's law) that frequent linguistic items are more prone to economical coding. It seems reasonable to assume that, regardless of the cognitive centrality, in the exponents of the concepts that are less frequently expressed, the degree of polyfunctionality (both semantic and syntactic) may be negotiated more strongly than in cases of exponents of the concepts that are more frequent in communicative situation.

In any case, the lower level of grammaticalization translates immediately also onto a lower level of lexicalization. Again, as has been noticed in chapter 9, lexicalization patterns of purpose reveal an interesting correlation – the highly lexicalized linkers occur almost without exception in the languages which are either the major ones (and where lexical clause linkers are the prototypical linking strategy) or are under the influence of the major ones. This explains in a straightforward way why among the 9 significant correlations with extra-linguistic factors discovered for purpose (see Fig.10.4.) 8 concern lexicalization.

Conditionality has been found to be similar to causality when it comes to the degree of lexicalization and grammaticalization but is less explicit (recall from chapter 8 that it has the lowest number of semantically monofunctional markers among the analysed relations). From the cognitive and developmental point of view, as explained in section 11.1., it is more complex than anteriority and causality. I argue that the cross-linguistically high degree of grammaticalization of this relation is explainable by its relatively high cognitive salience on the one hand and the influence of socio-cultural factors on the other (cf. 6 out of 9 parameters in Fig.10.4. are in the case of conditionality correlated with degree of grammaticalization). I believe that, as in the case of anteriority, the major factor is that of presence and length of written tradition although technical and social development too, beyond any doubt, leads to the increase in the level of abstraction employed in communication. Hypothesizing, which is the ultimate incarnation of abstraction, is, after all, the core of conditionality.

As for the degree of lexicalization, again, I believe that it is the extra-linguistic factors that explain the distribution of the most highly lexicalized conditionality linkers in the analysed sample. If we go back to the map in (Fig.9.19.) we will see that all the languages which have highly lexicalized M(w)M marking of conditionality are, once again, either the major languages of Europe and Asia or are under strong influence of those languages.⁵

Interestingly, the third parameter – degree of explicitness of encoding of conditionality – has been found to be correlated with all the socio-cultural factors

⁵ This concerns also Vitu – a language spoken northwest of the coast of West New Britain in Papua New Guinea. The speakers of Vitu use Tok Pisin as their *lingua franca* and schooling is provided exclusively in English.

considered except for the presence of the language in school teaching. Nonetheless, the number of semantically monofunctional markers of this relation is clearly smaller than for both purpose and causality (see Fig.9.26). If we consider the monofunctional conditionality markers together with cases of ‘if’/‘when’ polyfunctionality, the number rises significantly but is still smaller than for all the other relations in the same category (see cumulative values for c-glossemes covering two circumstantial meanings in Fig.9.26). One more finding supporting the claim that conditionality marking is strongly culture-centered is the fact that, as discussed in section 9.3.4., the relation has the highest record of borrowings among the four concepts investigated (and the effects of borrowing are more significant for the overall cross-linguistic explicitness of conditionality than for the other three relations).

It is often very difficult to see what the motivations for grammaticalization or lexicalization were once the processes occur. It is even more difficult to put these observations into a cross-linguistic perspective aiming to explain certain aspects of variations in language systems. Nonetheless, I believe that I have managed to prove here that, by taking into account a variety of potential explanations including cognitive, pragmatic and socio-cultural factors as well as areal/genetic patterning we are able to propose a quite coherent picture of the reasons behind the cross-linguistic variation in the encoding of circumstantial relations.

There are two more general conclusions that we may draw from the discussion presented here. The first one is that one has to remember that certain factors (including both extra and intra-linguistic) are autocorrelated and this should be taken into consideration while analyzing the influence of these factors on language structure. The second one is that one has to be careful not to draw too far-reaching conclusions from phenomena that may be geographically/genetically constrained such as degree of lexicalization of particular relations.

FINAL WORDS

“Unfortunately, or luckily, no language is tyrannically consistent. All grammars leak” wrote Edward Sapir in 1921. The metaphorical leaking has been used in this thesis to gain an insight into the origin and functioning of clause linkers that in the languages of the world mark the relations of anteriority, causality, purpose and conditionality. Each of the languages analysed has been read through as a palimpsest which reveals, unintentionally, the work of earlier generations of scribes and allows us to answer the questions about historical origin and functioning of the linkers. Moreover, when interpreted in the broader context which takes into account cognitive, pragmatic and language-external functions, it allows us also to get an insight into the motivations for the development of the fascinating group of linguistic items and the reasons behind cross-linguistic variation in the encoding of the aforementioned relations.

In the first part of the thesis I scrutinized the scope of the research as well as its theoretical foundations and the methodology applied. In the first four sections of chapter 1 I discussed the concept of state of affairs, clause and symmetrical and asymmetrical relations they may form as well as the idea of circumstantial relations. I then moved to the discussion on a variety of strategies of encoding of the relations (such as adverbial subordinators, converbs, clause chaining, special verb forms, coordination and juxtaposition) and proposed their function-based classification. I argued for the usefulness of introduction of a functional definition of clause linkers since, considering the structural differences in the analysed languages and problems with universal application of existing terms such as adverbial subordinator, converb (or even subordination and non-finiteness), only such a definition allow us to cover the full range of items designated to act as clause-linking devices. Hence, the proposed category of *circumstantial-glossemes*, as well as the four circumstantial relations looked at in this study, has been defined in structure-independent terms.

In chapter 2 the functional approach to the investigated issues was elaborated on and the inspirations drawn from functional-typological framework were emphasized. Language sample and data collection strategies were also discussed within that chapter with emphasis on the practical problems that everyone working on a research project involving a world choice of languages has to face. By applying a statistical method of language sampling I have ensured the representativeness of my sample and so I am confident that the results obtained are reliable. Data collection and data collection

strategies (with an emphasis on the role of language consultants in verification and supplementation of the data) were also presented within chapter 2.

Part II has been devoted mainly to the problem of grammaticalization of clause linkers and analysis of semantic affinities of the four circumstantial relations revealed by semantic polyfunctionalities of c-glossems. Some attention has been devoted to the analysis of forms of the linkers too. In the introductory chapter to part II, theory of grammaticalization and its main components were discussed. It was emphasized that the present study relies on exclusively synchronic material to reveal the pathways of development of clause linkers – an approach which is by no means new to grammaticalization studies:

Much what we have learned about grammaticalization in the last 30 years, let alone much of the pioneering work of the 19th century linguistics, rests on our ability, or willingness, to interpret synchronic structural distortions as relic fossils of prior diachronic change. (Givón 2002:39)

No other study, according to my knowledge, has used the synchronic evidence in a systematic study as extensively as I have done in this work. Following a set of *grammaticalization heuristics* built on our knowledge on grammaticalization processes and analyses of frequencies of occurrences of particular patterns of polysemy a number of analyses were performed and reported in chapters 4-7. The results were used in the reconstruction of the most common sources of particular semantic types of clause linkers. The internal structure of polymorphemic c-glossems was also scrutinized and the outcomes of the analysis were used to support and complement the data obtained from the analysis of patterns of polysemy. On the whole, the synchronic method has proved to be very successful – it has not only confirmed what has been known from other studies but provided us with new material for hypothesizing about grammaticalization pathways which have been previously not reported in the literature. The results have revealed an interesting picture of the main sources of the four groups of markers, two of which – anteriority and causality c-glossems – have been so far very poorly discussed in grammaticalization literature. One of the many interesting findings of this part of the thesis, which was possible to obtain only through a quantitative analysis of data, has been the rather striking difference between the four relations in the proportion of the categories that in the linguistic literature are viewed as the most common sources of clause linkers: adpositions, case markers, nouns and verbs. For instance, case markers were found to overlap twice more frequently with purpose linkers than with causality linkers, and adpositions to be three times more frequent sources of causality than of conditionality. Moreover, the existence of certain groups of relation-specific sources was also confirmed – this includes, for instance,

complementizers for purpose linkers and adverbs for conditionality and anteriority. On the pages of chapters 4-7 I argued also for a number of development pathways involving either extension of certain clause linkers over other clause linkers or separate pathways of grammaticalization where a given c-glosseme overlaps with another one as an epiphenomena of the fact that they both have been derived from a common source. The important conclusion I have drawn from these observations is that in the discussions on grammaticalization of clause linkers, more attention should be devoted to the fact that grammaticalization processes, based on principles of iconicity, operate also within the group of c-glossemes.

The second major component of the analyses in part II has been the reconstruction of the semantic affinities in the domain of circumstantial relations by looking at the types and frequencies of most common semantic overlaps of the c-glossemes. The analyses verified positively those that have been observed by Kortmann (1997) for European languages. The advantage of this study in this domain is the reconstruction of detailed semantic maps allowing us to gain an insight into the core as well as the margins of the conceptual space of the analysed relations and to observe complex-multiple overlaps. Again, interesting discrepancies have been found in the density of the maps reconstructed for anteriority and causality on the one hand and for conditionality and purpose on the other.

The four groups of linkers have been analysed also from the point of view of their degree of syntactic and semantic polyfunctionalities and the ratio of monomorphemic-to-polymorphemic items. For all the groups a tendency for more morphologically complex markers to be semantically and syntactically monofunctional has been recorded. Moreover, all the four groups have revealed a similar picture of semantic polyfunctionality and similar proportions of monomorphemic and polymorphemic items. As for syntactic polyfunctionality, conditionality markers have been found to possess lower number of polysemes than the linkers of the other three relations.

Finally, regarding the morphosyntactic forms of the markers, causality c-glossemes have been found to display the highest proportion of words among the analysed relations. The proportion of affixes, in turn, has been discovered to be highest for anteriority linkers. Certain geographic preferences concerning the forms of the markers have been found but their distribution, as I have shown, is quite inconsistent casting doubts on significance of the parameter of lexicalization for discussion on cognitive centrality of certain relations as Kortmann (1997) proposed for European languages.

In part III of the dissertation attention has focused on the cross-linguistic variation in degrees of grammaticalization, lexicalization and explicitness of particular

linkers and the motivations for both – the origin of the markers and their variation have been considered. In chapter 9 the results of the quantitative analyses were discussed. Of the four groups of circumstantial linkers causality c-glossemes have been found to be cross-linguistically most strongly grammaticalized (displaying the highest proportion of monomorphemic, monofunctional semantically and syntactically markers), most strongly lexicalized and most explicit. Markers of conditionality have been found to reveal similarly high levels of grammaticalization, lexicalization and explicitness which contrast directly with the low levels revealed by anteriority and purpose markers. Moreover, the markers of causality and conditionality turned out also to be more frequently borrowed than the markers of purpose or anteriority. Interestingly, the analysis has revealed quite strong geographic patterns in the variation of the three parameters – the regions with lowest scores for almost all of the investigated domains are Oceania, New Guinea, Australia and Indonesia, while Europe and Mainland Asia reveal the highest scores.

In chapter 10, the data reported in chapter 9 were correlated with a number of socio-cultural traits. The goal of the correlation analyses was to verify the potential influence that language external factors may have on the availability of clause linkers and the degree of their grammaticalization, lexicalization and explicitness in particular languages. The analyses were set against a growing body of hypotheses and works which suggest that extra-linguistic factors may affect language structure to a significant degree. For the group of clause linkers hypotheses of this kind has been most commonly proposed in relation to the level of written form development but in the statistical tests performed on the data I have considered also other factors such as population size, presence and characteristics of the language in school teaching, radio and TV broadcasting, type of society etc. Some intriguing results emerged from the analysis – encoding of causality and, quite surprisingly, purpose, seem to be much less susceptible to the influence of socio-cultural factors than encoding of the other two relations. On the whole, however, none of the relations seems to be completely immune to the influence of extra-linguistic factors. To my knowledge, apart from the study on the influence of cultural traits on the richness of deictic systems by Perkins (1992), my work is the only one that has looked at the issue of influence of a variety of language-external factors on language systems from a cross-linguistic point of view.

Finally, gathering all the information presented earlier, in chapter 11 I have attempted to explain the cross-linguistic variation combining, in the spirit of functional approach to language change, the cognitive, pragmatic and socio-cultural clues.

Although the work has yielded many interesting cross-linguistic generalizations and shed empirical light on issues which have been previously under-researched or only

hypothesized about, a number of issues have not been investigated here and some others have been treated only very generally. There is a number of directions and a number of ways in which this study can be developed.

The first prospect that comes to mind is the verification of many of the pathways of grammaticalization hypothesized here which have not been reported earlier in the literature. Such an endeavour would, of course, require diachronic studies on a number of languages. Given the fact that historical materials are not readily available (and for many languages will never be) in many cases such investigation would have to involve extensive data collection process including historical reconstructions.

Due to the size of the sample, nothing has been said here about the areal patterns of grammaticalization or external motivation for grammaticalization. This, without any doubt very interesting subject, would be worth pursuing, too. The same concerns in-depth analyses of patterns of borrowing of *c-glossemes* and motivation behind the process of borrowing – a topic that has been only very briefly treated in this thesis.

Certainly, it would also be interesting to enrich the analyses of semantic polyfunctionality of clause linkers by considering the methods of disambiguation that languages employ to distinguish between the various meanings. This would, naturally, require analyses of entire constructions with all their formal features (such as configurations of TAM markers, for instance) on the one hand and in-depth analyses of context on the other. Taking into account that such study would have to face the reality of working with a number of very diverse languages both of these tasks are a challenge in their own right.

The interesting topic or richness and diversity of systems of clause linkers would be a good candidate for a research topic, providing one has access to detailed grammatical descriptions of comparable depth and quality, for a variety of genetically and geographically diverse languages. One of the directions in which the research could go is the influence of literature and writing on the systems. Kortmann (1997:256) has argued, the languages with a long literary tradition are also those with the most elaborate sets of clause linkers. I would propose to extend the investigation beyond the languages of Europe and control not only for the length of literary but also written tradition in general and, ideally, to consider vitality of oral literature too.

Finally, the natural direction of extending the present work would be to look at a much larger set of markers of circumstantial relations in a much larger set of languages. Such additional analyses would surely contribute to our understanding of human language and human cognition, although, on the whole, they would certainly retain the picture of language as emerging from the analyses presented here – a constantly changing structure where diachrony penetrates synchrony, where cognition

and pragmatics dictate certain solutions and where the fascinating tool of communication gets reshaped by the socio-cultural environment in which it is used.

REFERENCES

PRINTED MATERIALS

ABBOTT, M. (1985): Subordinate clauses in Macushi. In: *Porto Velho Workpapers*, D. L. Fortune (ed.). Brazil: SIL. 254-67.

— (1991): Macushi. In: *Handbook of Amazonian Languages*. Vol. 3., D.C. Derbyshire, G.K. Pullum (eds.). Berlin: de Gruyter, 23-160.

AGNITHI, R.K. (2007): *Hindi: An Essential Grammar*. London: Routledge.

AIKHENVALD, A.Y. (2006): Serial verb constructions in a typological perspective. In: A.Y. Aikhenvald, R.M.W. Dixon (eds.), 1-87.

— (2008): Versatile Cases. *Journal of Linguistics* 44, 565-603.

— (2009): Semantics and Grammar in Clause Linking. In: R.M.W. Dixon, A.Y. Aikhenvald (eds.), 380-402.

—, DIXON, R.M.W. (2001a): Introduction. In: A.Y. Aikhenvald, R.M.W. Dixon (eds.) (2001b), 1-26.

—, DIXON, R.M.W. (eds.) (2001b): *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: OUP.

—, DIXON, R.M.W. (eds.) (2006): *Serial Verb Constructions. A Cross-linguistic Typology*. Cambridge: CUP.

AKATSUKA, N. (1986): Conditionals are discourse bound: In: E.C. Traugott et al. (eds.), 333-52.

ALPATOV, V.M., PODLESSKAYA, V.I. (1995): Converbs in Japanese. In: M. Haspelmath, E. König (eds.), 465-485.

AMHA, A. (2001): *The Maale language*. Leiden: CNWS.

ANDERSEN, G. (2001): *Pragmatic markers and sociolinguistic variation*. Amsterdam: Benjamins.

ANDERSON, S.R. (1977): On mechanisms by which languages become ergative. In: Ch.N. Li (ed.), 317-363.

— (1985): Typological distinctions in word formation. In: T. Shopen (ed.), vol. 3., 3-56.

ANISFELD, M. (1984): *Language Development from Birth to Three*. New Jersey: Lawrence Erlbaum Associates Inc.

AOKI, H. (1970): *Nez Perce grammar*. Berkeley: University of California Press.

— (1994): *Nez Perce dictionary*. Berkeley: University of California Press.

- APRESJAN, J. (1974): Regular polysemy. *Linguistics* 142, 5-33.
- ÅRSJÖ, B. (1994): Topic in Ama discourse. *Language and Linguistics in Melanesia* 25, 1-25.
- ASHTON, E.O. (1947): *Swahili grammar (including intonation)*. London: Longmans.
- ATHANASIADOU, A., DIRVEN R. (eds.) (1997): *On Conditionals Again*. Amsterdam: Benjamins.
- AUWERA VAN DER, J., Ó BAOILL, D.P. (eds.) (1998): *Adverbial Constructions in the Languages of Europe*. Berlin: de Gruyter.
- BACKHOUSE, A.E. (1994): *The Japanese Language: An Introduction*. Oxford: OUP.
- BAKKER, D. (forthcoming): Language sampling. In: *Handbook of Linguistic Typology*. J. Jung Song (ed.). Oxford: OUP.
- BALMER, W.T., GRANT, F.C.F. (1942): *A grammar of the Fante-Akan*. London: Atlantis Press.
- BANDURA, A. (1997): *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- BAŃKO, M. (ed.) (2000): *Inny słownik języka polskiego*. Warsaw: PWN.
- BEATON, A. C. (1968): *A grammar of the Fur language*. Khartoum: University of Khartoum.
- BELL, A. (1978): Language samples. In: *Universals of Human Language*, J. Greenberg, C.A. Ferguson, E.A. Moravcsik (eds.). Cambridge: MIT Press, 123-156.
- BENNETT, J. (1988): *Events and Their Names*. Oxford: Clarendon Press.
- (1996): *What Events Are*. In: R. Casati, A.C. Varzi (eds.), 137-151.
- BERG VAN DEN, R., BACHET, P. (2006): *Vitu grammar sketch*. Ukarumpa: SIL.
- BERGER, H. (1998): *Die Burushaski-Sprache von Hunza und Nager. Teil I: Grammatik. Teil II: Texte mit Übersetzungen. Teil III: Wörterbuch*. Wiesbaden: Harrassowitz.
- BERNSTEIN, B. (1974): *Class, Codes, and Control. Theoretical Studies towards a Sociology of Language. Vol. 1*. London: Routledge.
- BICKERTON, D. (1981): *Roots of Language*. Ann Arbor: Karoma.
- (1990): *Language and Species*. Chicago: University of Chicago Press.
- BIRNBAUM, H. (1984): Notes on syntactic change: cooccurrence vs. substitution, stability vs. permeability. In: *Historical syntax*, J. Fisiak (ed.). Berlin: de Gruyter, 25-46.

