

A quantitative and typological study of
Early Slavic participle clauses
and their competition



Nilo Pedrazzini
St Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2023

egda jug̃ dušq̃št̃ glete ěko

var̃ b̃q̃det̃.

i byvaat̃.

‘And when the south wind blows you say

“it’s going to be hot”.

And it is.’

(Luke 12:55)

Kua roa e takahia ana tēnei huarahi / This path has been trodden for a long time

Atu i konei, ki hea atu rā? / Where to from here?

(Lorde, *Mata Kohore*)

Acknowledgements

Jeje, no amount of thank-yous can possibly convey how shockingly lucky I am to have you. But here we go. Thank you for putting up with me when we were confined multiple times in lockdown in our tiny flat, only you and I, and our two gigantic cats, as I was writing away chapters and working six or seven jobs on the side. Thank you for letting me talk non-stop as we go for our walks, even though these are meant for you to have a relaxing break. Thank you for being so patient, for bearing with my ever-changing mood and both my flimsy and obsessive motivations, habits, hobbies, tastes and routines.

My deepest gratitude goes to Hanne Eckhoff, for being the kindest and most encouraging supervisor I could ever have asked for. I have learnt so much from the way you work, and your optimism was exactly what I needed to counteract my relentless catastrophism, to push on and finish this thing that is now for the (extremely tiny, but extremely intimidating) world to read.

I'm immensely grateful to Dag Haug, for letting me come to Oslo for two months exactly three years ago now. That started an extremely cool collaboration, without which this thesis would look very, very different.

Thanks to the Economic and Social Research Council for fully funding my DPhil well into the end of my second year (grant number ES/P000649/1), when I got the wonderful opportunity of a full-time (pre-doc) post-doc for the Living with Machines project at The Alan Turing Institute.

My thanks also go to the several people who have helped me (intentionally or unintentionally) get to this point. Mary MacRoberts, for the crucial feedback on earlier drafts of the first two Chapters. Ida Tóth, na svim savetima na samom početku mog DPhil-a i na zabavnim language exchanges preko Zoom-a, što me je držalo pri zdravoj pameti tokom pandemije. Carlo Dalle Ceste, for introducing me to PROIEL over ten years ago and for those precious early tips which hugely influenced the choices I've made in my academic wanderings.

Thanks to the Living with Machines team at the Turing Institute, particularly Federico Nanni, for being the best manager and mentor, and to Ruth Ahnert, Barbara McGillivray, Kalle Westerling, Daniel Wilson, Kaspar Beelen, Mariona Coll Ardanuy, Katie McDonough, Giorgia Tolfo for the trust, the discussions, and one of the most defining years of this journey. I said no to more beers than I would have wanted to in

order to write this up, but that won't happen again.

To Maria Cristina Bragone, for being a constant for the last twelve years, regularly checking up on me, mentoring me, and making sure that I don't forget I'm a Slavonic philologist, first.

Ai miei nonni: grazie di cuore per l'entusiasmo mostrato ad ogni (pessimo) tentativo di spiegarvi cosa faccio. Come vedete, è un lavoro che richiede di scrivere tanto. Non credo che avrò mai la forza di tradurre tutto in italiano, ma queste tre righe sono solo per voi. Vi voglio bene.

Oxford,
October 7, 2023

Abstract

This thesis investigates the semantic and pragmatic properties of Early Slavic participle constructions (conjunct participles and dative absolutes) to understand the principles motivating their selection over one another and over their main finite competitor (*jegda*-clauses). The issue is tackled by adopting two broadly different approaches, which inform the division of the thesis into two parts.

The first part of the thesis uses detailed linguistic annotation on Early Slavic corpora at the morphosyntactic, dependency, information-structural, and lexical levels to obtain indirect evidence for different potential functions of participle clauses and their main finite competitor. The goal of this part of the thesis is to understand the roles of compositionality and default discourse reasoning as explanations for the distribution of participle constructions and *jegda*-clauses in the Early Slavic corpus. The investigation shows that the competition between conjunct participles, absolute constructions, and *jegda*-clauses occurs at the level of discourse organization, where the main determining factor in their distribution is the distinction between *background* and *foreground* content of an (elementary or complex) discourse unit. The analysis also shows that the major common denominator between the three constructions is that all of them can function as frame-setting devices (i.e. background clauses), albeit to very different extents. In fact, conjunct participles are more typically associated with the foreground constituent of a discourse unit, whereas dative absolutes and *jegda*-clauses are typically associated with the background content.

The second part of the thesis uses massively parallel data, including Old Church Slavonic and Ancient Greek, and analyses typological variation in how languages express the semantic space of English *when*, whose scope encompasses that of Early Slavic participle constructions and *jegda*-clauses. To do so, probabilistic semantic maps are generated and statistical methods (including Kriging, Gaussian Mixture Modelling, precision and recall analysis) are used to induce cross-linguistically salient dimensions from the parallel corpus and to study conceptual variation within the semantic space of the hypothetical concept WHEN. Clear typological correspondences and differences with Early Slavic from linguistic phenomena in other languages are then exploited to corroborate and refine observations made on the core semantic-pragmatic properties of participle constructions and *jegda*-clauses on the basis of annotated Early Slavic data.

The analysis shows that ‘null’ constructions (juxtaposed clauses such as participles and converbs, or independent clauses) consistently cluster in particular regions of the semantic map cross-linguistically, which clearly indicates that participle clauses are not equally viable as alternatives to any use of WHEN, but carry particular meanings that make them less suitable for some of its functions. The investigation helped identify genealogically and areally unrelated languages that seem typologically very similar to Old Church Slavonic in the way they divide the semantic space of WHEN between overtly subordinated and ‘null’ constructions. Comparison with these languages reveals great similarities between the functions of Early Slavic participle constructions and of linguistic phenomena in some of these languages (particularly clause chaining, bridging, insubordination, and switch reference). Crucially, new clear correspondences are found between these phenomena and ‘non-canonical’ usages of participle constructions (i.e. coreferential dative absolutes, syntactically independent absolutes and conjunct participles, and participle constructions with no apparent matrix clause), which had often been written off as ‘aberrations’ by previous literature on Early Slavic.

Contents

Acknowledgements	v
List of Figures	xiii
List of Tables	xix
List of Abbreviations	xxiii
Introduction	1
0.1 The problem	1
0.2 Background	4
0.2.1 Defining conjunct participles and dative absolutes	4
0.2.2 Conjunct participles according to Bary & Haug (2011)	8
0.2.3 Frame setters and fronted adjuncts	11
0.2.4 Segmented Discourse Representation Theory	14
0.3 What’s with <i>when</i> ?	16
0.4 A token-based typological approach to the study of language-internal variation	21
0.5 The Early Slavic corpus	22
0.5.1 Identification of the constructions in the corpora	24
0.6 Thesis overview	30
I The constructions in the corpus	35
1 Participle constructions in deeply-annotated treebanks	37
1.1 Introduction	37
1.2 Position in the sentence and aspect distribution	38
1.2.1 A remark about <i>bysto</i> -clauses	43
1.2.2 Summary	46
1.3 Subjects	47
1.3.1 Realization and position	49
1.3.2 Parts of speech	53

1.3.3	Summary	60
1.4	Lexical variation among participles	61
1.5	Greek-Old Church Slavonic mismatches	67
1.5.1	Conjunct participles	67
1.5.2	Absolute constructions	68
1.6	Remarks on the discourse-relation annotation on the Gospel of Luke . . .	75
1.7	Summary	81
2	Participle constructions in standard treebanks	85
2.1	Introduction	85
2.2	Dative absolutes and conjunct participles in standard treebanks	85
2.2.1	Position and aspect distribution	86
2.2.2	Properties of subjects	97
2.2.3	Lexical variation among participles	107
2.3	Conjunct participles and dative absolutes in strategically-annotated texts	120
2.3.1	Position, aspect and lexical variation of participles	120
2.3.2	Zooming in: dative absolutes and conjunct participles in a Middle Bulgarian text	124
2.4	Summary	132
3	Finite competitors: <i>jegda</i>-clauses	135
3.1	<i>Jegda</i> -clauses as finite competitors	135
3.2	<i>Jegda</i> -clauses in the corpus	141
3.2.1	Position in the sentence and tense-aspect distribution	141
3.2.2	Properties of subjects	143
3.2.3	Aktionsart and lexical variation	151
3.2.4	Clause length	154
3.2.5	Summary	155
4	Temporal relations and discourse representation	159
4.1	<i>Jegda</i> -clauses and temporal relations interpretation	159
4.1.1	Summary	169
4.2	Rhetorical relations and the discourse structure of <i>jegda</i> - and participle clauses	170
4.2.1	Competing motivations, competing relations	170
4.2.2	INDEPENDENT RHEMES as triggers for Strong Narration	173
4.2.3	Formalizing FRAMES as Background-triggers	175
4.3	Summary	180

II	The constructions in cross-linguistic perspective	183
5	The semantic map of WHEN and its typological parallels	185
5.1	Massively parallel Bible corpora: beyond ancient Indo-European translations	185
5.2	Data	186
5.3	Methods	190
5.4	Analysis	198
5.4.1	Overtly-subordinated versus ‘null’ constructions	202
5.4.2	Classification	207
5.4.3	Overall precision and recall for the GMM clusters	216
5.5	Summary	224
6	Early Slavic participle clauses and their typological parallels	227
6.1	Participle constructions in the Old Church Slavonic WHEN-map	227
6.2	Lexified counterparts to Early Slavic participle constructions in the world’s languages	232
6.3	Null constructions in a cross-linguistics perspective	238
6.3.1	Clause chaining	242
6.3.2	Bridging	246
6.3.3	Switch reference	249
6.3.4	Insubordination	260
6.4	Summary	263
7	Conclusions	267
7.1	Overview of the main findings	267
7.2	Where to from here?	273
Appendices		
A	Massively parallel dataset: corpus representativeness	279
A.1	Language families	279
B	Language classification by pattern of co-expression	285
B.1	Pattern and subpattern assignment	285
	Bibliography	329

List of Figures

1	Discourse representation structure of a complex discourse from Asher & Lascarides (2003: 9)	15
2	Slash notation (secondary dependencies) in PROIEL/TOROT	25
3	PROIEL analysis versus linearization analysis of an XSUB	26
4	Dependency tree with a dative absolute and its subject	27
5	Dependency tree containing two coordinated conjunct participles	28
6	Dependency tree containing a dative absolute coordinated to the main clause	29
7	Dependency tree containing a <i>jegda</i> -clause.	30
8	Dependency tree containing two coordinated <i>jegda</i> -clauses	30
1.1	Average anaphoric distance between the subject and immediate antecedent in sentence-initial SV and VS conjunct participle	51
1.2	Average anaphoric distance between the subject and immediate antecedent in sentence-initial dative absolutes and conjunct participle (null and overt)	52
1.3	Average anaphoric distance between the subject and immediate antecedent in overt-subject and null-subject conjunct participles in sentence-initial position	53
1.4	Average anaphoric distance between the subject and immediate antecedent in sentence-initial dative absolutes and conjunct participle (overt subjects)	54
1.5	Pick-up rates of subject referents in the previous discourse: sentence-initial dative absolutes and conjunct participles (overt and null subjects)	55
1.6	Pick-up rates of subject referents in the previous discourse: sentence-initial dative absolutes and conjunct participles (overt subjects)	56
1.7	Average anaphoric distance between subject and immediate antecedent in sentence-initial dative absolutes and VS conjunct participle (overt subjects only)	60
2.1	Lexical variation among subject lemmas in sentence-initial dative absolutes and conjunct participles	103
2.2	Lexical variation among subject lemmas in sentence-initial dative absolutes in the SV and VS configuration	103