- BISANG, W. (1996): Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in East and Mainland South East Asian languages. *Studies in Language* 20(3), 519-97.
- (1998): Adverbiality: The view from the Far East. In: J. van der Auwera, D.P. Ó Baoill (eds.), 641-812.
- BLACKINGS, M.J., FABB, N. (2003): *A grammar of Ma'di*. Berlin: de Gruyter.
- BLAKE, B. (1988): Review of Meritt Ruhlen, *A Guide to the World's Languages*. Vol. 1: Classification. *Journal of Linguistics* 24(1), 261-262.
- BLANK, A. (2001): Pathways of lexicalization. In: M. Haspelmath et al. (eds.), vol. 2., 1596-1608.
- , KOCH, P. (eds.) (1999): *Historical Semantics and Cognition*. Berlin: de Gruyter.
- BLOOM, L. (2006): Language Acquisition in its Developmental Context. In: *Handbook of Child Psychology*. Vol 2. *Cognition, Perception, and Language*, W. Damon, D. Kuhn, R.S. Siegler (eds.). New Jersey: John Wiley and Sons Inc, 309-370.
- , CAPATIDES, J.B. (1987): Sources of Meaning in the Acquisition of Complex Syntax: The Sample Case of Causality. *Journal of Experimental Child Psychology* 43, 112-128.
- , LAHEY, M., HOOD, L., LIFTER K., FIESS, K. (1980): Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language* 7, 235-261.
- BLOOMFIELD, L. (1933): *Language*. New York: Henry Holt and Company.
- BLUTNER, R. (1998): Lexical pragmatics. *Journal of Semantics* 15, 115-162.
- BLÜHDORN, H. (2008): Subordination and coordination in syntax, semantics and discourse. Evidence from the study of connectives. In: *Subordination versus coordination in sentence and text. A cross-linguistic perspective*, C. Fabricius-Hansen, R. Wiebke (eds.). Amsterdam: Benjamins, 59–85.
- BOLINGER, J. (1984): Intonational Signals of Subordination. In: C. Brugman, M. Macaulay (eds.), 401-413.
- BOUQUIAUX, L. (1978): *Dictionnaire sango-français = Bākàrí sango-fàránzì*. Paris: Société d'études linguistiques et anthropologiques de France.
- BOWDEN, J. (1997): *Taba (Makian Dalam): description of an Austronesian language from Eastern Indonesia*. Melbourne: University of Melbourne.
- (2001): *Taba: description of a South Halmahera Austronesian language*. Canberra: Australian National University.

- BRADLEY, J., KIRTON, J. AND THE YANYUWA COMMUNITY (1992): *Yanyuwa Wuka: language from Yanyuwa country. A Yanyuwa dictionary and cultural resource*. Manuscript, Australian Institute for Aboriginal and Torres Strait Islander Studies
- BRINTON, L.J., TRAUOGOTT E.C. (2006): *Lexicalization and Linguistic Change*. Cambridge: CUP.
- BRODY, J. (1987): Particles borrowed from Spanish as discourse markers in Mayan languages. *Anthropological Linguistics* 29(4), 507-521
- BROWN, K. (ed.) (2005): *Encyclopedia of Language and Linguistics*. Vol. 1-14. New York: Elsevier.
- BRUGMAN, C., MACAULAY, M. (eds.) (1984): *Proceedings of the 10th Annual Meeting of the Berkeley Linguistic Society*. Berkeley: Berkeley Linguistic Society.
- BURENHULT, N. (2005): *A grammar of Jahai*. Canberra: Australian National University.
- BUSSMANN, H. (1996): *Routledge Dictionary of Language and Linguistics*. London: Routledge.
- BYBEE, J. (1985): *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.
- , DAHL, Ö. (1989): The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1), 51-103.
- , HOPPER, P.J. (2001): Introduction to frequency and the emergence of linguistic structure. In: *Frequency and the Emergence of Linguistic Structure*, J. Bybee, P.J. Hopper (eds.). Amsterdam: Benjamins, 1-26.
- , NOONAN, M. (2001): *Complex Sentences in Grammar and Discourse: Essays in honor of Sandra A. Thompson*. Amsterdam: Benjamins.
- , PERKINS R., PAGLIUCA W. (1994): *The evolution of grammar: tense, aspect and modality in the language of the world*. Chicago: University of Chicago Press.
- CAMPBELL, L. (2001): What's wrong with grammaticalization? *Language Sciences* 23(2-3), 113-161.
- , BUBENIK, V., SAXON, L. (1988): Word order universals: refinements and clarifications. *Canadian Journal of Linguistics* 33, 209-23.
- , JANDA, R. (2001): Introduction: conceptions of grammaticalization and their problems. *Language Sciences* 23(2-3), 93-112.
- , MITHUN, M. (1980): Syntactic reconstruction: priorities and pitfalls. *Folia Linguistica Historica* 1(1), 19-40.

—, RINGEN, J. (1981): Teleology and the explanation of sound change. In: *Phonologica 1980*, W.U. Dressler, O.E. Pfeiffer, J.R. Rennison (eds.). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 57-67.

CARLSON, R. (1994): *A grammar of Supyire*. Berlin: de Gruyter.

CASATI, R., VARZI, A. C. (eds.) (1996): *Events*. Aldershot: Dartmouth Publishing Group.

—, (1997): *Fifty Years of Events. An Annotated Bibliography 1947 to 1997*. Bowling Green: Philosophy Documentation Center.

CHAFE, W. (1982): Integration and Involvement in Speaking, Writing, and Oral Literature. In: *Spoken and Written Language. Exploring Orality and Literacy*, D. Tannen (ed.). Norwood: Ablex.

— (1984): How People Use Adverbial Clauses. In: C. Brugman, M. Macaulay (eds.), 437-449.

— (1987): Cognitive Constraints on Information Flow. In: *Coherence and Grounding in Discourse*, R. Tomlin (ed.). Amsterdam: Benjamins, 21-51.

— (1994): *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*: Chicago: University of Chicago Press.

CHAPPELL, H. (1994): Mandarin Semantic Primitives. In: A. Wierzbicka, C. Goddard (eds.), 109-148.

CHULUU, Ü. (1994): *Introduction, Grammar, and Sample Sentences for Dagur*. Philadelphia: University of Pennsylvania.

CLAUDI, U., HEINE, B. (1986): On the metaphorical base of grammar. *Studies in Language* 10, 297-335.

COMRIE, B. (1976): *Aspect*. Cambridge CUP.

— (1985): *Tense*. Cambridge CUP.

— (ed.) (1987): *The World's Major Languages*. London: Croom Helm.

— (ed.) (1990): *The major languages of East and South-East Asia*. London: Routledge.

— (ed.) (1990): *The major languages of Eastern Europe*. London: Routledge.

— (ed.) (1990): *The major languages of South Asia, the Middle East and Africa*. London: Routledge.

— (ed.) (1990): *The major languages of Western Europe*. London: Routledge.

- CONDORAVDI, C., FILIP, H. (2007): *A Bibliography on Event Semantics and Related Matters. Events: Modification, Aspect and Lexical Meaning*. Stanford: LSA Institute. (available at <http://grove.ufl.edu/~hfilip/LSA.315.bib.pdf>. Last accessed: 09/07/2010)
- CORVER, N., VAN RIEMSDIJK, H.C. (2001): *Semi-lexical categories: the function of content words and the content of function words*. Berlin: Mouton de Gruyter.
- COUPER-KUHLEN, E., KORTMANN, B. (eds.) (2000): *Cause – condition – concession – contrast: cognitive and discourse perspectives*. Berlin: Mouton de Gruyter.
- CRAIG, C.G. (1991): Ways to go in Rama: a case study in polygrammaticalization. In: E.C. Traugott, B. Heine (eds.), vol. 2., 455-492.
- CRASS, J. (2002): *Die Grammatikalisierung des Verbes "sagen" im Beria*. Typescript, University of Mainz.
- CRISTOFARO, S. (2003): *Subordination*. Oxford: OUP.
- CROFT, W. (2003)[1990]: *Typology and Universals*. Cambridge: CUP.
- (1991): *Syntactic categories and grammatical relations: the cognitive organisation of information*. Chicago: University of Chicago Press.
- (1995): Autonomy and functionalist perspective. *Language* 71(3), 490-532.
- (2000): *Explaining Language Change: An Evolutionary Approach*. Harlow- Essex: Longman.
- (2001): *Radical Construction Grammar*. Oxford: OUP.
- CUMMINGS, L. (2005): *Pragmatics: A Multidisciplinary Approach*. Edinburgh: Edinburgh University Press.
- CURNOW, T. J. (2001): What language features can be 'borrowed'?. In: A.Y. Aikhenvald, R.W.M. Dixon (eds.) (2001b), 412-436.
- CYFFER, N. (1998): *A sketch of Kanuri*. Köln: Köppe.
- CYSOUW, M. (2003) *The Paradigmatic Structure of Person Marking*. Oxford: OUP.
- DAHL**, Ö. (1985): *Tense and Aspect Systems*. Oxford: Blackwell.
- (ed.) (2000): *Tense and Aspect in the Languages of Europe*. Berlin: de Gruyter.
- (2004): *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: Benjamins.
- DALBY, A. (2006) [1998]: *Dictionary of Languages: The Definitive Reference to More than 400 Languages*. London: A&C Black Publishers Ltd.

- DANCYNGIER, B. (1993): Interpreting conditionals: time, knowledge and causation. *Journal of Pragmatics*, 19, 403-434.
- (1998): *Conditionals and prediction: time, knowledge, and causation in conditional constructions*. Cambridge: CUP.
- , SWEETSER, E. (2000): Constructions with *if*, *since* and *because*: Causality, Epistemic Stance, and Clause Order. In: E. Couper-Kuhlen, B. Kortmann (eds.), 111-142.
- DANIELSEN, S. (2007): *Baure: an Arawak language of Bolivia*. Leiden: CNWS.
- (forthcoming): Clause embedding strategies in Baure (Arawak). In: *Subordination in South American Indian languages*, K. Haude, R. van Gijn, P. Muysken (eds.). Amsterdam: Benjamins.
- DAVIDSON, D. (1967): The logical form of action sentences. In: *The Logic of Decision and Action*, N. Rescher (ed.). Pittsburgh: University of Pittsburgh Press, 81-95.
- (1980): *Essays on Actions and Events*. New York: OUP.
- DAVIES, H. (2005): Coordination and constituency in St'át'imcets. In: *Verb First: On the syntax of verb-initial languages*, A. Carnie, H. Harley, S.A. Dooley (eds.). Amsterdam: Benjamins, 31-64.
- DECLERCK R., REED, S. (2001): *Conditionals: a comprehensive empirical analysis*. Berlin: de Gruyter
- DEGRAFF, M. (2001): On the origin of Creoles: A Cartesian critique of “neo” Darwinian Linguistics. *Linguistic Typology* 5, 213-311.
- DEVRIENDT, B., GOOSSENS, L., AUWERA VAN DER, J. (eds.) (1996): *Complex Structures. A Functionalist Perspective*. Berlin: de Gruyter.
- DIESEL, H., TOMASELLO, M. (2001): The acquisition of finite complement clauses in English: A corpus-based analysis, *Cognitive Linguistics*, 12, 1-45.
- DIK, S.C. (1997a) [1989]: *The Theory of Functional Grammar. Part 1: The Structure of the Clause*, K. Hengeveld (ed.). Berlin: de Gruyter.
- (1997b) [1989]: *The Theory of Functional Grammar. Part 2: Complex and Derived Constructions*, K. Hengeveld (ed.). Berlin: de Gruyter.
- DIXON, R.M.W. (2009): The Semantics of Clause Linking in Typological Perspective. In: R.M.W. Dixon, A.Y. Aikhenvald (eds.), 1-55.
- , AIKHENVALD A.Y. (eds.) (2001): *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: OUP.

- , AIKHENVALD A.Y. (eds.) (2002): *Word: a typological framework*. In: *Word: A cross-linguistic typology*, R.M.W. Dixon, A.Y. Aikhenvald (eds.), Cambridge: CUP, 1-42.
- , AIKHENVALD, A.Y.(eds.) (2009): *The Semantics of Clause Linking: A Cross-Linguistic Typology*. Oxford: OUP.
- DONAGAN, A. (1987): *Choice, the essential element in human action*. London: Routledge.
- DONOHUE, M., ROQUE, L.S. (2002): *I'saka*. Canberra: Pacific Linguistics.
- DRYER, M.S. (1989): Large Linguistic Areas and Language Sampling. *Studies in Language* 13, 257-292.
- (1992): The Greenbergian Word Order Correlations. *Language* 68, 81-138.
- (1997): Are Grammatical Relations Universal? In: *Essays on Language Function and Language Type: Dedicated to T. Givón*, J. Bybee, J. Haiman, S.A. Thompson (eds.). Amsterdam: Benjamins, 115-143.
- DU BOIS, J.A. (1985): Competing motivations. In: J. Haiman (ed.), 343-365.
- (1987): The discourse basis of ergativity, *Language* 63, 805-855.
- DUNN, M.J. (1999): *A grammar of Chukchi*. PhD dissertation, Australian National University.
- E**BERT, K. (1991): Vom Verbum dicendi zur Konjunktion: ein Kapitel universaler Grammatikentwicklung. In: *Von Europa bis Ozeanien—von der Antonymie zum Rektivsatz: Gedenkschrift für Meinrad Scheller*, W. Bisang, R. Rinderknecht (eds.) Zürich: Arbeiten des Seminars für Allgemeine Sprachwissenschaft der Universität Zürich, 77-95.
- ECKERT, P., HUDSON, J. (1994): *Wangka Wiru: a handbook for the Pitjantjatjara language learner*. Underdale: Aboriginal Studies and Teacher Education Centre.
- EGLI, H. (1990): *Paiwangrammatik*. Wiesbaden: Harrassowitz.
- EIJK VAN, J. (1997): *The Lilloet language: phonology, morphology, syntax*. Vancouver: University of British Columbia.
- ENFIELD, N.J. (2002): Ethnosyntax: Introduction. In: *Ethnosyntax: Explorations in Grammar and Culture*, N.J. Enfield (ed.). Oxford: OUP, 3-30.
- ENGELENHOVEN VAN, A. (1995): *A description of the Leti Language (as spoken in Tutukei)*. Ph.D. dissertation, University of Leiden.
- (2004): *Leti: a language of Southwest Maluku*. Leiden: KITLV Press.

- ERELT, M. (ed.) (2003): *Estonian language*. Tallinn: Estonian Academy Publishers.
- FACUNDES DA SILVA, S. (2000): *The language of the Apurinã people of Brazil. (Maipure/Arawak)*. Buffalo: SUNY.
- FERGUSON, CH. A., SNITZER REILLY, J., TER MEULEN, A., TRAUOGOTT, E.C. (1986): Overview. In: E.C. Traugott et al. (eds.), 3-20.
- FISCHER, O., NORDE, M., PERRIDON, H. (eds.) (2004): *Up and down the cline*. Amsterdam: Benjamins.
- FOLEY, W.A. (1991): *The Papuan Languages of New Guinea*. Cambridge: CUP.
- (1991): *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- FORTESCUE, M. (1984): *West Greenlandic*. London: Cromm Helm.
- FRAJZYNGIER, Z. (1996): *Grammaticalization of the complex sentence: A case study in Chadic*. Amsterdam: Benjamins.
- FRAKE, CH. (1969): The Ethnographic Study of Cognitive Systems. In: *Cognitive Anthropology*, S. Tyler (ed.). New York: Holt Rinehart and Winston, 28-41.
- FRANK, P.S. (1990): *Ika syntax*. Dallas: SIL and the University of Texas at Arlington.
- (2003): *Una Gramática del Arhuaco*. Manuscript.
- Gabsi, Z. (2003): *An outline of the Shilha (Berber) vernacular of Douiret (Southern Tunisia)*. PhD dissertation, University of Western Sydney.
- GENETTI, C. (1986): The development of subordinators from postpositions in Bodic languages. In: *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, V. Nikiforidou, M. Van Clay, D. Feder (eds.). Berkeley: Berkeley Linguistics Society, 387-400.
- (1991): From postposition to subordinator in Newari. In: E.C. Traugott, B. Heine (eds.), vol. 2., 227-255.
- GEORG, S. (2007): *A descriptive grammar of Ket (Yenisei-Ostyak): Part 1, Introduction, Phonology, Morphology*. Folkestone: Global Oriental
- GERKEN, L. A. , LANDAU, B., REMEZ, R.E. (1990): Function morphemes in young children's speech perception and production. *Developmental Psychology* 26, 204-216.
- GEURTS, B. (2000): Explaining grammaticalization (the standard way). *Linguistics* 38(4), 781-788.
- GHOSH, A. (2008): Santali. In: *The Munda Languages*, G.D.S. Anderson (ed.). London: Routledge, 11-98.

- GIL, D. (1991): Aristotle goes to Arizona and finds a language without AND. In: D. Zaefferer (ed.) (1991b), 96–130.
- (2001): Escaping Eurocentrism: Fieldwork as a Process of Unlearning. In: *Linguistic Fieldwork*, P. Newman, M. Ratliff (eds.). Cambridge: CUP, 102-132.
- GIVÓN, T. (1979): *On Understanding Grammar*. New York: Academic Press.
- (1990): *Syntax: A Functional-Typological Introduction*. vol. 2. Amsterdam: Benjamins.
- (1991): The evolution of dependent clause morpho-syntax in Biblical Hebrew. In: E.C. Traugott, B. Heine (eds.), vol. 2., 257-310.
- (1995): *Functionalism and grammar*. Amsterdam: Benjamins.
- (2002): The visual information-processing. In: *The Evolution of Language out of Pre-language*, T. Givón, B.F. Malle (eds.). Amsterdam: Benjamins, 3-50.
- GODDARD, C. (1994): Semantic theory and semantic universals. In: C. Goddard, A. Wierzbicka (eds.), 7-30.
- (2001): *Pitjantjatjara/Yankunytjatjara to English dictionary*. Alice Springs: IAD Press.
- , WIERZBICKA, A. (eds.) (1994): *Semantic and Lexical Universals – Theory and Empirical Findings*. Amsterdam: Benjamins.
- GOLINKOFF, R., HIRSH-PASEK, K. (1995): Reinterpreting children's sentence comprehension: Towards a new framework. In: *Handbook of language acquisition*, P. Fletcher, B. MacWhinney (eds.). London: Blackwell, 430-461.
- GRAVELLE, G.G. (2004): *Meyah: an east Bird's Head language of Papua, Indonesia*. Ph.D. dissertation. Vrije Universiteit Amsterdam.
- GREENBERG, J. H. (1971): *Language, Culture, and Communication*. Stanford: University Press.
- GRICE, H.P. (1975): Logic and conversation. In: *Syntax and Semantics*, vol. 3., P. Cole, J. Morgan (eds.), New York: Academic Press, 41-58.
- (1989): *Studies in the Way of Words*. Cambridge: Harvard University Press.
- GRINEVALD, C.G. (1990): *A grammar of Rama*. Université de Lyon. Report to National Science Foundation.
- GRONDONA, V.A. (1998): *A grammar of Mocoví*. PhD dissertation, University of Pittsburgh.
- GRUZDEVA, E. (1998): *Nivkh*. München: Lincom.