2.3	Lexical variation among subject lemmas in sentence-initial conjunct participles in the SV and VS configuration	104
2.4	Lexical variation among participles, any configuration	109
2.5	Lexical variation among participles in sentence-initial conjunct participles (SV versus VS)	110
2.6	Lexical variation among participles in sentence-initial dative absolutes (SV versus VS)	110
3.1	Average anaphoric distance between the subject and immediate antecedent in pre-matrix non-coreferential/coreferential <i>jegda</i> -clauses and pre-matrix dative absolutes/conjunct participles	145
3.2	Average anaphoric distance between subject and immediate antecedent in sentence-initial <i>jegda</i> -clauses, conjunct participles and dative absolutes (all subjects)	146
3.3	Average topic score of subjects in (the leftmost) pre-matrix conjunct participles and pre-matrix coreferential <i>jegda</i> -clauses, by subject realization	149
3.4	Average topic score of subjects in pre-matrix dative absolutes and pre-matrix non-coreferential <i>jegda</i> -clauses, by subject realization	150
3.5	Lexical variation among verbs in <i>jegda</i> -clauses, conjunct participles and dative absolutes	153
4.1	Possible orders of a non-durative <i>when</i> -clause and its non-durative matrix clause according to Hinrichs (1986: 78)	161
4.2	Basic discourse structure of a discourse unit involving a Background _{forward} relation in SDRT according to Asher, Prévot & Vieu (2007: 14)	176
4.3	Discourse representation structure for a complex discourse unit in Old Church Slavonic involving several FRAMES contributing to a common Framing Topic for a foreground constituent	179
4.4	Discourse representation structure for a complex discourse unit involving an implicit Framing Topic	180
5.1	Approximate areal distribution of the languages in the massively parallel dataset among the world's languages	189
5.2	Basic semantic map of WHEN resulting from plotting the first two dimensions of the MDS matrix	191
5.3	Kriging map for Doyayo	192
5.4	Clusters identified via Gaussian Mixture Modelling and <i>k</i> -means initialization, based on the first two dimensions of the MDS matrix: comparison between 3, 4, 5 and 6 clusters.	193
5.5	Evaluation of the optimal number of clusters for a GMM model of the two dimensional MDS matrix (Silhouette, AIC and BIC)	194

5.6	Gaussian Mixture Model of the two-dimensional MDS matrix (6 clusters)	194
5.7	30-nearest-neighbour search using the balltree method, with an example of its application to Doyayo (Atlantic-Congo, Africa)	196
5.8a	Kriging maps for Southern Nambikuára (Nambiquaran, South America)	198
5.8b	Kriging maps for Southern Altai (Turkic, Eurasia)	198
5.9	Heatmap of languages expressing each observation in the semantic map with a null construction	202
5.10a	Kriging map for Serbian (Indo-European, Eurasia)	204
5.10b	Kriging map for Moose Cree (Algic, North America)	204
5.11a	Kriging map for Romanian (Indo-European, Eurasia)	204
5.11b	Kriging map for Lahu (Sino-Tibetan, Eurasia)	204
5.12	High-competition area in the semantic map	205
5.13a	Kriging map for Banggai (Austronesian, Papunesia)	205
5.13b	Kriging map for Gen (Atlantic-Congo, Africa)	205
5.14	Three-dimensional unlabelled MDS map from the perspective of different axes	206
5.15	Potential three-way split of non-null situations.	207
5.16	Potential four-way split of non-null situations.	207
5.17	Approximate null area across all target languages	208
5.18a	Kriging map for Hawaiian (Austronesian, Papunesia)	210
5.18b	Kriging map for Djambarrpuyngu (Pama-Nyungan, Australia)	210
5.18c	Kriging map for Poqomchi' (Mayan, North America)	210
5.18d	Kriging map for Somali (Afro-Asiatic, Africa)	210
5.19a	Kriging map for Bete-Bendi (Atlantic-Congo, Africa)	211
5.19b	Kriging map for Sepik Iwam (Sepik, Papunesia)	211
5.20a	Kriging map for Mi'kmaq (Algic, North America)	211
5.20b	Kriging map for Italian (Indo-European, Eurasia)	211
5.21a	Kriging map for Nobonob (Nuclear Trans New Guinea, Papunesia)	212
5.21b	Kriging map for Eastern Oromo (Afro-Asiatic, Africa)	212
5.22a	Kriging map for Tuwuli (Atlantic-Congo, Africa)	212
5.22b	Kriging map for Kako (Atlantic-Congo, Africa)	212
5.23a	Kriging map for Ancient Greek (Indo-European, Eurasia)	213
5.23b	Kriging map for Wandala (Afro-Asiatic, Africa)	213
5.23c	Kriging map for Chumburung (Atlantic-Congo, Africa)	213
5.23d	Kriging map for Tii (Austronesian, Papunesia)	213
5.24a	Kriging map for Old Church Slavonic (Indo-European, Eurasia)	214
5.24b	Kriging map for Eigham (Atlantic-Congo, Africa)	214
5.24c	Kriging map for Xavánte (Nuclear-Macro-Je, South America)	214
5.24d	Kriging map for Goan Konkani (Indo-European, Eurasia)	214

5.25a	Kriging map for Lomeriano-Ignaciano Chiquitano (Chiquitano, South America)	215
5.25b	Kriging map for Kiribati (Austronesian, Papunesia)	215
5.25c	Kriging map for Afrikaans (Indo-European, Africa)	215
5.25d	Kriging map for Siwu (Atlantic-Congo, Africa)	215
5.26	Precision and recall plots for the six GMM clusters	217
5.27a	Kriging map for Pular (Atlantic-Congo, Africa)	218
5.27b	Kriging map for Sierra de Juárez Zapotec (Otomanguean, North America)	218
5.28	The top-scoring languages (by precision and recall) for GMM cluster 3	219
5.29	The top-scoring languages (by precision and recall) for GMM cluster 4	220
5.30a	Kriging map for Batad Ifugao (Austronesian, Papunesia)	220
5.30b	Kriging map for Konobo-Eastern Krahn (Kru, Africa)	220
5.31a	Kriging map for Nawuri (Atlantic-Congo, Africa)	221
5.31b	Kriging map for Tagalog (Austronesian, Papunesia)	221
5.32a	Kriging map for German (Indo-European, Eurasia)	221
5.32b	Kriging map for Adhola (Nilotic, Africa)	221
6.1	Kriging maps for Old Church Slavonic (Indo-European, Eurasia) and Ancient Greek (Indo-European, Eurasia), repeated	228
6.1a	Kriging map for Old Church Slavonic (Indo-European, Eurasia), repeated	228
6.1b	Kriging map for Ancient Greek (Indo-European, Eurasia), repeated	228
6.2	Kriging map for Later Church Slavonic	228
6.3a	Kriging maps for Old Church Slavonic after preprocessing and with syntactic placeholders for absolute constructions and conjunct participles	230
6.3b	Kriging maps for Ancient Greek after preprocessing and with syntactic placeholders for absolute constructions and conjunct participles	230
6.4a	Kriging map for Old Church Slavonic after adding the parallels to English <i>while</i> (in addition to <i>when</i>) and syntactic placeholders for absolute constructions and conjunct participles	230
6.4b	Kriging map for Ancient Greek after adding the parallels to English <i>while</i> (in addition to <i>when</i>) and syntactic placeholders for absolute constructions and conjunct participles	230
6.5	Kriging map for English (<i>when</i> and <i>while</i> situation).	231
6.6a	Overlap between the Kriging area for Old Church Slavonic absolute constructions and the Kriging area for English <i>while</i>	231
6.6b	Overlap between the Kriging area for Ancient Greek absolute constructions and the Kriging area for English <i>while</i>	231
6.7	Kriging map for Pular (Atlantic-Congo, Africa), repeated	233
6.8	Temporal connectives meaning ‘when’ in Pular according to Evans (2017: 61)	234

6.9	Pular WHEN-map with Kriging based on the functional annotation of Pular connectives	237
6.10a	Ancient Greek WHEN-map with Kriging based on functional labels, namely typical INDEPENDENT RHEMES and typical FRAMES	240
6.10b	Highlights of any point found inside the Kriging areas for FRAMES and INDEPENDENT RHEMES (red) in the Kriging map for Ancient Greek with added functional labels	240
6.11a	MDS maps of WHEN, highlighting all <i>jegda</i> -clauses, absolute constructions and conjunct participles in Old Church Slavonic	241
6.11b	MDS maps of WHEN, highlighting all <i>hote/hotan</i> -clauses, absolute constructions and conjunct participles in Ancient Greek	241
6.12	Location of the average observation for each different construction in Old Church Slavonic and Ancient Greek	242
6.13a	Kriging map for Avar (Nakh-Daghestanian, Eurasia)	245
6.13b	Kriging map for Mbyá Guaraní (Tupian, South America)	245
6.14	Kriging map for Cofán (Isolate, South America)	247
6.15a	Kriging map for Amele (Nuclear Trans New Guinea, Papunesia)	258
6.15b	Kriging map for Huichol (Uto-Aztecan, North America)	258
6.16a	MDS map of WHEN highlighting overt subordinators and switch reference markers in Amele (Nuclear Trans New Guinea, Papunesia)	259
6.16b	MDS map of WHEN highlighting overt subordinators and switch reference markers in Huichol (Uto-Aztecan, North America)	259
6.16c	MDS map of WHEN highlighting overt subordinators and switch reference markers in Cofán (Isolate, South America)	259