- GRZEGORCZYKOWA, R. (1996): *Wykłady z polskiej składni*. Warsaw: PWN.
- GÜLDEMANN, T. (2001): *Quotative constructions in African languages: a synchronic and diachronic survey*. Habilitation dissertation, University of Leipzig.
- HAIMAN, J. (1978): Conditionals are topics. *Language* 54, 564-589.
- (1980): The Iconicity of Grammar: Isomorphism and Motivation. *Language* 56, 515-540.
- (ed.) (1985): *Iconicity in Syntax*. Amsterdam: Benjamins.
- , KUTEVA, T. (2001): The symmetry of counterfactuals. In: J.L. Bybee, M. Noonan (eds.), 101-124.
- , THOMPSON, S.A. (1988): *Clause combining in grammar and discourse*. Amsterdam: Benjamins.
- HALE, K. (1971): Gaps in grammar and culture. In: *Linguistics and anthropology in honor of C. F. Voegelin*, D. Kinkade, K. Hale, O. Werner (eds.). Lisse: Petter de Ridder, 295-315.
- (1983): Warlpiri and the Grammar of Non-Configurational Languages, *Natural Language and Linguistic Theory*, 1(1), 5-47.
- HARMS, R. T. (1962): *Estonian Grammar*. The Hague: Mouton.
- HARPER, K. (1974): *Some aspects of the grammar of the Eskimo dialects of Cumberland Peninsula and North Baffin Island*. Ottawa: National Museum of Canada.
- HARRIS, M. (1986): The historical development of *si*-clauses in Romance. In: Traugott, E.C., et al. (eds.), 265-284.
- (1988): Concessive clauses in English and Romance. In: J. Haiman, S.A. Thompson (eds.), 71-100.
- (1989): Markedness and clause structure. In: *Markedness in synchrony and diachrony*, O. Mišeska Tomić (ed.). Berlin: de Gruyter, 333-358.
- (1990): The evolution of certain patterns in subordination in Romance and English. In: *Towards a typology of European languages*, J. Bechert, G. Bernini, C. Buridant (eds.), Berlin: de Gruyter, 307-315.
- , CAMPBELL, L. (1995): *Historical syntax in cross-linguistic perspective*. Cambridge: CUP.
- HASPELMATH, M. (1993): *A grammar of Lezgian*. Berlin: de Gruyter.
- (2003): The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In: *The new psychology of language*, M. Tomasello (ed.), New York: Erlbaum, vol. 2., 211-243.

- (2004a): Coordinating constructions: an overview. In: *Coordinating constructions*, M. Haspelmath (ed.). Amsterdam: Benjamins, 3–39.
- (ed.) (2004b): *Coordinating constructions*. Amsterdam: Benjamins, 3–39.
- (2004c): On directionality in language change with particular reference to grammaticalization. In: O. Fischer, M. Norde, H. Perridon (eds.), 17-44.
- (2005): Nominal and Verbal Conjunction. In: M. Haspelmath et al. (eds.), 262-265.
- (2007): Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology* 11(1), 119-132.
- , GIL, D., DRYER, M., COMRIE, B. (eds.) (2005): *The World Atlas of Language Structures*. Oxford: OUP.
- , KÖNIG, E. (eds.) (1995): *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms – adverbial participles, gerunds*. Berlin: de Gruyter.
- , KÖNIG, E., OESTERREICHER, W., RAIBLE, W. (eds.) (2001): *Language typology and language universals: An international handbook*. Vol. 1-2. Berlin: de Gruyter.
- HAWKINS, J.A. (1983): *Word order universals*. New York: Academic Press.
- HAY, J., BAUER, L. (2007): Phoneme inventory size and population size. *Language* 2, 388-400.
- HAYS, D. G. (1978): Phonology, Cultural Evolution, and Geographic Region. Manuscript.
- HEATH, J. (1978): *Linguistic diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- (1985): Discourse in the field: clause structure in Ngandi. In: *Grammar Inside and Outside the Clause: Some Approaches to Theory from the Field*, J. Nichols, A.C. Woodbury (eds.). Cambridge: CUP, 89-112.
- (2004): Coordination: an adaptationist view. In: M. Haspelmath (ed.) (2004b), Amsterdam: Benjamins, 67-88.
- HEESCHEN, V. (1998): *An ethnographic grammar of the Eipo language: spoken in the central mountains of Irian Jaya (West New Guinea), Indonesia*. Berlin: Dietrich Reimer.
- , SCHIEFENHÖVEL, W. (1983): *Wörterbuch der Eipo-Sprache: Eipo-Deutsch-Englisch*. Berlin: Dietrich Reimer.
- HEINE, B. (1994): Areal influence on grammaticalization. In: *Language contact and language conflict*, M. Piitz (ed.). Amsterdam: Benjamins, 56-68.

- , CLAUDI, U., HÜNNEMEYER, F. (1991a): *Grammaticalization: a conceptual framework*. Chicago: University of Chicago Press.
- , CLAUDI, U., HÜNNEMEYER, F. (1991b): From cognition to grammar – evidence from African languages. In: E.C. Traugott, B. Heine (eds.), vol. 1., 149-87.
- , KUTEVA, T. (2001): Convergence and divergence in the development of African languages. In: A.Y. Aikhenvald, R.M.W. Dixon (eds.) (2001b), 393-411.
- , KUTEVA, T. (2002): *World Lexicon of Grammaticalization*. Cambridge: CUP.
- , KUTEVA, T. (2003): Contact-induced grammaticalization. *Studies in Language* 27(3), 529-72.
- , KUTEVA, T. (2005): *Language contact and grammatical change*. Cambridge: CUP.
- , KUTEVA, T. (2006): *The changing languages of Europe*. Oxford: OUP.
- , KUTEVA, T. (2007): *The Genesis of Grammar: A Reconstruction*. Oxford: OUP
- HENGEVELD, K. (1993): Semantic type, Factivity and the expression of adverbial clauses. In: *The internal structure of adverbial clauses. EURO TYP Working Papers*, vol. 5., K. Hengeveld (ed.), 119-132.
- (forthcoming): The grammaticalization of tense and aspect. In: *The Oxford Handbook of Grammaticalization*, B. Heine, H. Narrog (eds.). Oxford: OUP.
- HERCUS, L.A. (1994): *A grammar of the Arabana-Wangkangurru language: Lake Eyre Basin, South Australia*. Canberra: Australian National University.
- HERSHBERGER, W. (1988): Psychology as a conative science. *American Psychologist* 43(10), 823-824.
- HILGARD, E. R. (1980): The trilogy of mind: Cognition, affection, and conation, *Journal of the History of the Behavioral Sciences* 16, 107-117.
- HILL, J. H., HILL, K.C. (1981): Variation in relative clause construction in modern Nahuatl. In: *Nahuatl Studies in Memory of Fernando Horcasitas*, F. Karttunen (ed.). Austin: University of Texas Press, 89-103.
- HIRSCH, E.D., JR (1977): *The Philosophy of Composition*. Chicago: University of Chicago Press.
- HOCK, H.H. (1986): *Principles of Historical Linguistics*. Berlin: de Gruyter
- HOLMER, A. (1996): *A parametric grammar of Seediq*. Lund: Lund University Press.
- (2002): The morphology and syntax of Seediq focus. In: *The history and typology of western Austronesian voice systems*, F. Wouk, M. Ross (eds.). Canberra: Pacific Linguistics, 333-354.

- HOPPER, P.J. (1979): Aspect and foregrounding in discourse. In: *Discourse and Syntax*, T. Givón (ed.). New York: Academic Press, 213–241.
- (1987): Emergent grammar. In: *Proceedings of the 13th Annual Meeting of the Berkeley Linguistic Society*, J. Aske, N. Beery, L. Michaelis, H. Filip (eds.). Berkeley: Berkeley Linguistic Society, 139-157.
- (1991): On some principles of grammaticization. In: E. C. Traugott, B. Heine (eds.). vol. 1., 17-36.
- (1995): The Category ‘Event’ in Natural Discourse and Logic. In: *Discourse, Grammar and Typology: Papers in Honor of John WM Verhaar*, A. Werner, T. Givón, S.A. Thompson (eds.). Amsterdam: Benjamins, 139-150
- , THOMPSON, S. (1980): Transitivity in grammar and discourse. *Language* 56(2), 251-299.
- , TRAUOGOTT, E.C. (2003)[1993]: *Grammaticalization*. Cambridge: CUP.
- HORIE, K. (2001): Complement clauses. In: Haspelmath et al. (eds.), vol. 2., 979–993.
- HUALDE, J. I., ORTIZ DE URBINA, J. (eds.) (2003): *A grammar of Basque*. Berlin: de Gruyter.
- HUDDLESTON, R. (1984): *Introduction to the Grammar of English*. Cambridge: CUP.
- , PULLUM, G.K. ET AL. (2002): *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- HUITT, W. (1996): The mind. *Educational Psychology Interactive*. Valdosta: Valdosta State University.
(available at <http://www.edpsycinteractive.org/topics/summary/mind.html>. Last accessed: 09/07/2010).
- , Cain, S. (2005): An overview of the conative domain. *Educational Psychology Interactive*. Valdosta: Valdosta State University.
(available at <http://teach.valdosta.edu/whuitt/brilstar/chapters/conative.doc>. Last accessed: 09/07/2010).
- HUTCHISON, J.P. (1976): *Aspects of Kanuri syntax*. PhD dissertation, Indiana University.
- (1981): *A reference grammar of the Kanuri language*. Madison: University of Wisconsin-Madison.
- HUTTAR, G.L., HUTTAR M.L. (1994): *Ndyuka*. London: Routledge.
- IWASAKI, SH. (2002): *Japanese*. Amsterdam: Benjamins.
- , INGKAPHIROM, P. (2005): *A reference grammar of Thai*. New York: CUP.

- J**ACKENDOFF, R. (1999) Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3(7), 272-279.
- JACKSON, H. (1995): *Grammar and Meaning: A Semantic Approach to English Grammar*. London: Longman.
- JACOBSON, S.A. (1996): *Yup'ik Eskimo dictionary*. Fairbanks: Alaska Native Language Center.
- JACQ, P., SIDWELL, P. (1999): *Sapuan (Sepuar)*. München: Lincom.
- JAGGAR, P.J. (2001): *Hausa*. Amsterdam: Benjamins.
- JAHN, E.H. (1989): Language planning and language change. In: *Language change: contribution to the study of its causes*, L.E. Breivik, E.H. Jahr (eds.). Berlin: de Gruyter, 99-113.
- JANZEN, T., SHAFFER, B. (2002): Gesture as the substrate in the process of ASL grammaticization. In: *Modality and structure in signed and spoken languages*, R.P. Meier, K. Cormier, D. Quinto-Pozos (eds.). Cambridge: CUP, 199-223.
- JOHNSON, H. L., CHAPMAN, R. S. (1980): Children's Judgment and Recall of Causal Connectives: A Developmental Study of "Because," "So," and "And". *Journal of Psycholinguistic Research* 9(3), 243-260.
- JONES, CH. (1991): *Purpose clauses: syntax, thematics, and semantics of English purpose constructions*. Dordrecht: Kluwer Academic.
- JONES, R. (1998): *The Boko/Busa language cluster*. München: Lincom.
- JOSEPH, B. D., JANDA, R. D. (1988): The how and why of diachronic morphologization and demorphologization. In: *Theoretical Morphology: Approaches in Modern Linguistics*, M. Hammond, M. Noonan (eds.). New York: Academic Press, 193-210.
- JUNG, D. (2002): Clause combining in Apache. In: *Problems of polysynthesis*, N. Evans, H.J. Sasse (eds.). Berlin: Akademie Verlag, 167-83.
- K**ALMAR, I. (1985): Are there really no primitive languages? In: *Literacy, Language and Learning. The nature and consequences of reading and writing*, D.R. Olson, N. Torrance, A. Hildyard, A. (eds.). Cambridge: CUP.
- K**ARLSSON, F., MIESTAMO, M., SINNEMÄKI, K. (eds.) (2008): *Linguistic Complexity. Typology, Contact, Change*. Amsterdam: Benjamins.
- KAUFMANN, S. (2005): Conditionals. In: K. Brown (ed.), vol. 3., 6-9.
- KAY, P. (1976): Discussion of papers by Kiparsky and Wescott. *Origins and Evolution of Language and Speech. Annals of the New York Academy of Sciences* 280. New York: The New York Academy of Sciences, 117-119.

- KAZENIN, K. I. (1994): Split syntactic ergativity: toward an implicational hierarchy. *Sprachtypologie und Universalienforschung* 47, 78-98.
- KÄMPFE, H.R. (1995): *Abriß der tschuktschischen Grammatik: auf der Basis der Schriftsprache*. Wiesbaden: Harrassowitz.
- KHRAKOVSKIJ, V.S. (ed.) (2005): *Typology of conditional constructions*. München: Lincom.
- KIELHOFNER, G (2002): Volition. In: *Model of Human Occupation: Theory and application*, G. Kielhofner (ed.). Baltimore: Lippencott Williams and Wilkins, 33–50.
- KIHM, A. (1994): *Kriyol syntax: the Portuguese-based Creole language of Guinea-Bissau*. Amsterdam: Benjamins
- KILIAN-HATZ, CH. (2008): *A Grammar of Modern Khwe (Central Khoisan)*. Cologne: Rüdiger Köppe.
- KIPARSKY, P., KIPARSKY, C. (1970): Fact. In: *Progress in Linguistics*, M. Bierwisch, K.E. Heidolph (eds.). The Hague: Mouton, 143-173.
- KIRTON, J., CHARLIE, B. (1996): *Further aspects of the grammar of Yanyuwa, Northern Australia*. Canberra: Australian National University.
- KLAMER, M. (2000): How report verbs become quote markers and complementizers. *Lingua* 110, 69-98.
- KLOSS, H. (1967): ‘Abstand languages’ and ‘ausbau languages’. *Anthropological Linguistics* 9, 29–41.
- KOPTJEVSKAJA-TAMM, M. (1994): Finiteness. In: *The Encyclopedia of Language and Linguistics*, R.E. Asher (ed.) Oxford: Pergamon Press, vol. 3., 1245-1248.
- KOROMA, R. (1994): *Die Morphosyntax des Gola*. Köln: Institute für Afrikanistik.
- KORTMANN, B. (1997): *Adverbial subordination: a typology and history of adverbial subordinators based on European languages*. Berlin: de Gruyter.
- (1999): Iconicity, Typology and Cognition. In: *Form Miming Meaning: Iconicity in Language and Literature*, M. Nänny, O. Fischer (eds.). Amsterdam: Benjamins, 375-392.
- (2001): Adverbial Conjunctions. In: M. Haspelmath et al. (eds.), vol.1., 842-854.
- KOUL, O.N. (2008): *Modern Hindi Grammar*. Springfield: Dunwoody Press.
- KÖNIG, E. (1985a): On the history of concessive connectives in English. Diachronic and synchronic evidence. *Lingua* 66, 1-19.
- (1985b): Where do concessives come from? On the development of concessive connectives. In: *Historical semantics. Historical word formation*, Jacek Fisiak (ed.). Berlin: de Gruyter, 263-282.

— (1986): Conditionals, concessive conditionals and concessives: areas of contrast, overlap, and neutralization. In: E.C., et al. (eds.), 229-246.

— (1988): Concessive connectives and concessive sentences: crosslinguistic regularities and pragmatic principles. In: *Explaining Language Universals*, J.A. Hawkins (ed.). Oxford: Blackwell, 145-166.

— (1991). Concessive relations as the dual of causal relations. In: D. Zaefferer (ed.) (1991b), 190-209.

KÖVECSES, Z., RADDEN, G. (1998) Metonymy: developing a cognitive linguistic view. *Cognitive Linguistics* 9, 37-77.

KRISTOFFERSEN, L. (1992): Predication and derivation in West Greenlandic. In: *Layered clause structure and reference in a functional perspective*. M. Fortescue, P. Harder, L. Kristoffersen (eds.). Amsterdam: Benjamins.

KROLL, B. (1977): Combining ideas in written and spoken English: A look at subordination and coordination. In: *Discourse across Time and Space*, E. Keenan, T.L. Bennett (eds.). Los Angeles: University of Southern California, 69-108.

KUNO, S. (1973): *The structure of the Japanese language*. Cambridge: MIT.

KURYŁOWICZ, J. (1949): La nature des procès dits ‘analogiques’. *Acta Linguistica* 5, 15-37. Reprinted in: (1966) *Readings in linguistics II*, E. P. Hamp, F. W. Householder and R. Austerlitz (eds.). Chicago: University of Chicago Press, 158-174.

KUTEVA, T. (1998): Large linguistic areas in grammaticalization: Auxiliation in Europe. *Language Sciences* 20(3), 289-311.

— (2000): Areal grammaticalization: The case of the Bantu-Nilotic borderland. *Folia Linguistica* 34(3-4), 267-83.

LANGACKER, R. W. (1977): Syntactic reanalysis. In Ch.N. Li (ed.), 57-139

— (1990): Subjectification. *Cognitive Linguistics* 1, 5-38.

LEGATE, J.A. (2002): *Warlpiri: theoretical implications*. PhD dissertation, Massachusetts Institute of Technology.

LEHMANN, CH. (1982): Thoughts on Grammaticalization. A programmatic Sketch. Vol. I., *Arbeiten des Kölner Universalien-Projekts* 48. Köln: Institut für Sprachwissenschaft der Universität.

— (1988): Towards a typology of clause linkage. In: J. Haiman, S.A.Thompson (eds.), 181-225.

LEHMANN, T. (1993): *A grammar of modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.

- LENKER, U., MEURMAN-SOLIN, A. (eds.) (2007): *Connectives in the History of English*. Amsterdam: Benjamins.
- LESSAU, D.A. (1994): *A Dictionary of Grammaticalization*. Bochum: Universitätsverlag.
- LEVINSON, S. (2000): *Presumptive Meaning: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.
- LEWIS, M.P. (ed.) (2009): *Ethnologue: Languages of the World*. Dallas: SIL. (available at <http://www.ethnologue.com/>. Last accessed: 09/07/2010).
- LI, CH.N. (ed.) (1977): *Mechanisms of Syntactic Change*. Austin: University of Texas Press.
- , THOMPSON, S.A. (1989): *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- LIBERMAN, P. (1984): *The biology and evolution of language*. Cambridge: Harvard University Press.
- LICHTENBERK, F. (1979): Syntactic iconism, coordination, subordination and language evolution. *Hawaii Working Papers in Linguistics* 11, 79-89.
- (1991): On the gradualness of grammaticalization. In: E.C. Traugott, B. Heine (eds.), vol. 1., 37-80.
- LIEBERMAN, P. (1984): *The Biology and Evolution of Language*. Cambridge: Harvard University Press.
- LIN, H.T. (1984): *Essential grammar for modern Chinese*. Boston: Cheng and Tsui.
- (2001): *A grammar of Mandarin Chinese*. München: Lincom.
- LONGACRE, R.E. (2007) [1985]: Sentences as combinations of clauses. In: T. Shopen (ed.), vol. 2., 372-420.
- LOOGMAN, A. (1956): *Swahili Grammar and Syntax*. Pittsburgh: Duquesne University Press.
- LORD, C. D. (1976): Evidence for syntactic reanalysis: from verb to complementizer in Kwa. In: *Papers from the parasession on diachronic syntax*, S.B. Steever, C.A. Walker, S.S. Mufwene, (eds.). Chicago: Chicago Linguistic Society, 179-191.
- LUST, B., MERVIS C.A. (1980): Development of coordination in the natural speech of young children. *Journal of Child Language* 7, 219-304
- MACWHINNEY, B., BATES, E. (eds.) (1989): *The crosslinguistic study of sentence processing*. New York: CUP.
- MADAN, J.F. (1991): *A Standard Swahili-English Dictionary*. Oxford: OUP

- MAINWARING, G.B. (1971): *A grammar of the Lepcha language*. New Delhi: Manjusri Publishing House.
- MARLETT, S.A. (1981): *The structure of Seri*. San Diego: University of California
- MARSCHARK, M., PETERSON, R., WINSTON, E.A. (eds.) (2005): *Sign Language Interpreting and Interpreter Education Directions for Research and Practice*. Oxford: OUP.
- MARTIN, S.A. (1997): *Dagur Mongolian grammar, texts, and lexicon: based on the speech of Peter Onon*. Richmond: Curzon.
- MATHER, E., MEADE, M., MIYAOKA, O. (2002): *Survey of Yup'ik grammar*. Kyoto: ELPR.
- MATHIOT, M. (ed.) (1979): *Ethnolinguistics: Boas, Sapir and Whorf revisited*. The Hague: Mouton.
- MATRAS, Y., SAKEL, J. (eds.) (2007): *Grammatical borrowing: a cross-linguistic survey*. Berlin: de Gruyter.
- MAURI, C. (2008): *Coordination Relations in the Languages of Europe and Beyond*. Berlin: de Gruyter.
- MAW, J. (1974): *Swahili style: a study*. London: University of London.
- MCGREGOR, R.S. (1995): *Outline of Hindi Grammar with Exercises*. Oxford: OUP.
- MCSHANE, J. (1991): *Cognitive Development. An information Processing Approach*. Oxford: Blackwell Publishers.
- MCWHORTER, J.H. (2001a): *The Power of Babel: A Natural History of Language*. New York: Times Books/Henry Holt.
- (2001b): The world's simplest grammars are Creole grammars. *Linguistic Typology*, 6, 125-66.
- (2008): Why does a language undress? Strange cases in Indonesia. In: F. Karlsson, M. Miestamo, K. Sinnemäki, K. (eds.), 167-191.
- MEILLET, A. (1914): Le problème de la parenté des langues. *Scientia* 35(15). Reprinted in: (1921) *Linguistique historique et linguistique générale*. Paris: Champion, 76-101.
- MELÉNDEZ, M.A. (1998): *La lengua achagua: estudio gramatical*. Bogotá: Universidad de las Andes.
- MILLER, J. (2006): Clause Structure in Spoken Discourse. In: K. Brown (ed.), vol. 2., 481-483.

- , WEINERT, R. (1998): *Spontaneous spoken language. Syntax and Discourse*. Oxford: Clarendon.
- MISHRA, M. (2006): *A Santhali grammar*. Delhi: Shipra Publications.
- MITHUN, M. (1980): A functional approach to syntactic reconstruction. In: *Papers from the fourth International Conference on Historical Linguistics*, E.C. Traugott, R. LaBrum, S. Shepherd (eds.). Amsterdam: Benjamins, 87-96.
- (1984): How to Avoid Subordination. In: C. Brugman, M. Macaulay (eds.), 493-509.
- (ed.) (1996): *Prosody, grammar, and discourse in central Alaskan Yup'ik*. Santa Barbara: University of California.
- MOLENCKI, R. (2007): The evolution of “*since*” in medieval English. In: U. Lenker, A. Meurman-Solin (eds.), 97–113.
- MONTAUT, A. (2004): *A grammar of Hindi*. München: Lincom.
- MORENO CABRERA, J.C. (1998): On the relationship between grammaticalization and lexicalization. In G. Ramat, P.J. Hopper, A. Hopper (eds.), 209-227.
- MORSE, N.L., MAXWELL, M.B. (1999): *Cubeo grammar*. Dallas: SIL.
- MOSER, M.B. (1996): *Seri de Sonora*. México: El Colegio de México.
- MOUS, M., ODA, O. (2009): The Semantics of Clause Linking in Konso. In: R.M.W. Dixon, A.Y. Aikhenvald (eds.), 336-355.
- MYACHINA, E.N. (1981): *The Swahili language: a descriptive grammar*. London: Routledge.
- MYHILL, J., HIBIYA, J. (1988): The discourse function of clause-chaining. In: J. Haiman, S.A. Thompson (eds.), 361-398.
- NASH, D.G. (1980): *Topics in Warlpiri grammar*. Cambridge: MIT.
- NEDJALCOV, V.P. (1990): Osnovnye tipy deepricastij. [Typological parameters of converbs]. In: *Tipologija i grammatika*, V.S. Xrakovskij (ed.). Moscow: Nauka, 36-59. Translated in: M. Haspelmath, E. König (eds.) (1995), 97-136.
- , NEDJALCOV, I.V. (1987): On the typological characteristics of converbs. In: *Symposium on language universals*, T. Help (ed.). Tallinn, 75-79.
- NERLICH, B., CLARKE, D.D. (1999): *Synecdoche as a cognitive and communicative strategy*. In: Blank, A., Koch, P. (eds.). 197-214.
- NETTLE, D. (1999): *Linguistic diversity*. Oxford: OUP.

- NEUKOM, L. (2001): *Santali*. München: Lincom.
- NEVIS, J.A. (1984): A non-endoclititic in Estonian. *Lingua* 64, 209-224.
- (1985): Language-external evidence for clitics as words: Lappish particle clitics. *Chicago Linguistic Society* 21, 289-305.
- NEWMAN, P. (2000): *The Hausa language: an encyclopedic reference grammar*. New Haven: Yale University Press.
- NEWMeyer, F.J. (1998): *Language form and language function*. Cambridge: MIT Press.
- (2002): Uniformitarian Assumptions and Language Evolution Research. In: *Transitions to Language*, A. Wray (ed.). Oxford: OUP, 359-375.
- NICHOLS, J. (2009): Linguistic complexity: a comprehensive definition and survey. In: . Sampson, D. Gil, P. Trudgill (eds.), 110-125.
- NIKOLAEVA, I. (ed.) (2007): *Finiteness: Theoretical and empirical foundations*. Oxford: OUP.
- NOONAN, M. (1992): *A grammar of Lango*. Berlin: de Gruyter.
- (2007)[1985]: Complementation. In: T. Shopen (ed.), vol. 2., 52-150.
- NORDLINGER, R. (1998): *A grammar of Wambaya, Northern Territory (Australia)*. Canberra: Australian National University.
- O**DA, O. (2000): *An overview of complex sentences and complement clauses in Konso*. MA thesis, Addis Abba University.
- OFORI, S.A. (2006): *Topics in Akan grammar*. PhD dissertation, Indiana University.
- OLSON, D.R. (1977): From utterance to text: the bias of language in speech and writing. *Harvard Educational Review* 47, 257-81.
- (1988): From utterance to text. In: *Perspectives on literacy*, E. R. Kintgen, B. M. Kroll, M. Rose (eds.). Carbondale: Southern Illinois University Press, 175-189.
- ONISHI, M. (1994): Semantic Primitives in Japanese. In: C. Goddard, A. Wierzbicka (eds.), 361–386.
- OSAM, E.K.A. (1994): *Aspects of Akan grammar: a functional perspective*. PhD dissertation, University of Oregon.
- P**ALMER, F. R. (1986): *Mood and modality*. Cambridge: CUP.
- PARSONS, T. (1989): The Progressive in English: Events, States and Processes. *Linguistics and Philosophy*, 12, 213–41.

- (1990): *Events in the Semantics of English. A Study in Subatomic Semantics*. Cambridge: MIT.
- (1991): Tropes and Supervenience. *Philosophy and Phenomenological Research* 51, 629–32.
- PENSALFINI, R.J. (1997): *Jingulu grammar, dictionary, and texts*. Cambridge: MIT.
- (2003): *A grammar of Jingulu: an Aboriginal language of the Northern Territory*. Canberra: Australian National University.
- PÉREZ QUINTERO, M.J. (2002): *Adverbial subordination in English: A functional approach*. Amsterdam: Rodopi.
- PERICLIEV, V. (2004): There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. *Linguistic Typology* 8, 376-83.
- PERKINS, R.D. (1988): The covariation of culture and grammar. In: *Studies in Syntactic Typology*, M. Hammond, E.A. Moravcsik, J. Wirth (eds.). Amsterdam: Benjamins, 359-378.
- (1992): *Deixis, Grammar and Culture*. Amsterdam: Benjamins.
- PETHŐ, G. (2001): What is polysemy? a survey of current research and results. In: *Pragmatics and the flexibility of word meaning*, E. Németh, K. Bibok (eds.): Amsterdam: Elsevier. 175–224.
- PISARKOWA, K. (1984): *Historia składni języka polskiego*. Wrocław: Ossolineum.
- PLAISIER, H. (2006): *A grammar of Lepcha*. Leiden: Brill.
- PO-CHING, Y., RIMMINGTON, D. (2000): *Chinese. An Essentials grammar*. London: Routledge.
- PODLESSKAYA, V. (2001): Conditional constructions. In. M. Haspelmath et al. (eds.), vol. 2., 998-1009.
- POOLE, M. E., AND FIELD, T.W. (1976): A comparison of oral and written code elaboration. *Language and Speech* 19, 305-11.
- POST, M.W. (2007): *A grammar of Galo*. PhD dissertation. La Trobe University.
- (2009): The Semantics of Clause Linking in Galo. In: R.M.W. Dixon, A.Y. Aikhenvald (eds.), 74-95
- QUIRK, R., GREENBAUM, S., LEECH, G., SVARTVIK, J. (1985): *A comprehensive grammar of the English language*. London: Longman.

- R**AMAT, A. G., HOPPER, P.J., HOPPER, A. (eds.) (1998): *The Limits of Grammaticalization*. Amsterdam: Benjamins.
- RAMAT, P. (1992): Thoughts on degrammaticalization. *Linguistics* 30, 549-60.
 — (1998): Typological comparison and linguistic areas: some introductory remarks, *Language Sciences* 20 (3), 227–240.
- (1999): Linguistic categories and linguists' categorizations. *Linguistics* 37, 157–180.
- RAU, D. V. (2005): Iconicity, tense, aspect, and mood morphology in Yami, Concentric. *Studies in Linguistics* 31(1), 65-94.
- , DONG, M. (2006): *Yami texts with reference grammar and dictionary*. Taipei: Academia Sinica.
- RAVIN, Y., LEACOCK, C. (eds.) (2000): *Polysemy*. Oxford: OUP.
- REED, I., MIYAOKA, O., JACOBSON, S., AFCAN, P., KRAUSS, M. (1997): *Yup'ik Eskimo grammar*. Fairbanks: University of Alaska.
- REESINK, G.P. (1999): *A grammar of Hatam, Bird's Head Peninsula, Irian Jaya*. Canberra: Australian National University.
- (ed.) (2002): *Languages of the eastern Bird's Head*. Canberra: Australian National University.
- REH, M. (1985): *Die Krongo-Sprache (Niino mó-di: Beschreibung, Texte, Wörterverzeichnis)*. Berlin: Dietrich.
- RIJKHOFF, J., BAKKER, D., HENGEVELD, K., KAHREL, P. (1993): A method of language sampling. *Studies in Language* 17, 169-203.
- RIJKHOFF, J., BAKKER, D. (1998): Language sampling. *Linguistic Typology* 2-3, 263-314.
- ROSATO, M. SANTANDREA, S. (1980): *Didinga (= Didina) grammar and dictionary: (Sudan-Equatoria)*. S. Santandrea (ed.). Rome.
- RUBINO, C.R.G. (2000): *Ilocano dictionary and grammar: Ilocano – English, English – Ilocano*. Honolulu: University of Hawaii Press.
- RUDE, N.E. (1985): *Studies in Nez Perce grammar and discourse*. PhD dissertation, University of Oregon.
- RUHL, C. (1989): *On monosemy*. Albany: Suny Press.
- RUHLEN, M. (1987):. *A Guide to the World's Languages, Volume 1: Classification*. Stanford: Stanford University Press.

- SALONI, Z., ŚWIDZIŃSKI, M. (2001): *Składnia współczesnego języka polskiego*. Warsaw: PWN.
- SAMPSON, G., GIL, D., TRUDGILL, P. (eds.) (2009): *Language Complexity as an Evolving Variable*. Oxford: OUP.
- SAPIR, E. (1921): *Language*. New York: Harcourt, Brace and World.
- SAWKA, K.S. (2001): *Aspects of Mayogo grammar*. MA thesis, University of Texas at Arlington.
- SAXENA, A. (1988): On syntactic convergence: The case of the verb *say* in Tibeto-Burman. In: *Proceedings of the fourteenth annual meeting of the Berkeley Linguistics Society*, Sh. Axmaker, A. Jaisser, H. Singmaster (eds.). Berkeley: University of California, 375–388
- (1995): Unidirectional grammaticalization: Diachronic and cross-linguistic evidence, *Sprachtypologie und Universalienforschung* 48 (4), 350–372.
- SCHACHTER, J. (1971): *Presupposition and counterfactual sentences*. Ph.D dissertation, University of California.
- SCHACHTER, P., SHOPEN T. (2007) [Schachter, P. 1985]: Parts-of-speech systems. In: T. Shopen (ed.), vol. 1., 1-61.
- SCHLOTTMANN, A., SURRIAN, L. (1999): Do 9-month-olds perceive causation-at-a-distance? *Perception* 28, 1105-1113.
- SCHMIDTKE-BODE, K. (2009): *A Typology of Purpose Clauses*. Amsterdam: Benjamins.
- SCORZA, D. (1973): *Sentence structures of the Au language*. Ukarumpa: SIL.
- (1976): *Au language word, phrase, clause*. Manuscript. (available at <http://www.sil.org/pacific/png/abstract.asp?id=50160>. Last accessed: 09/07/2010).
- (1985): A sketch of Au morphology and syntax, *Papers in New Guinea Linguistics* 22, 215-73.
- SEXTON, A. L. (1999): Grammaticalization in American sign language. *Language Sciences* 21, 105-41.
- SHOPEN T. (ed.) (2007)[1985]: *Language Typology and Syntactic Description*, vol. 1-3. Cambridge: CUP.
- SIEWIERSKA, A. (1991): *Functional grammar*. London: Routledge.
- SIMON, H. J., WIESE, H. (2002): *Pronouns: Grammar and Representation*. Amsterdam: Benjamins.
- SIMPSON, J. (1991): *Warlpiri morpho-syntax: a lexicalist approach*. Dordrecht: Kluwer.

- SIMPSON, J., WEINER, E. (eds.) (1989): *The Oxford English Dictionary*. 2nd edition. Oxford: Clarendon Press.
- SINNEMÄKI, K. (2009): Complexity and size of speech community. In: G. Sampson, D. Gil, P. Trudgill (eds.), 126-141.
- SMIRNOVA, M.A. (1982): *The Hausa language: a descriptive grammar*. London: Routledge.
- SMYTH, D. (2002): *Thai: an essential grammar*. London: Routledge.
- SOLNIT, D.B. (1997): *Eastern Kayah Li : grammar, texts, glossary*. University of Hawai'i Press.
- SPERBER, D., WILSON, D. (1987): *Precis of Relevance: Communication and Cognition. Behavioral and Brain Sciences* (10), 697-754.
- STASSEN, L. (1985): *Comparison and Universal Grammar*. Oxford: Basil Blackwell
- (2005). Noun Phrase Conjunction. In: M. Haspelmath et al. (eds.), 258-261.
- STOLZ, CH., STOLZ, T. (2001): Mesoamerica as a linguistic area. In: Haspelmath et al. (eds.), vol. 2., 1542-1553.
- STOLZ, T., BAKKER, D., PALOMO, R. (eds.) (2008): *Aspects of Language Contact*. Berlin: de Gruyter.
- STORM, H. (2003): *A handbook of Japanese grammar*. München: Lincom.
- STROM, C. (1992): *Retuarã syntax*. Dallas: SIL.
- SWADESH, M. (1971): *The Origin and Diversification of Language*. J.F. Sherzer (ed.). Chicago: Aldine-Atherton
- SWEETSER, E. E. (1984): *Semantic structure and semantic change: A cognitive linguistic study of modality, perception, speech acts and logical relations*. PhD dissertation, University of California at Berkeley.
- (1990): *From Etymology to Pragmatics*. Cambridge: CUP
- TALMY, L. (1978): Figure and ground in complex sentences. In: *Universals of Human Language: Syntax*, J.H. Greenberg, Ch.A. Ferguson E.A. Moravcsik (eds.). Stanford: Stanford University Press, vol. 4., 625–649.
- (2001): *Toward a Cognitive Semantics*. Vol. 1-2. Cambridge: MIT Press.
- TALLON, A. (1997): *Head and heart: Affection, cognition, volition as triune consciousness*. New York: Fordham University.
- TARPENT, M.L. (1987): *A grammar of the Nisgha language*. PhD dissertation, University of Victoria.