List of Tables

1	Main corpus breakdown	31
2	Breakdown of strategically annotated texts	32
1.1	Relative order of participle and matrix clause in the Codex Marianus . . .	39
1.2	Dative absolutes in the Codex Marianus: aspect distribution by position relative to the matrix clause	41
1.3	Conjunct participles in the Codex Marianus: aspect distribution by posi- tion relative to the matrix clause	41
1.4	Frequency of overt and null subjects among sentence-initial conjunct par- ticiples and dative absolutes in the Codex Marianus	49
1.5	Position of overt subjects relative to sentence-initial conjunct participles and dative absolutes in the Codex Marianus	50
1.6	Dative absolutes in the Codex Marianus: subject parts of speech	54
1.7	Dative absolutes in the Codex Marianus: ten most-frequent subject lemmas	57
1.8	Conjunct participles in the Codex Marianus: subject parts of speech . . .	57
1.9	Conjunct participles in the Codex Marianus: ten most-frequent subject lemmas	57
1.10	Dative absolutes and conjunct participles in the Codex Marianus: most frequent common nouns among subjects lemmas	58
1.11	Pre-matrix VS conjunct participles in the Codex Marianus: subject parts of speech	58
1.12	Pre-matrix SV conjunct participles in the Codex Marianus: subject parts of speech	58
1.13	Dative absolutes and conjunct participles in the Codex Marianus: ten most frequent lemmas, overall, by position of the participle relative to the matrix clause, and by position of the subject relative to the participle . .	62
1.14	Aktionsarten of pre-matrix conjunct participles and their matrix clause in the Codex Marianus	64
1.15	Aktionsarten of pre-matrix dative absolutes and their matrix clause in the Codex Marianus	65
1.16	Lexical variation among participles, overall, by position relative to the matrix clause, and by position of the subject relative to the participle . .	66

1.17	Old Church Slavonic constructions corresponding to a conjunct participle in Ancient Greek	67
1.18	Ancient Greek constructions corresponding to a conjunct participle in Old Church Slavonic	67
1.19	Old Church Slavonic constructions corresponding to a genitive absolute in Ancient Greek	69
1.20	Ancient Greek constructions corresponding to a dative absolute in Old Church Slavonic	71
1.21	Frequency of discourse relations tags among conjunct participles in the Codex Marianus	76
1.22	Frequency of discourse relations tags among dative absolutes in the Codex Marianus	76
1.23	Frequency of discourse relation tags among pre-matrix conjunct participles in the Codex Marianus	80
1.24	Frequency of discourse relation tags among post-matrix conjunct participles in the Codex Marianus	81
2.1	Relative order of participle and matrix clause in standard treebanks . . .	86
2.2	Dative absolutes in standard treebanks: aspect distribution by position relative to the matrix clause	89
2.3	Conjunct participles in standard treebanks: aspect distribution by position relative to the matrix clause	90
2.4	ELABORATIONS versus FRAME imperfective absolutes in post-matrix position	94
2.5	Frequency of ELABORATIONS and FRAME imperfective absolutes by variety, text (or manuscript), and by translated versus original Early Slavic texts	95
2.6	Overt and null subjects in sentence-initial conjunct participles and dative absolutes in standard treebanks	97
2.7	Position of overt subjects relative to sentence-initial conjunct participles and dative absolutes in standard treebanks	97
2.8	Dative absolutes in standard treebanks: subject parts of speech	97
2.9	Pre-matrix dative absolutes in standard treebanks: subject parts of speech	98
2.10	Post-matrix dative absolutes in standard treebanks: subject parts of speech	98
2.11	Subject parts of speech among conjunct participles that may head an overt subject in standard treebanks	99
2.12	Pre-matrix SV conjunct participles in standard treebanks: subject parts of speech	100
2.13	Pre-matrix VS conjunct participles in standard treebanks: subject parts of speech	100
2.14	Pre-matrix SV dative absolutes in standard treebanks: subject parts of speech	101

2.15	Pre-matrix VS dative absolutes in standard treebanks: subject parts of speech	101
2.16	Lexical variation among subjects in standard treebanks, by subcorpus and subject position	102
2.17	Sentence-initial dative absolutes and conjunct participles: ten most frequent subject lemmas by subcorpus and position of the subject relative to the participle	106
2.18	Lexical variation among participles by subcorpus	108
2.19	Dative absolutes and conjunct participles: ten most-frequent lemmas by subcorpus and position of the participle relative to the matrix clause . . .	114
2.20	Dative absolutes in strategically annotated treebanks: aspect distribution by sentence position	120
2.21	Conjunct participles in strategically annotated treebanks	121
2.22	Lexical variation among participles by subcorpus in strategically annotated treebanks	123
2.23	Dative absolutes and conjunct participles in strategically annotated treebanks: ten most-frequent lemmas by position of participle relative to the matrix clause	123
2.24	Dative absolutes in AQM: aspect distribution by sentence position	125
2.25	Conjunct participles in AQM: aspect distribution by sentence position . .	125
2.26	Dative absolutes in AQM: subject lemmas	125
2.27	Conjunct participles in AQM: subject lemmas	126
2.28	Conjunct participles in AQM: frequency of participle lemmas	128
2.29	Dative absolutes in AQM: frequency of participle lemmas	129
3.1	Ancient Greek constructions corresponding to a <i>jegda</i> -clause in Old Church Slavonic	138
3.2	Relative order of <i>jegda</i> -clause and matrix clause in the Codex Marianus .	142
3.3	Relative order of <i>jegda</i> -clause and matrix clause in standard treebanks . .	142
3.4	<i>Jegda</i> -clauses in the Codex Marianus: tense distribution by position relative to the matrix clause	142
3.5	<i>Jegda</i> -clauses in standard treebanks: tense distribution by position relative to the matrix clause	142
3.6	<i>Jegda</i> -clauses in the Codex Marianus: aspect distribution by position relative to the matrix clause	143
3.7	Overt and null subjects in <i>jegda</i> -clauses in the Codex Marianus	144
3.8	Overt and null subjects in <i>jegda</i> -clauses in standard treebanks	144
3.9	Subject coreference and non-coreference between <i>jegda</i> -clauses and their matrix clause in the Codex Marianus	144

3.10	Pick-up rates of subject referents in the previous discourse: pre-matrix non-coreferential <i>jegda</i> -clauses and pre-matrix dative absolutes, by subject realization and window of preceding sentences	147
3.11	Pick-up rates of subject referents in the previous discourse: pre-matrix coreferential <i>jegda</i> -clauses and (leftmost) pre-matrix conjunct participles, by subject realization and window of preceding sentences	147
3.12	Aktionsarten of <i>jegda</i> -clauses and their matrix clause in the Codex Marianus	151
3.13	Lexical variation among the verbs in <i>jegda</i> -clauses and participle constructions, by subcorpus and configuration	152
3.14	Mean, median and standard deviation of the length (in number of tokens) of participle constructions and <i>jegda</i> -clauses in different configurations . .	155
4.1	Tense of <i>jegda</i> -clauses and their matrix clause in the Codex Marianus and standard treebanks (combined)	160
5.1	The 10 most frequent language families in our dataset compared to the 10 most frequent families among the world's languages according to the Glottolog classification	188
5.2	The 10 most frequent families among the world's languages according to the Glottolog classification compared to the 10 most frequent language families in our dataset	188
5.3	Areal distribution of the languages in our dataset compared to the world's languages, following the classification into macro-areas from Glottolog. . .	189
5.4	Matrix of <i>when</i> and aligned tokens (sample).	190
5.5	Frequency of lexification patterns	208
5.6	Frequency of lexification patterns, including subpatterns within a main pattern	209
A.1	Language families represented in the massively parallel Bible dataset compared with the world's language families according to Glottolog.	284
B.1	Basic coexpression patterns	285
B.2	Subpatterns of coexpression	287
B.3	Classification of the languages in the massively parallel dataset into patterns and subpatterns.	328

List of Abbreviations

Glossing and abbreviations follow the Leipzig Glossing Rules, with the following additions.

A	transitive subject function; Actor macrorole; Actor
AOR	aorist
COLL	collective
DS	different subject
EXIST	existential
IMPF	imperfect
NARR	narrative
NSBJ	non-subject
PRED	predicate marker
PTC	discourse particle
REMP	remote past tense
RESULT	resultative
SEQ	sequence
SIM	simultaneous
SS	same subject
SUP	supine
TODP	today's past tense
VIS	visible, speaker's area
U	undergoer macrorole
UNIV	universal

Introduction

0.1 The problem

Participial clauses are employed widely in Early Slavic and are in evident competition with finite clauses introduced by the generic temporal subordinator *jegda* ‘when’ (1)-(2).¹

- (1) a. *viděv* *že* *feklu* *celova* *ju*
 see.PTCP.PFV.M.NOM.SG PTC Thecla.ACC kiss.AOR.3.SG 3.SG.F.ACC
 ‘When he saw Thecla, he kissed her’ (*Life of Thecla*, Bdinski Sbornik f. 50v)
- b. *Egda* *že* *se* *vidě* *Zosima,* *too* *vъ*
 when PTC this.N.ACC.SG see.AOR.3.SG Zosima.NOM then in
veštši *strach* *vpade*
 greater.ACC terror.ACC fall.AOR.3.SG
 ‘When Zosima saw this, he fell in a greater terror’ (*Life of Mary of Egypt*, Bdinski Sbornik ff. 169r-169v)
- (2) a. *tože takože i* *ot* *boljarъ* *mъnozi*
 also so PTC from Boyars.GEN.PL many.NOM.PL
prichodeš̃te *povědachutъ* *jemu* *gněvъ*
 come.PTCP.IPFV.NOM.PL speak.IMPV.3.PL 3.SG.M.DAT anger.ACC.SG
knežъ *na togo* *suštъ* *i*
 prince’s.ACC.SG on that.GEN.SG be.PTCP.IPFV.ACC.SG and
molechutъ *i* *ne suprotiviti se* *jemu*
 beg.IMPV.3.PL PTC NEG oppose REFL 3.SG.M.DAT
 ‘And many of the Boyars, when they came to him, would speak of the prince’s anger and begged not to oppose the prince.’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 58v)
- b. *i jegda* *že* *sii* *prichožaachu* *kъ* *njemu* *to*
 and when PTC this.NOM.PL come.IMPV.3.PL to 3.SG.M.DAT then
že si *i tako* *po bžbъstvъněmъ* *tomъ*
 PTC this.NOM.SG PTC so after divine.LOC.SG that.LOC.SG

¹Citations under the examples use italics for the (conventional) title of the text and regular font for the name of the manuscript. Citations of the Early Slavonic New Testament are all from the Old Church Slavonic Codex Marianus in PROIEL, so the manuscript name is omitted throughout the thesis. Citations of the Greek New Testament are instead all from von Tischendorf’s (1869–1872) version in PROIEL. Translations of biblical passages are mostly adapted from the King James Version and the New King James Version. Translations of all the examples from the *Life of Mary of Egypt* from the Bdinski Sbornik are by Coelho 2020. Translations of other examples, both Early Slavic and other languages, are my own unless otherwise noted.

učeníi *prědǫstavljaaše* *těmǫ* *trepezu* *ot*
 teaching.LOC.SG treat.IMPF.3.SG that.DAT.PL dining table.ACC.SG from
brašbnǫ *těchǫ* *manastyrbŭskychǫ*
 food.GEN.PL that.GEN.PL monastery's.GEN.PL

‘And when someone came to Theodosius, in the same way, after the spiritual conversation, he would treat those who came to dinner with the supplies of the monastery.’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 46b)

The focus of previous scholarship on Early Slavic participle clauses has been overwhelmingly on their origin and relationship with their Ancient Greek counterparts (e.g. Birnbaum 1958; Růžička 1958, 1961, 1963; Večerka 1961, 1997; Skupskij 1993), which show very similar functions, rather than on the motivations governing their distribution from the synchronic perspective in the Early Slavic corpus. Several works (e.g. Trost 1958; Růžička 1961; Andersen 1969; Berent 1973; Worth 1994; Corin 1995; Collins 2004, 2011) have also been dedicated to the specific functions of Early Slavic absolute constructions, a type of participle clause structurally akin to English absolutes (e.g. *all things considered* or *weather permitting*), but which receives special dative marking, both on the participle and on its subject, if overt. Absolute constructions are regularly found in distributional overlap with *jegda*-clauses, as in (3). As (4) shows, particularly sentence-initially, competition can also be observed between absolute constructions and agreeing participle clauses (‘conjunct participles’).

- (3) a. *i* ***koncejuštju*** *sę* *lět* *tomu*.
 and finish.PTCP.IPFV.N.DAT.SG REFL year.DAT.SG that.DAT.SG
vygnaša *žiroslava* *is* *posadnicŭstva*. *i*
 drive out.AOR.3.PL Žiroslav.GEN.SG from posadnichestvo.GEN.SG and
daša *zavidu* *nerevnicju*
 give.AOR.3.PL Zavid.DAT.SG Nerevenic.DAT.SG

‘And towards the end of that year [lit. when that year was ending], they drove Žiroslav out of the *posadnichestvo* and gave it to Zavid Nerevenic’ (*Novgorod First Chronicle*, Synodal Manuscript f. 40r)

- b. ***egda končaša*** *sę* *dnŭe* *.e. desętŭnii*.
 when finish.AOR.3.PL REFL day.GEN.PL fiftieth.NOM.PL
snide *dchǫ* *styi* *na aply*
 descend.AOR.3.SG spirit.NOM.SG holy.NOM.SG on apostle.ACC.PL
 ‘When fifty days had passed, the Holy Spirit descended on the apostles’
 (*Primary Chronicle*, Codex Laurentianus f. 35v)

- (4) a. *i* ***prišedšju*** *emu* *v korsunŭ*
 and arrive.PTCP.PFV.M.DAT.SG 3.SG.M.DAT in Korsun.ACC
uvidě *jako is* *korsunę* *blizŭ* *ustŭe*
 learn.AOR.3.SG that from Korsun.GEN close.NOM mouth.NOM
dněprŭskoe
 Dniepr's.NOM

‘When he arrived in Korsun, he learned that the mouth of the Dnieper was not far from Korsun’ (*Primary Chronicle*, Codex Laurentianus f. 3v)

- b. *i prišed v vizantiju vidě na*
 and arrive.PTCP.PFV.M.NOM.SG in Byzantium.ACC see.AOR.3.SG in
tomě městě z gorě i glušičь
 that.LOC.SG place.LOC.SG seven hill.GEN.PL and bay.GEN.PL
morskychь mnogo
 sea.GEN.PL much.ACC.SG
 ‘And when he arrived in Byzantium, he saw seven hills and many sea gulfs’
 (*The Tale of the Fall of Constantinople* f. 283r)

To pinpoint the factors explaining the distribution of competing constructions, one needs to draw from several levels of analysis at once. Consider (5a)-(5d).

- (5) a. Having finished his thesis, he cried with relief
 b. His thesis (being) finished, he cried with relief
 c. When he finished his thesis, he cried with relief
 d. He finished his thesis and cried with relief

To various extents, all these sentences relate two clauses in semantically underspecified ways. The nonfiniteness of the participle clause in (5a) and (5b) indicates dependency on the following (matrix) clause. In the absence of an explicit connective, however, their semantic relation must be pragmatically inferred, often similarly to syntactically coordinated independent clauses (5d). In (5c) the first finite clause is introduced by the temporal subordinating conjunction *when*. Yet, the precise temporal relation between the *when*-clause and the matrix clause is not established by the subjunction itself, since, unlike more explicit subjunctions, such as *after* or *before*, *when* is temporally underspecified and compatible with different core temporal interpretations (e.g. anteriority, simultaneity, inclusion). Whichever the surface realization, however, all these examples intuitively convey that the *crying* event is an effect of the *thesis-finishing* event, and that the second situation occurs at some point (sooner or later) after the first situation.

On one level, we may consider the compositional semantics of participle clauses at the clause- and sentence-level and understand how far compositionality can go in explaining functional differences between constructions. As we will see, the core temporal semantic properties of Early Slavic participle clauses, for example, can be accounted for, to a large extent, by the compositional semantics of the tense-aspect system of Early Slavic, specifically how the aspect of participles interacts with their matrix clause. On another level, the use of different participle clauses or *jegda*-clauses entails the strategic selection of one construction over another at the level of discourse structure. This involves temporal relations *across* sentences and the broader level of text organization, in which information structure and the rhetorical relations between *discourse* units play a crucial role.

The goal of this thesis is to investigate the semantic and pragmatic properties of Early

Slavic conjunct participles and absolute constructions to understand what principles motivate their selection over one another and over *jegda*-clauses at the synchronic level. The issue is tackled broadly using two different approaches. First, it uses detailed linguistic annotation on Early Slavic corpora at the morphosyntactic, dependency, information-structural, and lexical level, as well as parallel Ancient Greek data (when available) aligned at the token level, to obtain indirect evidence for different potential functions of participle clauses. This approach is broadly concerned with the aforementioned tension between compositionality and default discourse reasoning in accounting for the distribution of participle constructions and *jegda*-clauses in the Early Slavic corpus. Second, it uses ‘massively parallel’ data (Mayer & Cysouw 2014) and leverages typological variation in how languages express the semantic space covered by these constructions to assess the extent to which the division of labour observed among participle constructions and *jegda*-clauses is reflected cross-linguistically. Clear correspondences and differences with Early Slavic are then exploited to make generalizations about the core semantic-pragmatic properties of different competing temporal constructions in Early Slavic.

In the rest of this introduction, I first provide a preliminary formal definition of conjunct participles and absolute constructions (Section 0.2.1) and introduce the main theoretical frameworks to which this thesis will make consistent reference to. In particular, Section 0.2.2 summarizes Bary & Haug’s (2011) compositional-semantic account of the temporal semantics and discourse properties of conjunct participles in Ancient Greek, on which the first three chapters of the thesis will be based, while Section 0.2.3 touches on Charolles’s (2005) *framing adverbials*, Fabricius-Hansen & Haug’s (2012b) *fronted adjuncts*, and Chafe’s (1976) and Krifka’s (2007) *frame setters*. Section 0.2.4 introduces Segmented Discourse Representation Theory, a formal framework for discourse representation which will be used as a tool to reason about and formalize rhetorical relations between clauses beyond compositional-semantic accounts such as Bary & Haug’s (2011). Section 0.3 gives an overview of some of the main issues in the interpretation of *when*-clauses, on the basis of previous extensive research on event structure, temporal relations, and the anaphoric properties of tense, while Section 0.4 elaborates on the need for a gradient approach to the study of competition and introduces the concept of token-based typology, which is central to the methodology used in the second part of the thesis. Finally, Section 0.5 presents the Early Slavic corpus used in this work and Section 0.6 summarizes the structure of the thesis.

0.2 Background

0.2.1 Defining conjunct participles and dative absolutes

Under a Lexical-Functional Grammar (LFG) framework, the participle constructions under analyses are classified as *adjuncts*, namely grammatical functions that are not

subcategorizable (i.e. required as arguments) by a predicate (Dalrymple, Lowe & Mycock 2019: 12).

We can distinguish two main types of participial adjuncts in Early Slavic. In so-called (*co-*)*predicative*² or *conjunct* participles, the subject is coreferential with an argument of the matrix VP, most typically the subject, which determines the case of the participle form itself. In other words, conjunct participles are *functionally controlled* by an argument of the matrix clause. Control constructions are defined in LFG as those in which ‘either syntactic or lexical constraints require coreference between an argument of the matrix clause [controller] and an argument of a subordinate or modifying adjunct clause [controllee]’ (Dalrymple et al. 2019: 543). In functional (as opposed to anaphoric) control, the case of the controllee is determined by the grammatical function (e.g. SUBJECT, OBJECT, etc.) of the controller.

(6)-(8) are typical examples of conjunct participles.

- (6) *azъ prišedъ iscěljo i*
 1.SG.NOM come.PTCP.PFV.M.NOM.SG heal.PRS.1.SG 3.SG.M.ACC
 ‘I will come and heal him’ (Matthew 8:7)
- (7) *vykradše pervoe svętoslava po tomъ*
 kidnap.PTCP.PFV.M.NOM.PL first Svyatoslav.GEN after that.LOC.SG
ubiša kytana i družinu jeho vsju
 kill.AOR.3.PL Kytan.GEN and retinue.ACC.SG 3.SG.M.GEN entire.ACC.SG
izbiša
 thrash.AOR.3.PL
 ‘After kidnapping Svyatoslav first, they killed Kytan and thrashed his entire retinue’ (*Primary Chronicle*, Codex Laurentianus f. 75d)
- (8) *i padši kolěnoma i čelomъ*
 and fall.PTCP.PFV.F.NOM.SG knee.INST.DU and brow.INST.SG
tlbkjuštii [sic] *na zemli před ikonu*
 touch.PTCP.IPFV.F.NOM.SG on ground.LOC.SG before icon.ACC.SG
styje bce i sia slovesa
 holy.GEN.SG Mother of God.GEN.SG and this.ACC.PL word.ACC.PL
načech glati
 start.AOR.1.SG speak.INF
 ‘Falling on my knees and touching the ground before the icon of the holy Mother of God with my brow, I started speaking the following words’ (*Life of Mary of Egypt*, Bđinski Sbornik f. 177r).

Conjunct participles are thus adjuncts with an external argument, an ‘open slot’ which needs to be filled—in LFG terms, they are XADJUNCTS of verb-headed (as opposed to noun-headed, as in attributive uses) *f*-structures (as in (9), from example (6)), or verb-headed *open adjuncts* in other frameworks (e.g. Fabricius-Hansen & Haug 2012a).

²The term *copredicative* participles is used by Haspelmath (1995: 17–20), rather than only *predicative*, used, for example, by Eckhoff, Majer, Welo & Haug (2009), Bary & Haug (2011) and Haug (2011).

$$(9) \left[\begin{array}{l} \text{PRED} \quad \langle \text{icělit}i \langle \text{SUBJ,OBJ} \rangle \rangle \\ \text{ASP} \quad \text{PFV} \\ \text{TENSE} \quad \text{PRS} \\ \text{SUBJ} \quad \boxed{1} \left[\text{PRED} \quad \langle \text{az} \bar{\text{c}} \rangle \right] \\ \text{OBJ} \quad \left[\text{PRED} \quad \langle \text{i} \rangle \right] \\ \text{XADJ} \quad \left[\begin{array}{l} \text{PRED} \quad \langle \text{priti} \langle \text{SUBJ} \rangle \rangle \\ \text{ASP} \quad \text{PFV} \\ \text{SUBJ} \quad \boxed{1} \end{array} \right] \end{array} \right]$$

Absolute constructions, on the other hand, are not functionally controlled by any argument of the matrix VP. They thus have an internal subject which needs not be coreferential with any constituent in the superordinate clause. As in several other early Indo-European languages (e.g. Ancient Greek, Latin, Gothic, Sanskrit), both the participle and its subject generally appear in oblique cases, dative in Early Slavic, which is why we often refer to this construction as *dative absolute*. (10)-(12) are typical examples of dative absolutes from the same sources as (6)-(8).