- TAULI, V. (1983): *Standard Estonian grammar*. Uppsala: Almqvist and Wiksell.
- TERRILL, A. (2003): *A grammar of Lavukaleve*. Berlin: de Gruyter.
- THOMASON, S.G., KAUFMAN, T. (1988): *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- THOMPSON, S.A. (1985): 'Subordination' in Formal and Informal Discourse. In: *Meaning, Form, and Use in Context: Linguistic Applications*, D. Schiffrin (ed.). Georgetown: Georgetown University Press, 85-94.
- , LONGACRE R.E., HWANG, SH.J.J. (2007) [Thompson, S.A., Longacre R.E 1985]: Adverbial clauses. In: T. Shopen (ed.), vol. 2., 237-300.
- THORNELL, CH. (1997): *The Sango language and its lexicon (Senda-yanga ti sängö)*. Lund: Lund Univeristy Press.
- THURGOOD, G. (1986): Lolo-Burmese subordinators from case postpositions: Several partial etymologies. In: *Proceedings of the Second Annual Meeting of the Pacific Linguistics Conference*, S. DeLancey R.S. Tomlin (eds.), Eugene: University of Oregon, 449-454.
- TIKKANEN, B. (1995): Burushaski Converbs in their South and Central Asian Areal Context. In: M. Haspelmath, E. König (eds.). Berlin: de Gruyter, 487-528.
- TIMBERLAKE, A. (2007): *Aspect, tense, mood*. In: T. Shopen (ed.), vol. 3., 280-333.
- TINGSABADH, M.R.K., ABRAMSON, A.S. (eds.) (2001): *Essays in Tai linguistics*. Bangkok: Chulalongkorn University Press.
- TOMASELLO, M. (2005): *Constructing a language: a usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- TRAUGOTT, E.C. (1985): Conditional markers. In: J. Haiman (ed.), 289-307.
- (1989): On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 57, 33-65.
- (1990): From less to more situated in language: The unidirectionality of semantic change. In: *Papers from the 5th International Conference on English Historical Linguistics*, S.M. Adamson, V. Law, N. Vincent, S.M. Wright (eds.). Amsterdam: Benjamins, 497-517.
- (1994): Grammaticalization and lexicalization. In: *The Encyclopedia of Language and Linguistics*, R.E. Asher, J.M.Y. Simpson (eds.). Oxford: Pergamon Press, vol. 3., 1481-1486.
- (1999): The rhetoric of counter-expectation in semantic change: A study in subjectification. In: Blank, A., Koch, P. (eds.), 177-196.

- (2003): From Subjectification to Intersubjectification. In: *Motives for Language Change*, R. Hickey (ed.). Cambridge: CUP, 124-139.
- , HEINE, B. (1991a): Introduction. In: E.C. Traugott, B. Heine (eds.) (1991b), vol. 1., 1-14
- , HEINE, B. (eds.) (1991b): *Approaches to Grammaticalization*. Vol. 1-2. Amsterdam: Benjamins.
- , KÖNIG, E. (1991): The semantics-pragmatics of grammaticalization revisited In: E.C. Traugott, B. Heine (eds.), vol. 1., 189-218.
- , TER MEULEN, A., SNITZER REILLY, J., FERGUSON CH. A. (eds.) (1986): *On conditionals*. Cambridge: CUP.
- TRUDGILL, P. (1989): Contact and isolation in linguistic change. In: *Language Change: Contributions to the Study of its Causes*, L.E. Breivik, E.H. Jahr (eds.). Berlin: de Gruyter, 227-237.
- (1996): Dialect typology: Isolation, social network and phonological structure. In: *Towards a Social Science of Language: Papers in Honour of William Labov*, G. Guy, C. Feagin, D. Schiffrin, J. Baugh (eds.): Amsterdam: Benjamins, vol. 1., 3-21.
- (2004a): Linguistic and social typology: the Austronesian migrations and phoneme inventories. *Linguistic Typology* 8, 305-320.
- (2004b): On the complexity of simplification. *Linguistic Typology* 8, 384-388.
- TSUMAGARI, T. (2003): Dagur. In: *The Mongolic languages*, J. Janhunen (ed.). London: Routledge, 129-153.
- VALIN, VAN R.D. (1984): A typology of syntactic relations in clause linkage. In: C. Brugman, M. Macaulay (eds.), 542-58.
- (2005): *Exploring the Syntax-Semantics Interface*. Cambridge: CUP.
- , LAPOLLA, R.J. (1997): *Structure, meaning and function*. Cambridge: CUP.
- VOEGELIN, C.F., VOEGELIN, F.M. (1977): *Classification and index of the world's languages*. New York: Elsevier.
- WATAHOMIGIE, L.J., BENDER, J., YAMAMOTO, A.Y., MAPATIS, E., MANAKAJA, J., POWSKEY, M. (2001): *Hualapai Reference Grammar*. Kyoto: Endangered Languages of the Pacific Rim Project.
- WATSON, J.C.E. (1993): *A syntax of San'ani Arabic*. Wiesbaden: Harrassowitz.
- (2002): *The phonology and morphology of Arabic*. Oxford: OUP.

- WEBER, D.J. (1989): *A grammar of Huallaga (Huánuco) Quechua*. Berkeley: University of California Press.
- (1996): *Una gramática del quechua del Huallaga (Huánuco)*. Lima: Ministerio de Educación and Instituto Lingüístico de Verano.
- WERNER, H. (1997): *Die ketische Sprache*. Wiesbaden: Harrassowitz.
- WHORF, B. (1956): *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, J.B. Carroll (ed.). Cambridge: MIT.
- WIERZBICKA, A. (1979): Ethno-syntax and the philosophy of grammar. *Studies in Language* 3 (3), 313-83.
- (1996): *Semantics: primes and universals*. Oxford: OUP
- (1998): Anchoring linguistic typology in universal semantic primes. *Linguistic Typology* 2, 141-194.
- WILBUR, R.B. (1999): Metrical structure, morphological gaps, and possible grammaticalization in ASL. *Sign Language and Linguistics* 2, 217-44.
- WILLETT, T.L. (1980a): Clause types in Southeastern Tepehuan. *Workpapers of the Summer Institute of Linguistics (University of North Dakota)* 24, 51-72.
- (1980b): Sentence components in Southeastern Tepehuan. *Workpapers of the Summer Institute of Linguistics (University of North Dakota)* 24, 73-96.
- (1984): Subordination in Southeastern Tepehuan discourse". *SIL-México Workpapers* 5. México: Instituto Lingüístico de Verano, 119-130.
- (1987): Discourse strategies in Southeastern Tepehuan. *SIL-México Workpapers* 8. México: Instituto Lingüístico de Verano, 31-98.
- (1991): *A reference grammar of Southeastern Tepehuan*. Dallas: SIL.
- (2000): Las conjunciones de subordinación en el tepehuán del sureste. *Memorias del Quinto Encuentro Internacional de Lingüística en el Noroeste*. Hermosillo-Sonora: Universidad de Sonora.
- WING, C., KOFISKY SCHOLNIC, E. (1981): Children's comprehension of pragmatic concepts expressed in 'because', 'although', 'if' and 'unless. *Journal of Child Language* 8, 347-365.
- WISCHER, U. (2000): Grammaticalization versus lexicalization – “methinks” there is some confusion. In: *Pathways of Change: Grammaticalization in English*, O. Fischer, A. Rosenbach, D. Stein (eds.). Amsterdam: Benjamins, 355-370.
- WITTGENSTEIN, L. (1979): *Wittgenstein's Lectures, 1932-35*, Ambrose, A. (ed.). Oxford: Blackwell.

WOLFF, E. (1993): *Referenzgrammatik des Hausa: zur Begleitung des Fremdsprachenunterrichts und zur Einführung in das Selbststudium*. Hamburg: LIT.

WOOLLAMS, G. (1996): *A grammar of Karo Batak, Sumatra*. Canberra: Australian National University.

WORDICK, F.J.F. (1982): *The Yindjibarndi language*. Canberra: Australian National University.

WRAY, A., GRACE, G.W. (2007): The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543-578.

WRÓBEL, H. (2001): *Gramatyka języka polskiego*. Cracow: Od Nowa.

ZAEFFERER, D.: (1991a): Conditionals and Unconditionals: Cross-Linguistic and Logical Aspects. In: D. Zaefferer (ed.), 210-36.

— (ed.) (1991b): *Semantic Universals and Universal Semantics*. Dordrecht: Foris.

ZEITOUN, E. (2007): *A grammar of Mantauran (Rukai)*. Taipei: Academia Sinica.

INTERNET RESOURCES

The following websites and/or their sub-sites (last accessed: 09/07/2010) have been consulted in the process of collecting information on socio-cultural profiles of the sample languages:

Major websites

AFLANG DIRECTORY

<http://www.humnet.ucla.edu/humnet/aflang>

ALASKA NATIVE LANGUAGE CENTRE

<http://www.uaf.edu/anlc/index.html>

COUNTRIES AND THEIR CULTURES

<http://www.everyculture.com>

DOCUMENTING ENDANGERED LANGUAGES OF THE PACIFIC (DELP) PROJECT

http://www.arts.usyd.edu.au/research_projects/delp

ENDANGERED LANGUAGES OF SIBERIA

<http://lingsib.unesco.ru/en/languages>

ETHNOLOGUE, LANGUAGES OF THE WORLD

<http://www.ethnologue.com>

NATIVE AMERICAN LANGUAGE NET

<http://www.native-languages.org>

OMNIGLOT – THE GUIDE TO LANGUAGES, ALPHABETS AND OTHER WRITING SYSTEMS

<http://www.omniglot.com>

PAPUA NEW GUINEA LANGUAGE RESOURCES

<http://www.sil.org/pacific/png/index.asp>

THE RED BOOK OF THE PEOPLE OF RUSSIAN EMPIRE

<http://www.eki.ee/books/redbook/index1.shtml>

UCLA LANGUAGE MATERIALS PROJECT

<http://www.lmp.ucla.edu>

Websites dedicated to particular communities, their languages and cultures

APACHE JICARILLA

<http://www.jicarilla.net>

BASQUE

<http://www.cd.sc.ehu.es/DOCS/book.SS-G/v2/Euskara.html>

<http://www.buber.net/Basque/Euskara/lang1.html>

CHUKCHI

http://privatewww.essex.ac.uk/~spena/Chukchee/CHUKCHEE_HOMEPAGE.html

ESTONIAN

<http://www.einst.ee/publications/language/language.html>

LEPCHA

<http://www.lepcha.info>

YAMI

<http://yamiproject.cs.pu.edu.tw>

YANYUWA

<http://www.deakin.edu.au/arts-ed/diwurruwurru/yanyuwa/index.htm>

.

APPENDICES

APPENDIX I Genetic affiliation of the sample languages

Names of phyla (including isolated languages) are put in capitals and extinct phyla/families/languages – in italics. The numbers in the brackets indicate for each of the phylum and language family/subfamily how many of the languages required in the initial 100 sample have been included in the final sample.

AFRO-ASIATIC (5/6)

Chadic (1/1): **Hausa**
Omotic (1/1): **Maale**
Semitic (1/1): **Arabic (San'ani)**
Cushitic (1/1): **Konso**
Ancient Egyptian (0/1): -
Berber (1/1): **Shelha (Douiret dialect)**

ALTAIC (2/2)

Altaic Proper (1/1): **Dagur**
Korean-Japanese (1/1): **Japanese**

AMERIND (16/18)

Central Amerind (1/2): **Southeastern Tepehuan**
Ge-Pano-Carib/Macro-Carib (1/1): **Macushi**
Ge-Pano-Carib/Ge-Pano/Macro-Panoan (1/1): **Mocovi**
Ge-Pano-Carib/Ge-Pano/Macro-Ge (1/1): -
Northern Amerind/Hokan (2/2): **Hualapai, Hokan**
Northern Amerind/Penutian (2/2): **Nisga'a, Nez Perce**
Northern Amerind/Almosan-Keresiouan (1/1): **Lillooet (St'at'imcets)**
Equatorial-Tucanoan/Macro-Tucanoan (2/2): **Retuarã (Letuma), Cubeo**
Equatorial-Tucanoan/Equatorial (3/3): **Apurina, Achagua, Baure**
Chibchan-Paezan/Chibchan (1/1): **Rama**
Chibchan-Paezan/Paezan (1/1): **Ika (Arhuaco)**
Andean (1/1): **Quechua Huallaga**

AUSTRALIAN (7/7) Arabana, Jingulu, Pitjantjatjara, Wambaya, Warlpiri, Yanyuwa, Yindjibarndi

AUSTRIC (13/14)

Austro-Tai/Austronesian/Malayo-Polynesian/CE Malayo-Polynesian/E Malayo-Polynesian/S Halmahera-NW New G (1/1): **Taba**
Austro-Tai/Austronesian/Malayo-Polynesian/CE Malayo-Polynesian/C Malayo-Polynesian (1/1): **Leti**
Austro-Tai/Austronesian/Malayo-Polynesian/CE Malayo-Polynesian/E Malayo-Polynesian/Oceanic (1/1): **Vitu**
Austro-Tai/Austronesian/Malayo-Polynesian/W Malayo-Polynesian (3/3): **Batak Karo, Ilokoano, Yami**
Austro-Tai/Austronesian/Atayalic (1/1): **Seediq**
Austro-Tai/Daic (1/1): **Thai**
Austro-Tai/Austronesian/Paiwanic (1/1): **Paiwan**
Austro-Tai/Austronesian/Tsouic (1/1): **Rukai (Mantauran)**
Austroasiatic/Mon-Khmer (2/2): **Sapuan, Jahai**
Austroasiatic/Munda (1/1): **Santali**
Miao-Yao (0/1): -

CAUCASIAN (1/1): Lezgian

CHUKCHI-KAMCHATKAN (1/1): Chukchi

ELAMO-DRAVIDIAN (1/1): Tamil

ESKIMO-ALEUT (1/1): Central Alaskan Yup'ik

INDO-HITTITE (3/4)

Indo-European (3/3): **Polish, English, Hindi**
Anatolian (1): -

INDO-PACIFIC (8/13)

³³ Torricelli (1/1): **Au**
³⁴ Trans-New Guinea (1/1): **Eipo**
West Papuan (1/1): **Hattam**

Sepik-Ramu (1/1): **Yimas**
East Papuan (1/1): **Lavukaleve**
Arai (0/1): -
Sko (1/1): **I'saka (Krisa)**
Andaman Islands (1/1): -
Geelvink Bay (0/1): -
Kwomtari-Baibai (1/1)
Amto-Musian (1/1): **Ama**
East Bird's Head (1/1): **Meyah**
Tasmanian (0/1): -

KHOISAN (1/1): Khwe

SUMERIAN (0/1): -

KET (1/1): Ket

NAHALI (0/1): -

HURRIAN (0/1): -

BURUSHASKI (1/1): Burushaski

MEROITIC (0/1): -

BASQUE (1/1): Basque

ETRUSCAN (0/1): -

NIVKH (1/1): Nivkh

NA-DENE (1/1): Apache Jicarilla

NIGER-KORDOFANIAN (8/9)

Niger-Congo/Niger-Congo Proper/ Central Niger-Congo/ Central Niger-Congo/South C Niger-Congo/Eastern (1/1): **Swahili**
Niger-Congo/Niger-Congo Proper/ Central Niger-Congo/ Central Niger-Congo/South C Niger-Congo/Western (1/1): **Akan**
Niger-Congo/Niger-Congo Proper/ Central Niger-Congo/ Central Niger-Congo/South C Niger-Congo/Ijo-Defaka (0/1): -
Niger-Congo/ Niger-Congo Proper/ Central Niger-Congo/Central Niger-Congo/ North C Niger-Congo(3/3): **Mayogo, Sango, Suppyire**
Niger-Congo/Niger-Congo Proper/ West Atlantic (1/1): **Gola**
Niger-Congo/Mande (1/1): **Boko**
Kordofanian (1/1): **Krongo**

NILO-SAHARAN (4/5): Didinga, Fur, Kanuri, Lango, Ma'di

PIDGINS AND CREOLES (2/2): Kryiol, Ndyuka (Aukan)

SINO-TIBETAN (4/4)

Tibeto-Karen/ Tibeto-Burman (2/2) **Galo, Lepcha**
Tibeto-Karen/Karen (1/1): **Kayah Li**
Sinitic (1/1): **Mandarin**

URALIC-YUKAGHIR (1/1): Estonian

APPENDIX II List of languages with references and names of consultants

Language name	References	Consultant(s)
Achagua	Meléndez (1998)	
Akan	Balmer and Grant (1942) Osam (1994), Ofori (2006),	Charles Morfo
Ama	Årsjö (1994)	
Apache Jicarilla	Jung (2002)	
Apurina	Facundes da Silva (2000)	Sidney da Silva Facundes
Arabana	Hercus (1994)	Luise Hercus
Arabic San'ani	Watson (1993, 2002)	Janet Watson
Au	Scorza (1973, 1976, 1985)	
Basque	Hualde and Ortiz de Urbina (2003)	Xabier Artiagoitia
Batak Karo	Woollams (1996)	
Baure	Danielsen (2007, forthcoming)	Swintha Danielsen
Boko	Jones (1998)	Ross McCallum Jones
Burushaski	Tikkanen (1995), Berger (1998)	Bertil Tikkanen
Central Alaskan Yup'ik	Mithun (1996), Reed et al. (1997), Mather et al. (2002), Jacobson (1996)	Steven Jacobson
Chukchi (Telqep dialect)	Kämpfe (1995), Dunn, M.J. (1999)	Michael Dunn
Cubeo	Morse and Maxwell (1999)	Nancy Morse
Dagur	Chuluu (1994), Martin (1997), Tsumagari (2003)	
Didinga	Rosato, Santandrea (1980)	Nicky de Jong
Eipo	Heesch and Schiefenhövel (1983), Heesch (1998),	Volker Heesch
English	Quirk et. al (1985), Simpson, and Weiner (1989), Pérez Quintero (2002), Huddleston, Pullum et al. (2002),	Ronnie Cann, Ron Looker
Estonian	Harms (1962), Tauli (1983), Erelt (2003)	Mati Erelt
Fur	Beaton (1968)	
Galo	Post (2007, 2009)	Mark Post
Gola	Koroma (1994)	Regine Koroma
Hatam	Reesink (1999, 2002)	Ger Reesink