- (10) *i s̄chodeštem̄ im̄ s̄ gory*
 and come.down.PTCP.IPFV.M.DAT.PL 3.PL.DAT from mountain.GEN.SG
zapovědě im̄ is̄ gl̄e
 command.AOR.3.SG 3.PL.DAT Jesus.NOM say.PTCP.PRS.M.NOM.SG
 ‘As they came down from the mountain, Jesus commanded them saying’
 (Matthew 17:9)

- (11) *i umnoživšem̄ sē člvkom̄ na zemli*
 and multiply.PTCP.PFV.M.DAT.PL REFL person.DAT.M.PL on earth.LOC.SG
pomyšlīša sozdati stolp̄ do nbse v̄ dni
 plan.AOR.3.PL create.INF pillar.ACC.SG to heaven.GEN.SG in day.ACC.PL
nektana i faleka
 Nectan.GEN.SG and Peleg.GEN.SG
 ‘And when people had multiplied on earth, they planned to build a pillar to heaven in the days of Nectan and Peleg.’ (The *Primary Chronicle*, Codex Laurentianus f. 2v)

- (12) *lētu skončavšu se i priide paki*
 year.DAT.N.SG end.PTCP.PFV.N.DAT.SG REFL PTC come.AOR.3.SG again
zosima v̄ pustynju
 Zosima.NOM in desert.ACC.SG
 ‘When the year was ending Zosima came again to the desert’ (*Life of Mary of Egypt*, Bdinski Sbornik f. 187r)

If conjunct participles are open adjuncts, then absolutes can be considered *closed* adjuncts. In LFG terms, they are ADJuncts of verb-headed *f*-structures, as represented in (13) from example (10).

(13)	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 10px;">PRED</td> <td style="padding: 2px 10px;">‘zapověděti <SUBJ, OBJ, OBL_{GOAL}>’</td> </tr> <tr> <td style="padding: 2px 10px;">ASP</td> <td style="padding: 2px 10px;">PFV</td> </tr> <tr> <td style="padding: 2px 10px;">TENSE</td> <td style="padding: 2px 10px;">PST</td> </tr> <tr> <td style="padding: 2px 10px;">SUBJ</td> <td style="padding: 2px 10px;">[PRED ‘Isusъ’]</td> </tr> <tr> <td style="padding: 2px 10px;">OBL_{GOAL}</td> <td style="padding: 2px 10px;">[PRED ‘i*’₁]</td> </tr> <tr> <td style="padding: 2px 10px;">OBJ</td> <td style="padding: 2px 10px;">[...]</td> </tr> <tr> <td style="padding: 2px 10px;">ADJ</td> <td style="padding: 2px 10px;">[PRED ‘sъchoditi <SUBJ, OBL_{SOURCE}>’]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[ASP IPFV]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[SUBJ [PRED ‘i*’₁]]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[OBL_{SOURCE} [PRED ‘sъ <OBJ>’]]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[OBJ [PRED ‘gora’]]]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[XADJ [PRED ‘glagolati <SUBJ>’]]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[ASP IPFV]]</td> </tr> <tr> <td style="padding: 2px 10px;"></td> <td style="padding: 2px 10px;">[SUBJ 1]]]</td> </tr> </table>	PRED	‘zapověděti <SUBJ, OBJ, OBL _{GOAL} >’	ASP	PFV	TENSE	PST	SUBJ	[PRED ‘Isusъ’]	OBL _{GOAL}	[PRED ‘i*’ ₁]	OBJ	[...]	ADJ	[PRED ‘sъchoditi <SUBJ, OBL _{SOURCE} >’]		[ASP IPFV]		[SUBJ [PRED ‘i*’ ₁]]		[OBL _{SOURCE} [PRED ‘sъ <OBJ>’]]		[OBJ [PRED ‘gora’]]]		[XADJ [PRED ‘glagolati <SUBJ>’]]		[ASP IPFV]]		[SUBJ 1]]]
PRED	‘zapověděti <SUBJ, OBJ, OBL _{GOAL} >’																												
ASP	PFV																												
TENSE	PST																												
SUBJ	[PRED ‘Isusъ’]																												
OBL _{GOAL}	[PRED ‘i*’ ₁]																												
OBJ	[...]																												
ADJ	[PRED ‘sъchoditi <SUBJ, OBL _{SOURCE} >’]																												
	[ASP IPFV]																												
	[SUBJ [PRED ‘i*’ ₁]]																												
	[OBL _{SOURCE} [PRED ‘sъ <OBJ>’]]																												
	[OBJ [PRED ‘gora’]]]																												
	[XADJ [PRED ‘glagolati <SUBJ>’]]																												
	[ASP IPFV]]																												
	[SUBJ 1]]]																												

Traditionally, subject non-coreferentiality (or *switch reference*) between the absolute construction and the matrix clause has been considered obligatory, but, as it has often been pointed out, this is by no means always the case in the Early Slavic sources (cf. Worth 1994; Collins 2004, 2011), similarly to what we also independently observe in Ancient Greek (cf. Fuller 2006). (14) and (15) are examples of subject co-reference in Early Slavic dative absolutes.

- (14) *i bystъ idōštemъ imъ. ištistišę*
 and happen.AOR.3.SG go.PTCP.IPFV.DAT.PL 3.PL.DAT cleanse.AOR.3.PL
sę
 REFL
 ‘And it came to pass that, as they went, they were cleansed’ (Luke 17:14)

- (15) *i všedšim v gorodъ utěšista*
 and enter.PTCP.PFV.DAT.PL in city.ACC.SG console.AOR.3.DU
volodimercę
 people of Vladimir.ACC.PL
 ‘And when they entered the city, they consoled the people of Vladimir’ (*Suzdal Chronicle*, Codex Laurentianus f. 126r)

Such cases of subject co-referentiality have sometimes been called ‘nonstandard’ (Worth 1994) or ‘non-canonical’ (Collins 2011), if not outright ‘aberrations’ or signs of ‘decay’ (e.g. Večerka 1961: 49; Večerka 1996: 190; Corin 1995: 268). There are other frequently

observed instances of ‘non-canonical’ absolutes (as well as conjunct participles), most notably those which are overtly coordinated to their matrix or which do not have a clear matrix clause they attach to. Conjoined absolutes have been commented on particularly in the literature on Old East Slavic (cf., for example, Nečásek 1957; Večerka 1961; Borkovskij & Kuznecov 1963; Alekseev 1987; Gebert 1987; Remneva 1989; Corin 1995), but it is in fact well attested in South Slavic recensions of Church Slavonic as well (cf. Grković-Major 2007 and Collins 2011). As argued by Collins (2011), several ‘non-canonical’ absolutes should be interpreted by taking into account the broader discourse context rather than the inter-clausal or inter-sentential level. As I will show in the second part of this thesis, these occurrences are not at odds with the core functions of dative absolutes as discourse devices. In fact, most, if not all, the non-canonical usages of absolutes previously observed in the literature are phenomena frequently observed to co-occur cross-linguistically on constructions with similar functions as Early Slavic absolutes.

0.2.2 Conjunct participles according to Bary & Haug (2011)

According to Bary & Haug (2011), conjunct participles in Ancient Greek fulfil one of three main discourse functions (labelled as *FRAMES*, *INDEPENDENT RHEMES*, *ELABORATIONS*), each involving a specific temporal relation to the matrix clause and certain discourse effects that can be modelled at the compositional level. As the corpus evidence in Haug (2012) shows, each function is more likely to surface in specific syntactic, semantic and information-structural configurations. Bary & Haug’s (2011) analysis is based on Ancient Greek, but we can easily exemplify the concept, in a nutshell, using Early Slavic participles, leaving the semantic formalization aside for simplicity. Since both Old Church Slavonic and Ancient Greek participles have been shown to express aspect, rather than relative tense (Eckhoff & Haug 2015; Kamphuis 2020), I will talk about the ‘aspect’ of participles and refer to what is traditionally called ‘past participles’ and ‘present participles’ as ‘perfective participles’ and ‘imperfective participles’ instead, which is reflected in the glossing criteria.

- *FRAMES* set the stage for and provide temporal anchoring for the matrix event.
 - They are strictly anaphoric/presuppositional: they depend on the context for their own temporal reference, thus referring to previously mentioned or easily inferable situations. In (16), for example, there is no explicit mention of Mstislav leaving for Pereyaslav, so the motion event is contextually inferred and used as anchor for the matrix event.
 - Because they are presuppositional, they tend to occur sentence-initially.
 - They are not modally dependent on the matrix verb, so that if the matrix verb is an imperative, for example, the participle is not interpreted as part of the command, as (17) shows.

- When perfective, they always induce *narrative progression*, whereby the matrix eventuality is interpreted as taking place (just) after the participle event. When imperfective, the adjunct event includes the matrix event, as in (18).
- Their dominant verbal aspect within a corpus is likely to vary with the genre: ‘both while *x*-ing and after *x*-ing are possible ways of linking to a previously mentioned or accessible event’ (Haug 2012: 311-312).
- They are more likely to constitute ‘predictable’ (Haug 2012) predications in the discourse in which they appear, which may be expected to be reflected as overall lower lexical variation (Haug 2012: 312).

(16) *i prišedŭ mstislavŭ v perejaslavŭ.*
 and arrive.PTCP.PFV.M.NOM.SG Mstislav.M.NOM in Pereyaslav.ACC
poima ženu svoju ide v
 take.PTCP.IPFV.M.NOM.SG wife.ACC.SG his.F.ACC.SG go.AOR.3.SG to
lučeskŭ.
 Luchesk.ACC

‘And when Mstislav arrived in Pereyaslav, he went to Luchesk, taking his wife with him’ (*Suzdal Chronicle*, Codex Laurentianus f. 115r)

(17) *ty že postę seř pomaži*
 2.SG.NOM PTC fast.PTCP.IPFV.M.NOM.SG REFL anoint.IMP.2.SG
glavŭ svojoř i lice tvoe umyi
 head.ACC.SG your.ACC.SG and face.ACC.SG your.ACC.SG wash.IMP.2.SG
 ‘But when you fast, put oil on your head and wash your face’ (Matthew 6:17)

(18) *iščodešće že obrętoř čka*
 go out.PTCP.IPFV.M.NOM.PL PTC receive.AOR.3.PL person.GEN.SG
kýřiněiska
 of Cyrene.GEN.SG

‘As they were going out, they met a man from Cyrene’ (Matthew 27:32)

- INDEPENDENT RHEMES are discourse-coordinated, therefore information-structurally very similar to independent clauses.
 - Like FRAMES, they relate to the preceding context for their temporal reference, but unlike FRAMES, they generally introduce new events.
 - They also provide the temporal reference for the next event in line and can be stacked up similarly to a series of independent clauses, as in (19).
 - Unlike FRAMES, they are modally dependent on the matrix verb. They are, for example, interpreted as part of the command if the matrix is an imperative, as in (20).
 - Like FRAMES, they also always produce narrative progression when perfective.

- They are predominantly perfective and precede the matrix verb.
- They are expected to encode new information: this could surface as overall more lexical variation among participles with this function.

- (19) *i abie tekъ edinъ otъ*
 and immediately run.PTCP.PFV.M.NOM.SG one.M.NOM.SG from
nichъ. i priemъ gobъ.
 3.PL.GEN and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG
isplъnъ ocъta. i vъznezъ
 fill.PTCP.PFV.M.NOM.SG vinegar.GEN.SG and put.PTCP.PFV.M.NOM.SG
na trъstъ. napaěše i
 on reed.ACC.SG give.to.drink.IMPF.3.SG 3.M.SG.ACC
 ‘Immediately one of them ran and took a sponge, filled it with sour wine and put it on a reed, and offered it to him to drink’ (Matthew 27:48)

- (20) *i ně li ti tamo povoica. a*
 and not be.3.SG.PRS.ACT if 2.SG.DAT there veil.SG.GEN then
kъvoěi prisoli
 buy.PTCP.PFV.F.NOM.SG send.IMP.2.SG
 ‘And if you don’t have a veil there, then buy one and send it.’ (Birch bark letter 731; translation by Schaecken 2018a)

- ELABORATIONS add granularity to the semantics of the main event and typically express manner or means/instrument.
 - They are more likely to be imperfectives since their function is to provide more information on the matrix event by describing “concomitant circumstances” that are cotemporal with the main event’ (Haug 2012: 312).
 - They are more likely to encode new information.
 - They are more likely to follow the matrix clause, because they are temporally dependent on the main verb and because they are non-topical and non-focused.
 - Like INDEPENDENT RHEMES they are modally dependent on the matrix verb, as (21) shows.

- (21) *ašte ty esi chъ. rъci namъ ne*
 if 2.SG.NOM be.PRES.2.SG Christ.NOM say.IMP.2.SG 1.PL.DAT NEG
obinъje se
 conceal.PTCP.IPFV.M.NOM.SG REFL
 ‘If You are the Christ, tell us plainly’ (John 10:24)

Bary & Haug’s (2011) framework is mostly built on the analysis of conjunct participles,³ which begs the question of whether absolute constructions also fulfil all three functions

³The authors only include one example containing an absolute participle, which is treated as a FRAME (see Bary & Haug 2011: 8).

with some frequency. The intuition is that absolute constructions should typically show properties of FRAMES, given the main characterization of dative absolutes in the literature as ‘backgrounding’ or ‘stage-setting’ devices (Worth 1994: 30; Corin 1995: 259; Collins 2011: 113).

As already mentioned, earlier scholarship on Early Slavic adjunct participle clauses has overwhelmingly focussed on their origin and relationship with Ancient Greek conjunct participles and genitive absolutes. More recently, Sakharova (2007, 2010) analysed the pragmatic motivations for the differences in the distribution of participle constructions and finite independent clauses for a number of verbs in Old East Slavic *Novgorod First Chronicle*. Sakharova suggests that the usage of participle constructions (dative absolutes and conjunct participles alike) mostly depends on their ‘secondary status’ (*второстепенности*) in the discourse (Sakharova 2010: 88), which in most cases has to do with discourse properties typical of ‘backgrounded’ units. Background and foreground, according to Sakharova, can be defined on the basis of ‘what information is to be considered more or less relevant to the rhetorical goal’:⁴ *foreground* refers to ‘unexpected, non-stereotypical’ situations, whereas *background* refers to predictable information about ‘stereotyped sequences of situations’.⁵ This characterization of background and foreground unit bears, in fact, great similarities with the properties of FRAMES and INDEPENDENT RHEMES, respectively, as described by Bary & Haug (2011) and outlined above.

0.2.3 Frame setters and fronted adjuncts

Much of the discussion in Chapters 1-4 revolves around the possible framing function of participle constructions and *jegda*-clauses, and the configurations in which their interpretation as FRAMES is, in fact, available. As we will see already at the beginning of Chapter 1, the sentence-initial position is where conjunct participles may be ambiguous between a FRAME and an INDEPENDENT RHEME function, and it is also where the vast majority of dative absolutes in the Gospels are found, which, I will argue, is in itself indicative of their typical function as FRAMES. As argued by Bary & Haug (2011), FRAMES are generally found sentence-initially, since they are presuppositional and because they set the stage and provide temporal anchoring for the matrix event. In this, they are similar to Charolles’s (2005) *framing adverbials*, Fabricius-Hansen & Haug’s (2012b) *fronted adjuncts*, and Chafe’s (1976) and Krifka’s (2007) *frame setters*, all of which are most naturally fronted because of their discourse-cohesive function of providing ‘the frame in which the following expression should be interpreted’ (Krifka 2008: 269), and because ‘they are topical (i.e. “anchored” in the preceding context)’ (Charolles 2005: 13) and locate ‘the matrix event relative to an eventuality that [...] has already been explicitly

⁴[...] *какую информацию считать менее соответствующей риторической цели* (Sakharova 2010: 89).

⁵*Ожидаемой и, следовательно, менее выделенной прагматически, оказывается информация о стереотипных последовательностях ситуаций.* (Sakharova 2010: 89).

established in or is inferable from the preceding context’ (Fabricius-Hansen & Haug 2012b: 54). It is important to keep in mind some of the properties in common to these notions and Bary & Haug’s (2011) FRAMES since they will be particularly useful to account for a group of absolute constructions that, outside of the dataset used in Chapter 1, occur in post-matrix position but whose function is still compatible with Bary & Haug’s (2011) FRAMES, despite these being overwhelmingly more frequent in pre-matrix position, as already mentioned.

Frame setters, according to Chafe (1976: 50), are used ‘to limit the applicability of the main predication to a certain restricted domain’. Krifka (2007) built on Chafe’s (1976) definition by adding that, where a frame setter is not explicit, the domain or precise evaluation of the main predicate is unspecified (hence presumably inferred). Explicit frame setters specify this domain by choosing one out of a set of possible frames, thereby stating that the main predicate holds within that frame—which is why, according to Krifka, frame setters always contain a separate focus, which is meant as ‘the presence of alternatives that are relevant for the interpretation of linguistic expressions’ (Krifka 2007: 247). Krifka’s treatment of frame setters (like that of co-eventive adjuncts in Fabricius-Hansen & Haug 2012b, as well as that of participle adjuncts in Bary & Haug 2011) is based on the assumption shared by several theories of information structure and non-linear prosody (e.g. Truckenbrodt 1995, 2007; Büring 2006; Féry 2007, 2008) that sentences present a double bipartition into *background-focus*⁶ and *topic-comment* (also called *theme-rheme*). While *focus* always refers to the presence of alternatives, *topic* can refer to *aboutness* topic, in broad terms corresponding to ‘the entity that a speaker identifies, about which then information, the comment, is given’ (Krifka 2007: 265), but also to frame setters, sometimes also referred to as *frame topics* (e.g. Fabricius-Hansen & Haug 2012b: 41) or *frame setting topics* (e.g. Féry 2007; Song 2017: 25). Both aboutness and frame-setting topics refer to entities that are already established or inferable from the preceding discourse—‘individuals in the case of aboutness topics, time and places in the case of frame topics’ (e.g. Fabricius-Hansen & Haug 2012b: 41). While sentences always have explicit foci, topics are not necessarily expressed (as in [*the HOUSE is on fire*]*Comment*; Krifka 2007: 267).⁷

The relation between *topic-comment* and *background-focus* is much more complex than the mere equivalences *topic = background* and *comment = focus*. Crucially, a topic can contain a focus. Though not pursuing the issue in depth, Fabricius-Hansen & Haug (2012b) suggest that closed adjuncts (i.e. absolutes), because of their internal subject, may have a topic of their own serving as a contrastive partial topic of a common topic (i.e. the referent introduced by the subject of the matrix). A contrastive topic, in Krifka’s

⁶Note that this is not the same as *background* as used so far, which is more in line with the usage of the term in formal theories of discourse representation such as SDRT.

⁷The example given by Krifka (2007) leverages prosody (as the all-capitals on *house* indicates), to make certain arguments stronger—something which we cannot obviously rely on for a dead language.

(2007) treatment, is ‘an aboutness topic that contains a focus, which is doing what focus always does, namely indicating an alternative’ (Krifka 2007: 267), which in the case of contrastive topic is an alternative aboutness topic, as in (22).

(22) (From Krifka 2007: 268)

A: Does your sister speak Portuguese?

B: [My [BROther]_{Focus}]_{Topic} [[DOES]_{Focus}]_{Comment}

Fabricius-Hansen & Haug (2012b) argue that clause-final restrictive adjuncts (which, they note, often receive either mildly or strongly contrastive readings) semantically correspond to frame-setting fronted adjuncts, so that the former must be syntactically analysed on a par with fronted adjuncts, namely adjoined high up in the sentence—‘above rather than below the subject of the matrix clause’ (p.54). In (23a), for instance, the closed adjunct *with the leaves on the trees* may receive the restrictive interpretation which is instead explicitly triggered in its paraphrasis in (23b). Similarly, Krifka (2007: 270) compared contrastive topics and frame setters, saying that what they have in common ‘is that they express that, for the communicative needs at the current point of discourse, the current contribution only gives a limited or incomplete answer’. This is one of the information structural differences between (23e), which has no adjunct, and the others in (23). The domain of (23e) has to be inferred from the context, since there is no frame-setter providing one out of a set of alternatives within which the eventuality holds.

(23) (a. and b. from Fabricius-Hansen & Haug 2012b: 48)

- a. The cave entrance is easy to miss *with the leaves on the trees*
- b. The cave entrance is easy to miss *when the leaves are on the trees*
- c. *With the leaves on the trees* the cave entrance is easy to miss
- d. *When the leaves are on the trees* the cave entrance is easy to miss
- e. The cave entrance is easy to miss

Crucially, in all of (23a)-(23d) the adjunct is very much FRAME-like from the temporal perspective. The (mildly) contrastive reading possible in at least (23a) and (23b) does not change our observations about the type of semantic contribution these make compared to that in (23c) and (23d).

0.2.4 Segmented Discourse Representation Theory

Bary & Haug’s (2011) framework intentionally leaves aside semantic effects that derive from default discourse reasoning (cf Bary & Haug 2011: 36-37) and only formally models the temporal dimension. Their formalization is thus not fully specified as far as the rhetorical links introduced by participle clauses are concerned. However, FRAMES, INDEPENDENT RHEMES and ELABORATIONS *are* primarily associated with specific discourse functions, as we have seen. FRAMES, as Bary & Haug (2011: 13) say, are similar to Krifka’s (2008: 269) *stage topics* or *frame setters*, which provide “the frame in which the following expression should be interpreted”. INDEPENDENT RHEMES are like independent clauses from the discourse perspective. ELABORATIONS provide more information (elaborate) on the matrix event, similarly to how the *Elaboration* relation is described in formal theories of discourse representation like Segmented Discourse Representation Theory (SDRT; Asher & Lascarides 2003), on which Bary & Haug’s (2011) discourse analysis is loosely based.