Language name	References	Consultant(s)
Hausa	Smirnova (1982), Wolff (1993), Newman (2000), Jaggar (2001)	Malami Buba, Philip
Hindi	McGregor (1995), Montaut (2004), Agnithi (2007), Koul (2008)	Jaggar, Paul Newman
Hualapai	Watahomigie et al. (2001)	Amar Singh
I'saka (Krisa)	Donohue and Roque (2002)	Akira Yamamoto
Ika (Arhuaco)	Frank (1990, 1993)	Paul Frank
Ilokano	Rubino (2000)	Carl Rubino
Jahai	Burenhult (2005)	Niclas Burenhult
Japanese	Kuno (1973), Onishi (1994), Alpatov and Podlesskaya (1995), Iwasaki (2002), Storm (2003)	Yuri Shinya, Kaori Miura-Dowman
Jingulu	Pensalfini (1997, 2003)	Robert Pensalfini
Kanuri	Hutchison (1981, 1986), Cyffer (1998)	John Hutchison
Kayah Li	Solnit (1997)	David Solnit
Ket	Werner (1997), Georg (2007)	Andrey Nefedov
Khwe	Kilian-Hatz (2008)	Christa Kilian-Hatz
Konso	Oda (2000), Mous and Oda (2009)	Ongaye Oda
Krongo	Reh (1985)	
Kryiol	Kihm (1994)	Alain Kihm
Lango	Noonan (1992)	
Lavukaleve	Terrill (2003)	Angela Terrill
Lepcha	Mainwaring (1971), Plaisier (2006)	Heleen Plaisier
Leti	Engelenhoven (1995, 2004)	Aone van Engelenhoven
Lezgian	Haspelmath (1993)	
Lillooet (St'át'imcets)	Eijk (1997), Davies (2005)	Jan van Eijk
Ma'di	Blackings and Fabb (2003)	
Maale	Amha (2001)	
Macushi	Abbott (1985, 1991)	

Language name	References	Consultant(s)
Mandarin	Lin (1984, 2001), Li and Thomson (1989), Chappell (1994), Po-Ching and Rimmington (2000)	Wenshan Li
Mayogo	Sawka (2001)	
Meyah	Gravelle (2004)	Gilles Gravelle
Mocoví	Grondona (1998)	
Ndyuka (Aukan)	Huttar and Huttar (1994)	George Huttar
Nez Perce	Aoki (1970, 1994), Rude (1985)	Noel Rude
Nisga'a	Tarpent (1987)	Marie Lucie Tarpent
Nivkh (Gilyak)	Gruzdeva (1998)	Ekaterina Gruzdeva
Paiwan	Egli (1990)	Chao-Lin Li, Chunming
Pitjantjatjara	Eckert and Hudson (1994), Goddard (2001)	
Polish	Bańko (2000), Grzegorzycykowa (1996), Saloni and Świdziński (2001), Wróbel (2001)	Maciej Grochowski
Quechua Huallaga	Weber (1989, 1996)	David Weber
Rama	Grinevald (1990), Craig (1991)	Colette Grinevald
Retuarã (Letuma)	Strom (1992)	Clay Strom
Rukai (Mantauran dialect)	Zeitoun (2007)	Elizabeth Zeitoun
Sango	Bouquiaux (1978), Thornell (1997)	Christina Thornell
Santali	Neukom (2001), Mishra (2006), Ghosh (2008)	Madhusudan Mishra
Sapuan	Jacq and Sidwell (1999)	Paul Sidwell
Seediq	Holmer (1996, 2002)	Arthur Holmer
Seri	Marlett (1981), Moser (1996)	Stephen Marlett
Shelha (Douiret dialect)	Gabsi (2003)	Zouhir Gabsi
Southeastern Tepehuan	Willett (1964, 1987, 1980a,b, 1991, 2000)	Thomas Willett
Suppyire	Carlson (1994)	
Swahili	Ashton (1947), Loogman (1956), Myachina, (1981), Madan (1991)	Assibi Amidu
Taba (East Makian)	Bowden (1997, 2001)	
Tamil (standard form)	Lehmann (1993)	Bhuvana Narasimhan

Language name	References	Consultant(s)
Thai	Tingsabath and Abramson (2001), Smyth (2002), Iwasaki and Ingkaphirom (2005)	Vipas Pothipath, Shoichi Iwasaki
Vitu	Berg and Bachtet (2006)	René van den Berg
Wambaya	Nordlinger (1998)	Rachel Nordlinger
Warlpiri	Nash (1980), Hale (1983), Simpson (1991), Legate (2002)	Jane Simpson
Yami	Rau (2005), Rau and Dong (2006)	Victoria Rau
Yanyuwa	Bradley, Kirton and the Yanyuwa Community (1992), Kirton and Charlie (1996)	John Bradley
Yimas	Foley (1991)	William Foley
Yindjibarndi	Wordick (1982)	

APPENDIX III Questionnaire/list of recurrent questions

As explained in section 2.3.3., the lists of questions sent to the consultants were tailored for each of the languages individually. The set of the questions depended on the information missing from the database after published materials and manuscripts available for a given language have been analysed. The following presents the list of the recurrent questions. If there was uncertainty, specific follow-up questions were asked.

PART A: QUESTIONS CONCERNING CLAUSE LINKERS

PART A1

(for clause linkers described in the published materials/manuscripts for which some information needed for the database was missing)

General information

Is the X clause-linker:

a) original to the language borrowed

What is the source language?.....

b) synchronically monomorphemic synchronically polymorphemic

c) a word a clitic an affix a combination of word(s) and affix(es)
distributed marker

d) if it's an affix please specify

prefix

suffix

circumfix

other please explain more

e) (for synchronically polymorphemic markers only)

give morpheme-by-morpheme glosses and mark all the borrowed morphemes incorporated

.....

f) are there any restrictions of usage of the marker (e.g. only in same- or different subject clauses, only with verbs of motion etc.)?

.....

Other syntactic functions

Is the X clausal marker used in any other syntactic function (cf. English *after* serving the function of adposition, adverb and clause linker)?

yes no

If yes explain what these other functions are and provide an example for each of them.

.....

Other clause-linking functions

Is the X clausal marker used to convey any other circumstantial meanings between clauses (cf. English *since* used as a clause linker in temporal and causal clauses)?

yes no

If yes explain what these meanings are and provide an example for each of them.

.....

PART A2

On numerous occasions the following question has also been asked:

Apart from the X, Y, and Z clause linkers are there any other morphemes (or polymorphemic structures) used for encoding of the relation of anteriority/causality/purpose/conditionality?

yes no

For each of the new clause linkers the consultant has been asked to provide an example of use and answer the standard list of questions from PART A1.

PART B: QUESTIONS CONCERNING SOCIO-CULTURAL ISSUES

Level of written form development

Mark all the options that apply to language XXX.

- the language has never had a written form or it is not used by the speakers
- orthography and primers have been developed for the language
- the written form is used for personal communication between the speakers
- there are printed materials being published in the language
 - the printed materials were published already before 1801
 - the printed materials were published before 1901
 - the printed materials have been published only since 1901
- there are newspapers published in the language
 - the newspapers were published already before 1801
 - the newspapers were published before 1901
 - the newspapers have been published only since 1901
- there are literary works being published in the language
 - the literary works were published already before 1801
 - the literary works were published before 1901
 - the literary works have been published only since 1901
- the language in its written form is used by the state authorities in their official letters/acts/directives etc.
 - the language in its written form was used by the state authorities already before 1801

- the language in its written form was used by the state authorities before 1901
- the language in its written form has been used by the state authorities only since 1901

Remarks:

Presence of the language in school teaching

Mark the option that best describes the presence of the XXX language in school teaching

- no school teaching in the language;
- language taught only as a foreign language or as a language of instruction only in some schools;
- language present as a language of instruction in first (and possibly also other) grade;
- language fully present at all stages of education, including higher education.

Remarks:

Radio broadcasting in the language

Mark the option that best describes the presence of the XXX language in radio broadcasting

- no broadcasts;
- modest amount of broadcasting (occasional programs);
- broadcasting fully present (most of the time stations broadcast in the native language).

Remarks:

TV broadcasting in the language

Mark the option that best describes the presence of the XXX language in TV broadcasting

- no broadcasts;
- modest amount of broadcasting (occasional programs);
- broadcasting fully present (most of the time stations broadcast in the native language).

Remarks:

APPENDIX IV Degrees of grammaticalization (1)

Summary for the MM group (monomorphemic and semantically monofunctional c-glossems).

Sample size: 67

1 = presence of the marker, 0 = absence of the marker

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	MM M	MM 1	MM 2	MM M	MM 1	MM 2	MM M	MM 1	MM 2	MM M	MM 1	MM 2
Akan	0	0	0	1	1	1	0	0	1	0	1	1
Arabana	0	0	0	0	0	0	0	0	0	0	0	0
Arabic	0	0	1	0	0	0	0	0	0	1	1	1
Au	1	1	1	0	1	1	0	0	0	0	0	0
Basque	1	1	1	1	1	1	0	0	1	1	1	1
Baure	0	0	0	0	0	0	0	0	0	0	0	1
Boko	0	1	1	1	1	1	0	0	0	0	0	0
Burushaski	1	1	1	1	1	1	0	0	0	0	0	0
Chukchi	0	1	1	1	1	1	0	0	0	0	0	0
Cubeo	1	1	1	0	1	1	0	0	0	1	1	1
Dagur	1	1	1	0	0	0	0	0	0	1	1	1
Didinga	0	0	0	0	1	1	0	0	0	1	1	1
Eipo	1	1	1	0	0	0	0	0	0	0	0	0
English	0	0	1	1	1	1	0	0	1	0	1	1
Estonian	1	1	1	1	1	1	0	0	0	0	0	0
Galo	1	1	1	0	0	0	0	0	0	0	1	1
Gola	0	0	0	0	0	0	0	0	0	0	0	0
Hattam	0	0	0	0	0	1	0	0	0	0	1	1
Hausa	0	0	0	0	0	0	0	0	0	0	0	0
Hindi	0	1	1	0	1	1	1	1	1	1	1	1
Hualapai	0	0	0	0	0	0	0	0	0	1	1	1
Ika	0	0	0	1	1	1	1	1	1	1	1	1
Ilokano	0	0	0	1	1	1	1	1	1	0	1	1
I'saka	0	0	0	0	0	0	0	0	0	0	0	0
Jahai	0	0	0	0	0	0	0	0	0	0	0	0
Japanese	0	1	1	0	1	1	0	0	0	1	1	1
Jingulu	0	0	0	0	0	0	1	1	1	0	0	0
Kanuri	0	0	0	0	0	0	0	0	0	0	0	1
Kayah Li	0	0	0	0	0	0	0	0	0	1	1	1
Ket	0	0	0	0	1	1	0	1	1	0	0	0
Khwe	0	0	0	0	0	0	0	1	1	0	0	0
Konso	0	1	1	0	0	0	0	0	0	1	1	1
Krongo	0	1	1	1	1	1	0	1	1	0	0	0
Lango	0	0	0	0	0	0	0	0	0	1	1	1

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	MM M	MM 1	MM 2	MM M	MM 1	MM 2	MM M	MM 1	MM 2	MM M	MM 1	MM 2
Lavukaleve	1	1	1	1	1	1	0	1	1	0	0	0
Lepcha	0	0	0	0	0	0	0	0	1	1	1	1
Leti	0	0	0	1	1	1	0	0	0	0	0	0
Lezgian	0	1	1	1	1	1	0	0	0	1	1	1
Lillooet	0	0	0	0	0	0	0	0	0	0	0	0
Madi	0	0	0	1	1	1	0	0	0	1	1	1
Mandarin	1	1	1	1	1	1	1	1	1	1	1	1
Mantauran Rukai	1	1	1	1	1	1	0	1	1	1	1	1
Meyah	0	0	0	0	1	1	0	1	1	0	0	0
Ndyuka	0	0	0	1	1	1	0	0	0	0	0	0
Nez Perce	0	0	0	1	1	1	1	1	1	1	1	1
Nisga'a	0	1	1	0	0	0	0	0	0	0	0	0
Nivkh	1	1	1	0	1	1	1	1	1	0	0	1
Polish	0	1	1	1	1	1	0	1	1	1	1	1
Quechua	0	0	0	0	0	0	0	0	1	0	0	0
Rama	0	0	0	0	1	1	0	0	0	0	0	0
Retuara	0	0	0	1	1	1	0	1	1	0	1	1
S Tepehuan	0	0	0	0	0	0	0	0	0	1	1	1
Sango	0	0	0	0	0	0	0	0	0	0	0	0
Sapuan	0	0	0	0	0	0	0	0	0	0	0	0
Seediq	0	0	0	1	1	1	1	1	1	0	1	1
Seri	0	0	0	1	1	1	0	0	0	0	0	0
Swahili	0	0	0	0	0	0	0	0	0	1	1	1
Taba	0	0	0	0	0	0	0	1	1	0	0	0
Tamil	0	1	1	0	0	0	0	1	1	0	0	0
Thai	0	0	1	1	1	1	0	1	1	1	1	1
Vitu	0	0	0	1	1	1	0	0	0	1	1	1
Wambaya	0	1	1	0	0	0	0	1	1	0	0	0
Warlpiri	0	1	1	0	0	0	0	1	1	1	1	1
Yami	0	0	0	0	0	0	0	0	0	0	1	1
Yimas	1	1	1	0	0	0	0	0	0	0	0	0
Yindjibarndi	0	0	0	1	1	1	0	0	0	0	1	1
Yup'ik	0	1	1	1	1	1	1	1	1	0	0	0
TOTAL	13	26	29	26	35	36	9	22	27	24	33	36

APPENDIX V Degrees of grammaticalization (2)

Summary for the MM group (monomorphemic and semantically monofunctional c-glossemes) in same- and different-subject purpose clauses.

Sample size: 70

1 = presence of the marker, 0 = absence of the marker

Language	PURPOSE SAME-SUBJECT			PURPOSE DIFFERENT-SUBJECT		
	MMM	MM1	MM2	MMM	MM1	MM2
Achagua	0	1	1	0	1	1
Akan	0	0	1	0	0	1
Arabana	0	0	0	0	1	1
Arabic	0	0	0	0	0	0
Au	0	0	0	0	0	0
Basque	0	0	1	0	0	1
Baure	0	0	0	0	0	0
Boko	0	0	0	0	0	0
Burushaski	0	1	1	0	0	0
Chukchi	0	0	0	0	0	0
Cubeo	0	0	0	0	1	1
Dagur	0	0	0	0	0	0
Didinga	0	0	0	0	0	0
Eipo	0	0	0	0	0	0
English	0	0	1	0	0	1
Estonian	0	0	0	0	0	0
Galo	0	0	0	0	0	0
Gola	0	0	0	0	0	0
Hattam	0	0	0	0	0	0
Hausa	0	0	0	0	0	0
Hindi	1	1	1	1	1	1
Hualapai	0	0	0	0	0	0
Ika	1	1	1	1	1	1
Ilokano	1	1	1	1	1	1
I'saka	0	0	0	0	0	0
Jahai	0	0	0	0	0	0
Japanese	0	0	1	0	0	0
Jingulu	1	1	1	1	1	1
Kanuri	0	0	0	0	0	0
Kayah Li	0	0	0	0	0	0
Ket	0	1	1	0	1	1
Khwe	0	1	1	0	1	1
Konso	0	0	0	0	0	0

Language	PURPOSE SAME-SUBJECT			PURPOSE DIFFERENT-SUBJECT		
	MMM	MM1	MM2	MMM	MM1	MM2
Krongo	0	1	1	0	0	1
Lango	0	1	1	0	0	0
Lavukaleve	0	1	1	0	1	1
Lepcha	0	0	1	0	0	1
Leti	0	0	0	0	0	0
Lezgian	0	0	0	0	0	0
Lillooet	0	0	0	0	0	0
Ma'di	0	0	1	0	0	0
Mandarin	1	1	1	1	1	1
Mantauran Rukai	0	1	1	0	1	1
Meyah	0	1	1	0	1	1
Ndyuka	0	0	0	0	0	0
Nez Perce	1	1	1	1	1	1
Nisga'a	0	0	0	0	0	0
Nivkh	1	1	1	1	1	1
Pitjatjantjara	0	1	1	0	1	1
Polish	0	1	1	0	1	1
Quechua	0	0	1	0	0	1
Rama	0	0	0	0	0	0
Retuarã	0	1	1	0	1	1
SE Tepehuan	0	0	0	0	0	0
Sango	0	0	0	0	0	0
Sapuan	0	0	0	0	0	0
Seediq	1	1	1	1	1	1
Seri	0	0	0	0	0	0
Swahili	0	0	0	0	0	0
Taba	0	1	1	0	1	1
Tamil	0	1	1	0	1	1
Thai	0	1	1	0	1	1
Vitu	0	0	0	0	0	0
Wambaya	0	1	1	0	1	1
Warlpiri	0	1	1	0	1	1
Yami	0	0	0	0	0	0
Yanyuwa	0	1	1	0	1	1
Yimas	0	0	0	0	0	0
Yindjibarndi	0	0	0	0	0	0
Yup'ik	1	1	1	1	1	1
TOTAL	9	27	34	9	26	32

APPENDIX VI Degrees of grammaticalization (3)

Summary for the M1 group (monomorphemic c-glossems encoding up to one additional circumstantial meaning).