SDRT is a theory of discourse interpretation which, like classic theories in dynamic semantics (Groenendijk & Stokhof 1991; Kamp & Reyle 1993; Kamp, Genabith & Reyle 2011), works under the assumption that ‘the interpretation of an utterance is made relative to the interpretation of the preceding utterance’ (Asher & Lascarides 2003: 39) so that every incoming sentence updates the discourse as it unfolds. SDRT models the semantics-pragmatics interface using rhetorical relations (such as the aforementioned *Elaboration* relation), which are logical links computed on the basis of commonsense reasoning (in the sense of Hobbs, Stickel, Martin & Appel 1993) with non-linguistic information (e.g. domain knowledge) (Asher & Lascarides 2003: 39), which allows accounting for semantically underspecified links between clauses. This captures what was mentioned above, namely that, much like a sequence of two main clauses, the precise semantic relation between participle clauses and their matrix clause (e.g. non-temporal effects such as causality, condition, instrument) must be pragmatically inferred.

An important aspect of SDRT is that its so-called Glue Logic (i.e. the set of inference rules applied to pragmatically enrich the ‘underspecified logical forms’ derived by the syntax and the compositional semantics) is non-monotonic. Namely, it captures *defeasible* inferences, meaning that the discourse representation structure can be modified if other cues emerge later in the discourse licensing different rhetorical relations than previously inferred. One of Bary & Haug’s (2011) reasons for avoiding a fully-fledged SDRT formalization of discourse is that, in SDRT, aspect (which SDRT broadly captures as an *eventive* versus *stative* distinction) is only used as a non-monotonic cue for certain rhetorical relations, whereas, as Bary & Haug (2011) showed, aspect can be represented to a large extent in the compositional semantics of Ancient Greek participles. In SDRT, for example, a sequence of two eventive discourse units may (non-monotonically) trigger a *Narration* relation (24), but not a *Background* relation, which instead requires a stative and an eventive unit (25).⁸

⁸?(α, β) simply indicates an underspecified link between the discourse units α and β .

(24) $Event(e_\beta) \wedge Event(e_\alpha) \wedge ?(\alpha, \beta) > \text{NARRATION}(\alpha, \beta)$

(25) $Event(e_\alpha) \wedge State(e_\beta) \wedge ?(\alpha, \beta) > \text{BACKGROUND}(\alpha, \beta)$
 $State(e_\beta) \wedge Event(e_\alpha) \wedge ?(\alpha, \beta) > \text{BACKGROUND}(\alpha, \beta)$

Unlike SDRT, Bary & Haug (2011) also do not make specific assumptions regarding the form a discourse structure should take, whereas in SDRT discourse structures are represented as graphs, which aptly captures the core concepts of defeasible inference and discourse update. An SDRT discourse representation structure is thus incrementally and hierarchically built and updated with each incoming discourse unit. To take a classic SDRT example, consider (26) from Asher & Lascarides (2003: 8).

- (26) a. John had a great evening last night.
 b. He had a great meal.
 c. He ate salmon.
 d. He devoured cheese.
 e. He then won a dancing competition.

A possible structure for (26) could be built by inferring that (26b) adds granularity to (26a), thus attaching to it via the SDRT relation *Elaboration*. Both (26c) and (26d), then, jointly elaborate on (26b). (26c) and (26d) are therefore linked by the rhetorical relation *Narration* and together form a complex discourse unit attaching to (26b) via *Elaboration*. Finally, (26e) continues the elaboration of (26a) in the same way as (26b) does. (26d) is therefore linked to (26b) via *Narration* and forms a complex discourse unit with it, so that they now *jointly* elaborate on (26a). Figure (1) shows the discourse structure in a simple graph form.

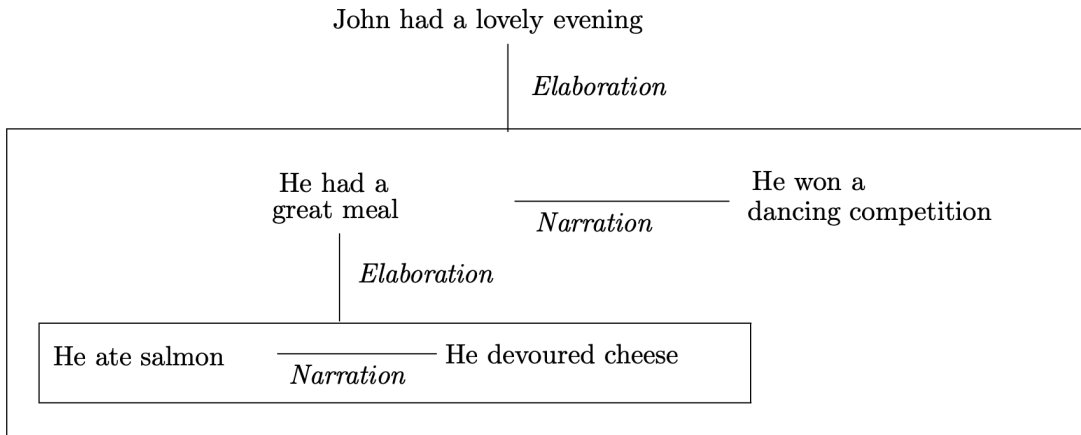


Figure 1: Discourse structure for (26), from Asher & Lascarides (2003: 9).

As we will see in the thesis, both the non-monotonic logic and the graph representation of discourse structure of SDRT are particularly useful to formally capture the discourse functions of participle constructions and *jegda*-clauses in more complex discourses. When several adjuncts co-occur in the same discourse, each adjunct can be interpreted as triggering a particular rhetorical relation at the clause-level, but may then be combined with subsequent units in the discourse structure to form complex discourse units that may attach to the subsequent discourse via a yet different relation.

0.3 What's with *when*?

There are several factors that heavily interact with tense-aspect in the interpretation of finite temporal clauses and their temporal relation to the matrix. These have been investigated specifically with regard to *when*-clauses since at least Steedman (1982) and include, most importantly, the distinction between *events* and *states* (DRT & SDRT), between *durative* and *non-durative* situations (Heinämäki 1972), between *point* and *interval* (Ritchie 1979), between *achievements*, *accomplishments*, *activities* and *states* (Vendler 1967), and, similarly, Comrie's (1976) higher-level categorization of these distinctions into three main Aktionsarten, namely TELIC, ATELIC and STATIVE. The precise differences between some of these categories are often marred by terminological discussions continuing since the sudden plethora of works, often very influential, published in the 1970s on tense and aspect (among others, Verkuyl 1972; Dowty 1972, 1977, 1979; Heinämäki 1972; Bennett 1975; Comrie 1976; Steedman 1977; Ritchie 1979; Isard 1974; Smith 1978) and in the 1980s on event structure, temporal relations, and the anaphoric properties of tense (e.g. Steedman 1982; Partee 1984; Comrie 1985; Nerbonne 1986; Passonneau 1987; Webber 1983, 1987, 1988; Dowty 1986; Hinrichs 1986; Moens & Steedman 1987, all drawing from Reichenbach 1947).

In common to many theories is that initial (i.e. pre-matrix) temporal adverbials, such as *when*-clauses, provide the reference time within which the situation described by the main clause holds temporally, a temporal frame that establishes how the main eventuality is to be interpreted. The main issue has long been explaining how different types of temporal referents introduced by an adverbial (e.g. a punctual event versus a period of time) interact with the temporal referent introduced by the matrix clause, which depends on the temporal adverbial for its interpretation.

It has long been observed (e.g. Kamp & Rohrer 1983; Partee 1984; Hinrichs 1986; Sandström 1993; Bonomi 1997; Sæbø 2012) that when either the temporal adverbial clause or the matrix clause (or both) describe a stative/imperfective eventuality there is a temporal symmetry between the two clauses, whereby if the role is switched the temporal interpretation is preserved, as in (27) from Sæbø (2012: 1422).

- (27) a. *My s Iroj gotovili dokumenty, kogda pozvonil Borja.*
 we with Ira prepared.IPFV documents when called.PFV Boris
 ‘Irina and I were preparing the documents when Boris called.’
- b. *Kogda my s Iroj gotovili dokumenty, pozvonil Borja.*
 when we with Ira prepared.IPFV documents called.PFV Boris
 ‘When Irina and I were preparing the documents Boris called.’

As pointed out by Sæbø (2012: 1422), (27a) and (27b) may have information-structural differences or may involve different discourse relations, but the temporal relation is unchanged. When both eventualities are eventive/perfective, the symmetry does not hold, as (28) shows.

- (28) a. When she died she left a massive doll collection.
 (Sæbø (2012: 1423))
- b. She died when she left a massive doll collection.

Sandström (1993) argues that the temporal referent introduced by a *when*-clause in English must be a bounded referent: either the *phase* of a state (‘when I was eight, I moved to Oslo’), a result state (‘when they had written the letter, they sighed with relief’), or an individual event (‘when he saw him, he smiled at him’). Even so, however, several other factors are at play in determining whether a *when*-clause can be employed or not. Compare the pairs in

- (29) a. She reached him, and said something in his ear.
 b. When she reached him, she said something in his ear.
- (30) a. They chose violence and attacked innocent people.
 b. ?When they chose violence, they attacked innocent people.
- (31) a. He burped loudly and put the glass back on the counter.
 b. ?When he burped loudly he put the glass back on the counter.
- (32) a. ?Sue killed her husband and put poison in his whiskey.
 b. When Sue killed her husband, she put poison in his whiskey.
- (33) a. ?When they built the fifth bridge, a man drowned in the river.
 b. When they were building the fifth bridge, a man drowned in the river.

These examples (from Sandström 1993, except (33) from Ritchie 1979 and (30) from me) show that sequences of two event-type predicates can be felicitously rendered either as a sequence of two main clauses, as *when*-structures, or both, even though the sentences

in each pair (29)-(33) are in fact structurally identical. Parallel structures like (30), if deemed equally acceptable, can hardly be considered to convey the same temporal relation between the events predicated of—(30a) suggests a temporal sequence, where they first chose violence and, as a result of it, they attacked. (30b) instead suggests that *every time* they chose violence they attacked innocent people (with the presupposition that they sometimes do not choose violence). In other instances, as in (32), the main clause is a ‘constitutive part’ (Moens 1987) or ‘subevent’ (Sandström 1993) of the event referent introduced by the *when*-clause, which ultimately shows how mere temporal inclusion of a main-clause event in a given temporal referent cannot be felicitously conveyed by an eventive *when*-clause (as in (33a), a reading that demands the progressive instead (33b).