Sample size: 67

1 = presence of the marker, 0 = absence of the marker

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M1M	M11	M12	M1M	M11	M12	M1M	M11	M12	M1M	M11	M12
Akan	0	0	0	1	1	1	0	0	1	0	1	1
Arabana	0	0	0	0	0	0	0	0	0	0	0	0
Arabic	0	0	1	0	1	1	0	1	1	1	1	1
Au	1	1	1	0	1	1	0	0	0	0	0	0
Basque	1	1	1	1	1	1	0	0	1	1	1	1
Baure	0	0	0	0	0	0	0	0	0	0	0	1
Boko	0	1	1	1	1	1	0	0	0	1	1	1
Burushaski	1	1	1	1	1	1	0	0	0	0	0	0
Chukchi	1	1	1	1	1	1	0	0	0	0	1	1
Cubeo	1	1	1	0	1	1	0	0	0	1	1	1
Dagur	1	1	1	0	0	0	0	0	0	1	1	1
Didinga	0	0	0	0	1	1	0	0	0	1	1	1
Eipo	1	1	1	0	1	1	0	1	1	0	0	0
English	0	0	1	1	1	1	0	0	1	0	1	1
Estonian	1	1	1	1	1	1	0	0	0	0	0	0
Galo	1	1	1	0	0	0	0	1	1	0	1	1
Gola	0	0	0	0	0	0	0	0	0	1	1	1
Hattam	0	0	0	0	0	1	0	0	0	0	1	1
Hausa	0	0	0	0	1	1	0	1	1	1	1	1
Hindi	0	1	1	0	1	1	1	1	1	1	1	1
Hualapai	0	0	0	0	0	0	0	0	0	1	1	1
Ika	0	0	0	1	1	1	1	1	1	1	1	1
Ilokano	0	0	0	1	1	1	1	1	1	0	1	1
I'saka	0	0	0	0	0	0	0	0	0	0	0	0
Jahai	0	0	0	0	0	0	0	0	0	0	0	0
Japanese	0	1	1	0	1	1	0	0	0	1	1	1
Jingulu	0	0	0	0	0	0	1	1	1	1	1	1
Kanuri	0	0	0	0	0	0	0	0	0	0	0	1
Kayah Li	0	0	0	0	0	0	0	0	0	1	1	1
Ket	0	0	0	0	1	1	0	1	1	0	0	0
Khwe	0	0	0	0	0	0	0	1	1	0	0	0
Konso	0	1	1	0	1	1	0	1	1	1	1	1
Krongo	0	1	1	1	1	1	0	1	1	0	1	1
Lango	0	0	0	0	0	0	0	0	0	1	1	1

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M1M	M11	M12	M1M	M11	M12	M1M	M11	M12	M1M	M11	M12
Lavukaleve	1	1	1	1	1	1	0	1	1	0	1	1
Lepcha	0	1	1	0	1	1	0	0	1	1	1	1
Leti	0	0	0	1	1	1	0	0	0	0	0	0
Lezgian	0	1	1	1	1	1	0	0	0	1	1	1
Lillooet	0	0	0	0	0	0	0	0	0	1	1	1
Madi	0	0	0	1	1	1	0	0	0	1	1	1
Mandarin	1	1	1	1	1	1	1	1	1	1	1	1
Mantauran Rukai	1	1	1	1	1	1	0	1	1	1	1	1
Meyah	0	0	0	0	1	1	0	1	1	0	0	0
Ndyuka	0	0	1	1	1	1	0	0	0	0	0	0
Nez Perce	0	0	0	1	1	1	1	1	1	1	1	1
Nisga'a	0	1	1	1	1	1	0	0	0	0	0	0
Nivkh	1	1	1	0	1	1	1	1	1	0	0	1
Polish	0	1	1	1	1	1	0	1	1	1	1	1
Quechua	0	0	0	0	0	0	0	0	1	0	0	0
Rama	0	0	0	0	1	1	0	0	0	0	1	1
Retuara	0	0	0	1	1	1	0	1	1	0	1	1
S Tepehuan	0	0	0	0	0	0	0	0	0	1	1	1
Sango	0	0	0	0	0	0	0	0	0	0	0	0
Sapuan	0	0	0	0	0	0	0	0	0	0	0	0
Seediq	0	0	0	1	1	1	1	1	1	0	1	1
Seri	0	0	0	1	1	1	0	0	0	0	0	0
Swahili	0	0	0	0	0	0	0	0	0	1	1	1
Taba	0	0	0	0	0	0	0	1	1	0	1	1
Tamil	0	1	1	0	0	0	0	1	1	0	1	1
Thai	0	0	1	1	1	1	0	1	1	1	1	1
Vitu	0	0	0	1	1	1	0	0	0	1	1	1
Wambaya	0	1	1	0	0	0	0	1	1	0	0	0
Warlpiri	0	1	1	0	1	1	0	1	1	1	1	1
Yami	0	0	0	0	1	1	0	1	1	0	1	1
Yimas	1	1	1	0	0	0	0	0	0	0	0	0
Yindjibarndi	0	0	0	1	1	1	0	0	0	0	1	1
Yup'ik	0	1	1	1	1	1	1	1	1	1	1	1
TOTAL	14	27	31	27	43	44	9	28	33	30	45	48

APPENDIX VII Degrees of lexicalization (1)

Summary for the M(w)M group (monomorphemic, free word, semantically monofunctional c-glossemes).

Sample size: 67

1 = presence of the marker, 0 = absence of the marker

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2
Akan	0	0	0	1	1	1	0	0	1	0	1	1
Arabana	0	0	0	0	0	0	0	0	0	0	0	0
Arabic	0	0	1	0	0	0	0	0	0	1	1	1
Au	1	1	1	0	1	1	0	0	0	0	0	0
Basque	0	0	0	1	1	1	0	0	0	0	0	0
Baure	0	0	0	0	0	0	0	0	0	0	0	1
Boko	0	0	0	1	1	1	0	0	0	0	0	0
Burushaski	1	1	1	1	1	1	0	0	0	0	0	0
Chukchi	0	0	0	1	1	1	0	0	0	0	0	0
Cubeo	0	0	0	0	1	1	0	0	0	0	0	0
Dagur	0	0	0	0	0	0	0	0	0	0	0	0
Didinga	0	0	0	0	1	1	0	0	0	1	1	1
Eipo	0	0	0	0	0	0	0	0	0	0	0	0
English	0	0	1	1	1	1	0	0	1	0	1	1
Estonian	0	0	0	1	1	1	0	0	0	0	0	0
Galo	0	0	0	0	0	0	0	0	0	0	0	0
Gola	0	0	0	0	0	0	0	0	0	0	0	0
Hattam	0	0	0	0	0	1	0	0	0	0	1	1
Hausa	0	0	0	0	0	0	0	0	0	0	0	0

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2
Hindi	0	1	1	0	1	1	1	1	1	1	1	1
Hualapai	0	0	0	0	0	0	0	0	0	0	0	0
Ika	0	0	0	0	0	0	0	0	0	0	0	0
Ilokano	0	0	0	1	1	1	1	1	1	0	1	1
I'saka	0	0	0	0	0	0	0	0	0	0	0	0
Jahai	0	0	0	0	0	0	0	0	0	0	0	0
Japanese	0	1	1	0	1	1	0	0	0	0	0	0
Jingulu	0	0	0	0	0	0	1	1	1	0	0	0
Kanuri	0	0	0	0	0	0	0	0	0	0	0	1
Kayah Li	0	0	0	0	0	0	0	0	0	1	1	1
Ket	0	0	0	0	1	1	0	1	1	0	0	0
Khwe	0	0	0	0	0	0	0	0	0	0	0	0
Konso	0	0	0	0	0	0	0	0	0	1	1	1
Krongo	0	1	1	1	1	1	0	0	0	0	0	0
Lango	0	0	0	0	0	0	0	0	0	1	1	1
Lavukaleve	0	0	0	1	1	1	0	0	0	0	0	0
Lepcha	0	0	0	0	0	0	0	0	1	1	1	1
Leti	0	0	0	1	1	1	0	0	0	0	0	0
Lezgian	0	0	0	1	1	1	0	0	0	0	0	0
Lillooet	0	0	0	0	0	0	0	0	0	0	0	0
Madi	0	0	0	1	1	1	0	0	0	1	1	1
Mandarin	1	1	1	1	1	1	1	1	1	1	1	1
Mantauran	0	1	1	1	1	1	0	1	1	1	1	1
Rukai	0	0	0	0	1	1	0	1	1	0	0	0
Meyah	0	0	0	0	1	1	0	1	1	0	0	0
Ndyuka	0	0	0	1	1	1	0	0	0	0	0	0
Nez Perce	0	0	0	1	1	1	0	0	0	1	1	1
Nisga'a	0	1	1	0	0	0	0	0	0	0	0	0
Nivkh	0	0	0	0	0	0	0	0	0	0	0	0

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2	M(w)MM	M(w)M1	M(w)M2
Polish	0	1	1	1	1	1	0	1	1	1	1	1
Quechua	0	0	0	0	0	0	0	0	1	0	0	0
Rama	0	0	0	0	0	0	0	0	0	0	0	0
Retuara	0	0	0	0	0	0	0	0	0	0	1	1
S Tepehuan	0	0	0	0	0	0	0	0	0	1	1	1
Sango	0	0	0	0	0	0	0	0	0	0	0	0
Sapuan	0	0	0	0	0	0	0	0	0	0	0	0
Seediq	0	0	0	1	1	1	1	1	1	0	1	1
Seri	0	0	0	1	1	1	0	0	0	0	0	0
Swahili	0	0	0	0	0	0	0	0	0	0	0	0
Taba	0	0	0	0	0	0	0	1	1	0	0	0
Tamil	0	1	1	0	0	0	0	1	1	0	0	0
Thai	0	0	1	1	1	1	0	1	1	1	1	1
Vitu	0	0	0	1	1	1	0	0	0	1	1	1
Wambaya	0	0	0	0	0	0	0	0	0	0	0	0
Warlpiri	0	0	0	0	0	0	0	1	1	1	1	1
Yami	0	0	0	0	0	0	0	0	0	0	1	1
Yimas	0	0	0	0	0	0	0	0	0	0	0	0
Yindjibarndi	0	0	0	0	0	0	0	0	0	0	0	0
Yup'ik	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	3	10	13	22	29	30	5	13	17	16	23	25

Summary for the M(w)1 group (monomorphemic, free word c-glosseme encoding up to one additional circumstantial meaning).

Sample size: 67

1 = presence of the marker, 0 = absence of the marker

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12
Akan	0	0	0	1	1	1	0	0	1	0	1	1
Arabana	0	0	0	0	0	0	0	0	0	0	0	0
Arabic	0	0	1	0	1	1	0	1	1	1	1	1
Au	1	1	1	0	1	1	0	0	0	0	0	0
Basque	0	0	0	1	1	1	0	0	0	0	0	0
Baure	0	0	0	0	0	0	0	0	0	0	0	1
Boko	0	0	0	1	1	1	0	0	0	1	1	1
Burushaski	1	1	1	1	1	1	0	0	0	0	0	0
Chukchi	0	0	0	1	1	1	0	0	0	0	1	1
Cubeo	0	0	0	0	1	1	0	0	0	0	0	0
Dagur	0	0	0	0	0	0	0	0	0	0	0	0
Didinga	0	0	0	0	1	1	0	0	0	1	1	1
Eipo	0	0	0	0	1	1	0	1	1	0	0	0
English	0	0	1	1	1	1	0	0	1	0	1	1
Estonian	0	0	0	1	1	1	0	0	0	0	0	0
Galo	0	0	0	0	0	0	0	1	1	0	0	0
Gola	0	0	0	0	0	0	0	0	0	1	1	1
Hattam	0	0	0	0	0	1	0	0	0	0	1	1
Hausa	0	0	0	0	1	1	0	1	1	1	1	1

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12
Hindi	0	1	1	0	1	1	1	1	1	1	1	1
Hualapai	0	0	0	0	0	0	0	0	0	0	0	0
Ika	0	0	0	0	0	0	0	0	0	0	0	0
Ilokano	0	0	0	1	1	1	1	1	1	0	1	1
I'saka	0	0	0	0	0	0	0	0	0	0	0	0
Jahai	0	0	0	0	0	0	0	0	0	0	0	0
Japanese	0	1	1	0	1	1	0	0	0	0	0	0
Jingulu	0	0	0	0	0	0	1	1	1	1	1	1
Kanuri	0	0	0	0	0	0	0	0	0	0	0	1
Kayah Li	0	0	0	0	0	0	0	0	0	1	1	1
Ket	0	0	0	0	1	1	0	1	1	0	0	0
Khwe	0	0	0	0	0	0	0	0	0	0	0	0
Konso	0	0	0	0	1	1	0	1	1	1	1	1
Krongo	0	1	1	1	1	1	0	0	0	0	0	0
Lango	0	0	0	0	0	0	0	0	0	1	1	1
Lavukaleve	0	0	0	1	1	1	0	0	0	0	0	0
Lepcha	0	0	0	0	0	0	0	0	1	1	1	1
Leti	0	0	0	1	1	1	0	0	0	0	0	0
Lezgian	0	0	0	1	1	1	0	0	0	0	0	0
Lillooet	0	0	0	0	0	0	0	0	0	0	0	0
Madi	0	0	0	1	1	1	0	0	0	1	1	1
Mandarin	1	1	1	1	1	1	1	1	1	1	1	1
Mantauran	0	1	1	1	1	1	0	1	1	1	1	1
Rukai	0	0	0	0	1	1	0	1	1	0	0	0
Meyah	0	0	0	1	1	1	0	1	1	0	0	0
Ndyuka	0	0	0	1	1	1	0	0	0	0	0	0
Nez Perce	0	0	0	1	1	1	0	0	0	1	1	1
Nisga'a	0	1	1	1	1	1	0	0	0	0	0	0
Nivkh	0	0	0	0	0	0	0	0	0	0	0	0

Language	ANTERIORITY			CAUSALITY			PURPOSE			CONDITIONALITY		
	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12	M(w)1M	M(w)11	M(w)12
Polish	0	1	1	1	1	1	0	1	1	1	1	1
Quechua	0	0	0	0	0	0	0	0	1	0	0	0
Rama	0	0	0	0	1	1	0	0	0	0	0	0
Retuara	0	0	0	0	0	0	0	0	0	0	1	1
S Tepehuan	0	0	0	0	0	0	0	0	0	1	1	1
Sango	0	0	0	0	0	0	0	0	0	0	0	0
Sapuan	0	0	0	0	0	0	0	0	0	0	0	0
Seediq	0	0	0	1	1	1	1	1	1	0	1	1
Seri	0	0	0	1	1	1	0	0	0	0	0	0
Swahili	0	0	0	0	0	0	0	0	0	0	0	0
Taba	0	0	0	0	0	0	0	1	1	0	1	1
Tamil	0	1	1	0	0	0	0	1	1	0	0	0
Thai	0	0	1	1	1	1	0	1	1	1	1	1
Vitu	0	0	0	1	1	1	0	0	0	1	1	1
Wambaya	0	0	0	0	0	0	0	0	0	0	0	0
Waripiri	0	0	0	0	1	1	0	1	1	1	1	1
Yami	0	0	0	0	1	1	0	1	1	0	1	1
Yimas	0	0	0	0	0	0	0	0	0	0	0	0
Yindjibarndi	0	0	0	0	0	0	0	0	0	0	0	0
Yup'ik	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	3	10	13	23	37	38	5	19	23	20	29	31

APPENDIX IX Degrees of explicitness with borrowed c-glossemes included

The values assigned to the languages follow the customized hierarchy of explicitness of encoding circumstantial relations presented in (Fig.9.24.) in section 9.3.2

The languages in which loanwords contributed to the increase of explicitness have been put in italics.

	ANTERIORITY	CAUSALITY	CONDITIONALITY	PURPOSE SAME-SUBJECT	PURPOSE DIFFERENT-SUBJECT	MEAN
Achagua	4	4	4	1	1	2.8
Akan	1	1	1	0	0	x
Ama	3	0	3	0	0	x
Apache	6	3	3	3	0	x
Apurina	7	0	0	0	0	x
Arabana	6	6	7	1	1	4.2
Arabic	1	2	1	2	2	1.6
Au	1	1	4	4	0	x
Basque	1	1	1	1	1	1.0
Batak Karo	1	1	1	1	1	1.0
Baure	7	1	1	1	1	2.2
Boko	1	1	2	2	2	1.6
<i>Burushaski</i>	1	1	1	1	1	1.0
Chukchi	1	1	2	0	0	x
Cubeco	1	1	1	6	1	2.0
Dagur	1	3	1	0	0	x
Didinga	1	1	1	1	1	1.0
Eipo	1	2	3	2	2	2.0
English	1	1	1	1	1	1.0
Estonian	1	1	1	1	1	1.0

	ANTERIORITY	CAUSALITY	CONDITIONALITY	PURPOSE SAME-SUBJECT	PURPOSE DIFFERENT-SUBJECT	MEAN
Fur	2	1	2	1	1	x
Galo	1	2	1	2	2	1.6
Gola	4	1	2	1	0	x
Hattam	6	1	1	1	1	2.0
Hausa	1	2	2	2	2	1.8
Hindi	1	1	1	1	1	1.0
Hualapai	5	5	1	5	5	4.2
Ika	6	1	1	1	1	2.0
Ilokano	1	1	1	1	1	1.0
I'saka	6	7	7	7	7	6.8
<i>Jahai</i>	1	1	1	7	7	3.4
Japanese	1	1	1	1	1	1.0
Jingulu	7	6	2	1	1	x
Kanuri	2	1	1	4	4	x
Kayah Li	5	5	1	5	5	4.2
Ket	1	1	1	1	1	1.0
<i>Khwe</i>	1	1	1	1	1	1.0
Konso	1	2	1	1	1	1.2
Krongo	1	1	1	1	2	1.2
Kryjol	1	1	1	0	1	x
Lango	3	1	1	1	0	x
Lavukaleve	1	1	3	1	1	1.4
Lepcha	1	2	1	1	1	1.2
<i>Leti</i>	1	1	1	1	1	1.0
Lezgian	1	1	1	1	1	1.0
Lillooet	2	2	2	1	1	1.6
Maale	1	1	1	0	0	x
Macushi	0	0	2	0	0	x

	ANTERIORITY	CAUSALITY	CONDITIONALITY	PURPOSE SAME-SUBJECT	PURPOSE DIFFERENT-SUBJECT	MEAN
Ma'di	6	1	1	1	0	x
Mandarin	1	1	1	1	1	1.0
Mantauran Rukai	1	1	1	1	1	1.0
Meyah	6	1	2	1	1	2.2
Mayogo	1	1	1	1	0	x
Mocovi	6	1	1	0	0	x
Ndyuka	1	1	2	2	2	1.6
Nez Perce	1	1	1	1	1	1.0
Nisga'a	1	2	1	3	3	2.0
Nivkh	1	1	1	1	1	1.0
Paiwan	3	0	3	0	3	x
Pitjatantjara	2	2	5	1	1	2.2
Polish	1	1	1	1	1	1.0
Quechua Huallaga	4	4	5	1	1	3.0
Rama	2	1	2	1	1	1.4
Letuma	1	1	1	1	1	1.0
SE Tepehuan	5	1	1	1	1	1.8
Sango	6	1	1	2	2	2.4
Santali	2	1	1	1	1	1.4
Sapuan	7	1	1	1	1	2.2
Seediq	6	1	1	1	1	2.0
Seri	5	1	5	0	6	x
Shelha	1	1	0	0	0	x
Supyire	6	1	2	1	1	2.2
Swahili	1	1	1	1	1	1.0
Taba	6	1	2	1	1	2.2
Tamil	1	1	3	1	1	1.4
Thai	1	1	1	1	1	1.0

	ANTERIORITY	CAUSALITY	CONDITIONALITY	PURPOSE SAME-SUBJECT	PURPOSE DIFFERENT-SUBJECT	MEAN
Vitu	6	1	1	1	1	2.0
Wambaya	1	2	7	1	1	2.4
Warlpiri	1	2	1	1	1	1.2
Yami	6	1	1	1	1	2.0
Yanyuwa	0	0	3	1	1	x
Yimas	1	5	5	6	6	4.6
Yindjibarndi	7	1	1	7	6	4.4
Yup'ik	1	1	1	1	1	1.0
TOTAL	194	113	125	125	119	
Number of languages with values present	82	80	81		72	
Mean	2.37	1.41	1.54		1.74	

APPENDIX X Socio-cultural profiles of the sample languages

The table below shows values of the parameters discussed in section 10.2. for each of the language in the sample.

Language	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
Akan	1	3	2	2	7	4	2	0.61	0.58
Arabana	1	2	1	1	1	1	1	0.25	0.27
Arabic	1	2	2	2	7	4	2	0.57	0.53
Au	1	1	1	1	4	2	1	0.32	0.32
Basque	3	3	3	3	6	3	3	0.75	0.58
Baure	1	1	1	1	2	1	1	0.25	0.21
Boko	1	1	1	1	6	3	1	0.39	0.42
Burushaski	1	3	2	1	5	3	1	0.74	0.47
Chukchi	2	3	2	1	4	2	1	0.46	0.47
Cubeo	2	3	1	1	4	2	1	0.43	0.47
Dagur	2	1	2	2	6	3	1	0.5	0.47
Didinga	1	1	1	1	5	3	1	0.36	0.37
Eipo	1	1	1	1	3	2	1	0.29	0.26
English	6	4	3	3	9	5	3	1	1
Estonian	4	4	3	3	7	4	3	0.86	0.68
Galo	1	1	2	1	3	3	1	0.32	0.26
Gola	1	1	1	1	6	3	1	0.39	0.42
Hattam	1	1	1	1	5	3	1	0.36	0.37
Hausa	4	3	3	3	8	4	2	0.82	0.79
Hindi	4	4	3	3	9	5	2	0.89	0.74
Hualapai	1	2	1	1	3	2	2	0.36	0.32
Ika	1	3	1	1	5	3	1	0.43	0.47
Ilokano	2	2	2	2	7	4	2	0.61	0.58
I'saka	1	1	1	1	3	2	1	0.29	0.26

Language	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
Jahai	1	1	1	1	4	2	1	0.32	0.32
Japanese	5	4	3	3	9	5	3	0.96	0.95
Jingulu	1	1	1	1	2	1	1	0.25	0.21
Kanuri	4	3	2	2	7	4	1	0.68	0.74
Kayah Li	1	1	1	1	6	3	1	0.39	0.42
Ket	1	1	1	1	3	2	1	0.29	0.26
Khwe	1	1	1	1	4	2	1	0.32	0.32
Konso	1	1	1	1	6	3	1	0.39	0.42
Krongo	1	1	1	1	5	3	1	0.36	0.37
Lango	2	2	2	1	7	4	1	0.54	0.58
Lavukaleve	1	1	1	1	4	2	1	0.32	0.32
Lepcha	2	3	2	1	5	3	1	0.5	0.53
Leti	1	1	1	1	4	2	1	0.32	0.32
Lezgian	2	3	2	1	6	3	2	0.57	0.58
Lilloet	1	3	1	1	3	2	2	0.39	0.37
Ma'di	1	1	1	1	6	3	1	0.39	0.42
Mandarin	5	4	3	3	9	5	3	0.96	0.95
Mantauran									
Rukai	1	1	1	1	5	3	2	0.39	0.37
Mayah	1	2	1	1	5	3	1	0.39	0.42
Ndyuka	1	1	2	1	5	3	2	0.43	0.37
Nez Perce	1	1	1	1	3	2	2	0.32	0.26
Nisga'a	1	2	1	1	3	2	1	0.32	0.32
Nivkh	1	1	1	1	3	2	1	0.29	0.26
Polish	6	4	3	3	8	4	3	0.96	0.95
Quechua	1	1	1	1	5	3	1	0.36	0.37
Rama	1	1	1	1	2	1	1	0.25	0.21
Retuarã	1	3	1	1	3	2	1	0.36	0.37
Sango	2	3	2	2	6	3	2	0.61	0.58
Sapuan	1	1	1	1	3	2	1	0.29	0.26

Language	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
SE									
Tepehuan	1	2	2	1	5	3	1	0.43	0.42
Seediq	1	1	1	1	4	2	1	0.32	0.32
Seri	1	1	1	1	3	2	1	0.29	0.26
Swahili	4	3	3	3	6	3	2	0.75	0.68
Taba	1	1	1	1	5	3	1	0.36	0.37
Tamil	5	3	3	3	8	4	2	0.86	0.84
Thai	5	4	3	3	8	4	3	0.93	0.89
Vitu	1	2	1	1	4	2	1	0.36	0.37
Wambaya	1	1	1	1	1	1	1	0.21	0.16
Waripiri	1	2	2	1	4	2	1	0.39	0.37
Yami	1	1	2	1	4	2	1	0.36	0.32
Yimas	1	1	1	1	3	2	1	0.29	0.26
Yindjibarndi	1	1	1	1	3	2	1	0.29	0.26
Yup'ik	2	3	3	2	5	3	2	0.61	0.53

APPENDIX XI Correlations between socio-cultural factors and degree of grammaticalization

For each of the degrees of grammaticalization (MMM, M1M etc.) the higher number refers to the value of Spearman's rho and the lower one to the statistical significance. Those correlations which are statistically significant have been put in bold.

Relation	Degree of gram.	Socio-cultural factors									
		Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2	
ANTERIORITY	MMM	.136	.028	.113	.122	.014	.052	.092	.072	.019	
	MM1	.271	.823	.364	.325	.908	.677	.460	.565	.878	
		.313	.209	.285	.221	.164	.157	.188	.236	.200	
		.010	.090	.019	.073	.184	.204	.127	.055	.106	
	MM2	.346	.267	.355	.313	.241	.236	.279	.303	.266	
		.004	.029	.003	.010	.049	.055	.022	.013	.029	
	M1M	.212	.155	.159	.137	.008	.033	.167	.145	.086	
		.085	.211	.198	.269	.948	.793	.176	.241	.491	
	M11	.323	.269	.282	.182	.133	.141	.201	.25	.215	
		.008	.028	.021	.139	.282	.256	.102	.037	.081	
	M12	.336	.299	.374	.257	.215	.232	.327	.336	.275	
		.005	.014	.002	.036	.080	.059	.007	.005	.024	
	M2M	.212	.155	.159	.137	.008	.033	.167	.145	.086	
		.085	.211	.198	.269	.948	.793	.176	.241	.491	
	M21	.302	.241	.304	.164	.138	.153	.236	.268	.208	
		.013	.050	.012	.185	.267	.218	.054	.028	.092	
	M22	.336	.299	.374	.257	.215	.232	.327	.336	.275	
		.005	.014	.002	.036	.080	.059	.007	.005	.024	

	Degree of gram.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
CAUSALITY	MMM	.137	.219	.164	.147	.209	.187	.325	.266	.236
		.269	.075	.186	.236	.089	.131	.007	.030	.055
		.150	.216	.085	.129	.170	.151	.265	.208	.190
	MM1	.227	.079	.496	.297	.170	.222	.030	.091	.123
		.130	.189	.060	.112	.175	.164	.245	.198	.185
		.294	.126	.628	.369	.157	.186	.046	.109	.135
	M1M	.116	.227	.138	.128	.174	.158	.302	.239	.212
		.350	.065	.266	.303	.160	.202	.013	.052	.085
		.115	.232	.160	.120	.187	.154	.222	.232	.207
	M11	.355	.059	.196	.333	.129	.213	.071	.059	.092
		.095	.205	.136	.103	.194	.168	.202	.223	.203
		.444	.096	.273	.409	.115	.173	.101	.070	.100
M2M	.116	.227	.138	.128	.174	.158	.302	.239	.212	
	.350	.065	.266	.303	.160	.202	.013	.052	.085	
	.115	.232	.160	.120	.187	.154	.222	.232	.207	
M21	.355	.059	.196	.333	.129	.213	.071	.059	.092	
	.095	.205	.136	.103	.194	.168	.202	.223	.203	
	.444	.096	.273	.409	.115	.173	.101	.070	.100	
M22	.181	.156	.131	.215	.090	.122	.241	.110	.089	
	.143	.207	.291	.081	.469	.326	.050	.376	.472	
	.109	.104	.043	.149	.054	.074	.179	.053	.047	
MMM	.380	.403	.729	.230	.662	.550	.148	.670	.708	
	.140	.167	.114	.203	.121	.150	.224	.126	.124	
	.259	.176	.359	.100	.329	.226	.068	.308	.316	
PURPOSE	MMM	.137	.219	.164	.147	.209	.187	.325	.266	.236
		.269	.075	.186	.236	.089	.131	.007	.030	.055
		.150	.216	.085	.129	.170	.151	.265	.208	.190
	MM1	.227	.079	.496	.297	.170	.222	.030	.091	.123
		.130	.189	.060	.112	.175	.164	.245	.198	.185
		.294	.126	.628	.369	.157	.186	.046	.109	.135
	M1M	.116	.227	.138	.128	.174	.158	.302	.239	.212
		.350	.065	.266	.303	.160	.202	.013	.052	.085
		.115	.232	.160	.120	.187	.154	.222	.232	.207
	M11	.355	.059	.196	.333	.129	.213	.071	.059	.092
		.095	.205	.136	.103	.194	.168	.202	.223	.203
		.444	.096	.273	.409	.115	.173	.101	.070	.100
M2M	.116	.227	.138	.128	.174	.158	.302	.239	.212	
	.350	.065	.266	.303	.160	.202	.013	.052	.085	
	.115	.232	.160	.120	.187	.154	.222	.232	.207	
M21	.355	.059	.196	.333	.129	.213	.071	.059	.092	
	.095	.205	.136	.103	.194	.168	.202	.223	.203	
	.444	.096	.273	.409	.115	.173	.101	.070	.100	
M22	.181	.156	.131	.215	.090	.122	.241	.110	.089	
	.143	.207	.291	.081	.469	.326	.050	.376	.472	
	.109	.104	.043	.149	.054	.074	.179	.053	.047	
MMM	.380	.403	.729	.230	.662	.550	.148	.670	.708	
	.140	.167	.114	.203	.121	.150	.224	.126	.124	
	.259	.176	.359	.100	.329	.226	.068	.308	.316	

Degree of gram.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
M1M	.181	.156	.131	.215	.090	.122	.241	.110	.089
	.143	.207	.291	.081	.469	.326	.050	.376	.472
M11	.052	.029	.112	.164	.072	.122	.160	.043	.020
	.675	.818	.368	.184	.564	.325	.196	.729	.875
M12	.086	.094	.181	.221	.138	.197	.209	.116	.097
	.491	.451	.142	.073	.264	.110	.090	.348	.434
M2M	.181	.156	.131	.215	.090	.122	.241	.110	.089
	.143	.207	.291	.081	.469	.326	.050	.376	.472
M21	.052	.029	.112	.164	.072	.122	.160	.043	.020
	.675	.818	.368	.184	.564	.325	.196	.729	.875
M22	.086	.094	.181	.221	.138	.197	.209	.116	.097
	.491	.451	.142	.073	.264	.110	.090	.348	.434
MMM	.352	.334	.267	.263	.412	.339	.352	.446	.441
	.003	.006	.029	.031	.001	.005	.003	.000	.000
MM1	.293	.304	.301	.276	.390	.373	.341	.416	.403
	.016	.012	.013	.024	.001	.002	.005	.000	.001
MM2	.301	.280	.276	.279	.346	.336	.281	.364	.357
	.013	.022	.024	.022	.004	.005	.021	.002	.003
M1M	.339	.332	.256	.276	.410	.326	.384	.459	.452
	.005	.006	.037	.024	.001	.007	.001	.000	.000
M11	.277	.273	.260	.236	.366	.333	.317	.410	.403
	.023	.025	.034	.054	.002	.006	.009	.001	.001
M12	.294	.256	.242	.246	.329	.304	.261	.367	.366
	.016	.037	.049	.045	.006	.013	.033	.002	.002

CONDITIONALITY

Degree of gram.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(3)	Type of society	Indicator 1	Indicator 2
M2M	.319	.304	.231	.257	.375	.298	.363	.420	.413
	.009	.012	.060	.036	.002	.014	.003	.000	.001
M21	.320	.282	.291	.281	.384	.353	.340	.422	.417
	.008	.021	.017	.021	.001	.003	.005	.000	.000
M22	.326	.307	.306	.280	.361	.346	.267	.426	.408
	.007	.012	.012	.022	.003	.004	.029	.000	.001

APPENDIX XII Correlations between socio-cultural factors and degree of lexicalization

For each of the degrees of lexicalization (M(w)MM, M(w)1M etc.) the higher number refers to the value of Spearman's rho and the lower one to the statistical significance. Those correlations which are statistically significant have been put in bold.

Relation	Degree of lexical.	Socio-cultural factors										Indicator 1	Indicator 2
		Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society					
ANTERIORITY	M(w)MM	.039	.129	.114	.048	.096	.088	.036	.139	.105			
		.756	.297	.359	.699	.438	.477	.772	.263	.397			
	M(w)M1	.247	.265	.232	.265	.281	.303	.270	.288	.265			
		.044	.030	.058	.030	.021	.013	.027	.018	.030			
	M(w)M2	.345	.378	.360	.427	.414	.438	.429	.412	.390			
		.004	.002	.003	.000	.000	.000	.000	.001	.001			
M(w)1M		.039	.129	.114	.048	.096	.088	.036	.139	.105			
		.756	.297	.359	.699	.438	.477	.772	.263	.397			
	M(w)11	.247	.265	.232	.265	.281	.303	.270	.288	.265			
		.044	.030	.058	.030	.021	.013	.027	.018	.030			
	M(w)12	.345	.378	.360	.427	.414	.438	.429	.412	.390			
		.004	.002	.003	.000	.000	.000	.000	.001	.001			
CAUSALITY	M(w)MM	.171	.158	.197	.169	.279	.225	.365	.283	.249			
		.166	.203	.110	.171	.022	.068	.002	.020	.042			
	M(w)M1	.218	.208	.160	.182	.311	.260	.335	.303	.283			
		.076	.092	.196	.139	.010	.034	.006	.013	.020			
	M(w)M2	.198	.179	.135	.164	.314	.271	.313	.291	.276			
		.109	.147	.276	.185	.010	.027	.010	.017	.024			

Degree of lexical.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2
M(w)1M	.149	.166	.169	.149	.240	.194	.340	.254	.223
	.230	.181	.172	.230	.050	.116	.005	.038	.069
	.126	.155	.181	.168	.268	.193	.281	.250	.219
M(w)11	.311	.211	.144	.175	.028	.118	.021	.042	.075
	.106	.127	.157	.150	.274	.206	.262	.239	.213
M(w)12	.392	.304	.206	.224	.025	.095	.032	.051	.083
	.181	.103	.134	.216	.138	.124	.163	.091	.092
M(w)MM	.143	.405	.278	.080	.264	.319	.187	.462	.461
	.213	.153	.177	.241	.212	.211	.236	.180	.172
M(w)M1	.084	.215	.151	.049	.084	.086	.055	.144	.164
	.255	.236	.242	.285	.301	.325	.272	.266	.270
M(w)M2	.037	.054	.048	.019	.013	.007	.026	.030	.027
	.181	.103	.134	.216	.138	.124	.163	.091	.092
M(w)1M	.143	.405	.278	.080	.264	.319	.187	.462	.461
	.129	.057	.134	.216	.138	.241	.194	.150	.123
M(w)11	.297	.649	.278	.080	.264	.049	.115	.227	.321
	.179	.143	.296	.285	.296	.350	.238	.235	.221
M(w)12	.147	.249	.015	.020	.015	.004	.053	.055	.073
	.156	.167	.171	.153	.364	.329	.225	.304	.304
M(w)MM	.210	.179	.171	.221	.003	.007	.069	.013	.013
	.139	.181	.194	.188	.379	.346	.238	.314	.321
M(w)M1	.263	.142	.116	.127	.002	.004	.053	.010	.008
	.164	.182	.189	.205	.360	.329	.191	.293	.306
M(w)M2	.184	.141	.125	.095	.003	.007	.122	.016	.012

Degree of lexical.	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	Indicator 1	Indicator 2
M(w)1M	.137	.090	.124	.132	.371	.306	.169	.268	.282
	.270	.471	.318	.285	.002	.012	.171	.028	.021
M(w)11	.133	.125	.159	.146	.396	.331	.160	.305	.328
	.284	.314	.198	.237	.001	.006	.196	.012	.007
M(w)12	.159	.127	.157	.165	.382	.318	.118	.289	.317
	.198	.304	.203	.181	.001	.009	.342	.018	.009

APPENDIX XIII Correlations between socio-cultural factors and degree of explicitness

For each of the relations the higher number refers to the value of Spearman's rho and the lower one to the statistical significance. Those correlations which are statistically significant have been put in bold.

Relation	SOCIO-CULTURAL FACTORS										Indicator 1	Indicator 2
	Level of written form development	Presence in school teaching	Presence of radio broadcasts	Presence of TV broadcasts	NoS(9)	NoS(5)	Type of society	NoS(9)	NoS(5)	Type of society		
ANTERIORITY (79 languages)	.354	.239	.366	.362	.294	.318	.364	.294	.318	.364	.356	.312
	.001	.037	.001	.001	.009	.004	.001	.001	.004	.001	.001	.005
CAUSALITY (77 languages)	.206	.104	.159	.174	.333	.277	.165	.165	.294	.165	.294	.292
	.075	.375	.169	.132	.003	.015	.154	.154	.015	.154	.010	.011
CONDITIONALITY (78 languages)	.297	.210	.296	.306	.370	.372	.295	.295	.372	.295	.361	.336
	.008	.069	.009	.007	.001	.001	.009	.009	.001	.009	.001	.003
PURPOSE same-subject (70 languages)	.105	.129	.097	.105	.132	.084	.096	.096	.084	.096	.141	.154
	.389	.291	.423	.385	.276	.491	.428	.428	.491	.428	.246	.203
PURPOSE different-subject (65 languages)	.189	.220	.076	.131	.094	.000	.147	.147	.000	.147	.154	.172
	.133	.081	.545	.299	.458	.997	.242	.242	.997	.242	.220	.170