Even without wandering far from English in genealogical terms, it is clear that these (sometimes very subtle) constraints are not identical cross-linguistically. The scope of each *when*-equivalent varies widely across different languages, which may favour other constructions in situations where *when*-clauses are perhaps preferred or more commonly found in English. This is not necessarily an observation about the lack of particular translational equivalents in either English or any of the other languages, but rather about (also potentially very subtle) differences in the way each language categorizes meanings, and, therefore, in the division of labour among competing constructions. Any *when*-clause in an English version of the Bible may correspond to different constructions in other translations:

(34) Acts 27:38

- a. And **when they had eaten enough**, they lightened the ship
(English Standard Version)
- b. *En nadat hulle met voedsel versadig was, het hulle die skip
and after they with food filled were had they the ship
ligter gemaak
lighter made
(Afrikaans)*⁹
- c. *Quando si furono saziati di cibo, alleggerirono la nave
when REFL had filled of food, they-lightened the ship
(Italian)*¹⁰
- d. *I nasitivši se jela, olakšaše ladju
and having-filled REFL with-food they-lightened ship
(Serbian)*¹¹

⁹1933 [1953] South African Bible Society translation.

¹⁰1998 [1948-53] *Bibbia di Gerusalemme*.

¹¹1868 Daničić-Karadžić translation.

- e. *Luego, satisfechos de la comida, aligeraban la nave*
 later satisfied of the meal they-lightened the ship
 (Spanish)¹²
- f. *După ce s'au săturat, au ușurat corabia*
 After that self-have filled, they-have lightened ship
 (Romanian)¹³

(34) is only one of the possible scenarios, where the English version has a *when*-clause and the other languages show a range of different possible constructions, instead of the *when*-equivalents that do exist in each of those languages. Several other scenarios are obviously possible and are, in fact, widely attested, including those where all the languages in (34) use *when*-equivalents or where all but English use a *when*-equivalent. Even among Indo-European languages, the scope of *when*-equivalents can however be quite different from the one of English *when*-clauses. In Ancient Greek, where non-finite temporal adverbial clauses are much more extensively employed than in English, we find that there is no direct way of employing *when*-equivalents (*hóte/hótan*) to express simple temporal inclusion in the past, for which imperfective participle clauses are generally employed instead, as in (35b) where the English *when*-clause corresponds to a genitive absolute in Greek (and similarly a dative absolute in Old Church Slavonic (35c)).

(35) Matthew 28:11

- a. Now **when they were going**, behold, some of the watch came into the city, and shewed unto the chief priests all the things that were done.
 (King's James Version)
- b. *Poreuomenōn de autōn idou tines tēs*
 go.PTCP.IPFV.GEN.PL PTC 3.PL.GEN behold some the.GEN.SG
koustōdías elthontes eis tēn polin
 watch.GEN.SG come.PTCP.PFV.NOM.PL in the.acc city.acc
apēngeilan tois archiereusin hapanta
 announce.AOR.3.PL the.DAT.PL chief priest.DAT.PL all.N.ACC.PL
ta genomena
 the.N.ACC.PL happen.PTCP.PFV.N.ACC.PL
- c. *idōštama že ima. se edini otō*
 go.PTCP.IPFV.DAT.DU PTC 3.DU.DAT behold some.NOM.PL from
kustodiję prišedōše vō gradō.
 watch.GEN.SG come.PTCP.PFV.NOM.PL in city.ACC.SG
vzvěstišę archiereomō vōsě
 announce.AOR.3.PL chief priest.DAT.PL all.N.ACC.PL
byvōšaa.
 happen.PTCP.PFV.N.ACC.PL

¹²2015 *Reina Valera Actualizada* translation.

¹³1921 *Biblia Cornilescu* translation.

Based on our knowledge of Ancient Greek aspectual semantics, we might expect to find *when*-equivalents in the imperfect to enable mere temporal inclusion of an eventive main clause, similarly to how the progressive in English enables simple temporal inclusion in the *when*-clause, as in (33). Corpus data, however, indicates that such a configuration (i.e. an imperfect *when*-equivalent followed by an eventive–aorist–main clause) is exceedingly rare in the Ancient Greek corpus¹⁴. Both imperfective participles and imperfect *jegda*-clauses with an aorist main clause are instead found in Early Slavic, as the dative absolute in (35c) and the *jegda*-clause in (36c) show. Once again, Greek does not use a *hōte*- or *hōtan*-clause, but this time a nominalized accusative with infinitive (36c).

(36) Luke 11:27

a. And it came to pass, as he was saying these things a woman from the crowd called out

b. *Egeneto* *de en tōi* *legein* *auton*
 happen.AOR.3.SG PTC in the.M.DAT.SG say.INF.PRS 3.SG.M.ACC
tauta *eparasa* *tis*
 this.N.ACC.PL raise.PTCP.PFV.F.NOM.SG certain.M.NOM.SG
phônên *gunê* *ek tou* *okhlou*
 voice.F.ACC.SG woman.F.NOM.SG from the.GEN.SG crowd.GEN.SG
eipen *autōi*
 say.AOR.3.SG 3.SG.M.DAT

c. *bystǔ* *že egda glaaše* *se.*
 happen.AOR.3.SG PTC when say.IMPV.3.SG this.N.ACC.SG
vǔzdvigǔši *glasǔ* *edina* *žena*
 raise.PTCP.PFV.F.NOM.SG voice.ACC some.F.NOM.SG woman.F.NOM.SG
otǔ *naroda* *reče* *emu*
 from crowd.GEN.SG say.AOR.3.SG 3.SG.M.DAT

It is clear that the competition between participle adjuncts and *jegda*-clauses cannot simply be captured in terms of discrete, categorical variables, but it should be modelled as a continuum allowing a degree of overlap and should aim to reveal broader patterns in a probabilistic, rather than a fully deterministic way. This is what token-based typology can help us achieve.

¹⁴Only 1 occurrence of pre-matrix *hōte*-clause with an imperfect verb followed by an aorist main verb was found in PROIEL, and none of *hōtan*. The verb in the imperfect is a form of *eimí* ‘to be, to exist’ (*kai hōte ēn dekaētēs ho paīs, prēgma es hautōn toiōnde genómenon exéphēné min* ‘Now when the boy was ten years old, the truth about him was revealed in some such way as this’, Herodotus 1:114). There are 3 occurrences of post-matrix imperfect *hōte*-clauses, 2 of which are forms of *eimí* and 1 of the verb *záō* ‘to live’. In such cases, we could argue that the order of the eventualities involved is aorist-imperfect, not imperfect-aorist, which is somewhat less rare, although still relatively uncommon (13 occurrences).

0.4 A token-based typological approach to the study of language-internal variation

In our exploration of the functions of participle clauses and *jegda*-clauses we do not have access, *a priori*, to a (set of) comparative concept(s) (Haspelmath 2010) that can be leveraged for language comparison. We have simply observed that there is overlap in the usage of certain language-particular constructions, whose individual functional remit must still be defined—which is, in fact, the very goal of this thesis. As we have seen, there are good reasons to expect that the semantics of Early Slavic participial adjuncts, like their Ancient Greek counterparts, can be defined compositionally to a good extent, particularly with respect to the temporal-aspectual dimension. However, simply by virtue of being juxtaposed forms, there are, especially non-temporal, aspects of their semantics which are left underspecified. This is, as we saw, why Bary & Haug’s (2011) framework is partly integrated with SDRT as a discourse formalization tool, particularly for dealing with more complex discourse structures.

The closest we get to a comparative concept encompassing the constructions under study is what Cristofaro (2013b) refers to as WHEN-clauses¹⁵, defined in functional terms as a construction ‘encoding a temporal relation between two events, such that there is a temporal overlap between the two’ and in which ‘the exact extent of the overlap is unspecified and subject to variation’. Cristofaro’s definition encompasses both *deranked* verb forms (Stassen 1985; Croft 1990; Cristofaro 1998, 2019), that is, verb forms lacking marking of one or more tense, aspect, or mood distinctions compared to independent clauses in the same language (e.g. participles in Early Slavic and Ancient Greek) and *balanced* ones, which include not only finite clauses introduced by a subordinator (e.g. *jegda*-clauses or English *when*-clauses), but also juxtaposed independent clauses whose temporal relation needs to be pragmatically inferred.

We have seen in Section 0.3 that English *when*-clauses, as well as *jegda*-clauses, can express more than just overlap. Cristofaro’s WHEN-clauses may either reflect a misnomer, for lack of better generic alternatives, or their definition may refer to a more general sense of *overlap* as *co-temporality* in the sense of, for example, Behrens, Fabricius-Hansen & Solfjeld (2012). Whichever the case, the scope of English *when*-clauses *does* observably span both that of *jegda*-clauses and that of conjunct participles and absolute constructions, as can be evinced by the counterparts to each of these constructions in different English versions of the Bible, that is, all three OCS constructions can correspond to a *when*-clause in English, as the examples in this Introduction already suggested. As a preliminary umbrella term for the functional continuum in which conjunct participles, absolute constructions and *jegda*-clauses are set, we may therefore stick to ‘WHEN-clauses’.

With these assumptions in mind, we can use a token-based typological approach

¹⁵Small caps is mine. Cristofaro (2013b) uses inverted commas (‘when’) to refer to the concept WHEN to distinguish it from the English token *when*.

(Levshina 2019, 2022) to explore the semantic ground covered by English *when* and *induce* cross-linguistically common semantic dimensions from parallel corpora available for Old Church Slavonic. *Token-based typology* ‘makes generalizations and classifies languages using the tokens of specific linguistic units or structures observed in language use, as approximated by corpora’ (Levshina 2019: 534), allowing us to capture the gradience and overlap between linguistic phenomena like WHEN-clauses and the language-internal variation which is inherent to the very concept of competition. This is in contrast with traditional *type-based* approaches which compare languages based on pre-defined categorical (e.g. *balanced* versus *deranked*; SV versus VS) or ordinal variables (e.g. number of basic colour categories), as the discrete categories in the World Atlas of Language Structures (WALS; Dryer & Haspelmath 2013) indicate.

Like previous experiments in token-based typology that aimed to induce cross-linguistically salient dimensions from a parallel corpus (e.g. Wälchli & Cysouw 2012; Hartmann, Haspelmath & Cysouw 2014; Levshina 2015) probabilistic semantic maps are employed in this thesis as a data-driven method to study conceptual variation within the semantic space of the hypothetical concept WHEN. As we will see, the analysis based on this method provides us with clearer functional boundaries between the Early Slavic constructions under analysis, supporting and complementing the findings of the corpus-based and temporal-semantic analysis of the first part of the thesis, while also setting Early Slavic WHEN-clauses in typological perspective.

0.5 The Early Slavic corpus

Most of the Early Slavic data used in this thesis is extracted from the TOROT¹⁶ treebank. Table 1 shows a breakdown of the texts from which dative absolutes, conjunct participles and *jegda*-clauses were extracted, with an indication of the (high-level) variety of Early Slavic which each text can be ascribed to and the total number of manually annotated tokens.

All TOROT data is lemmatized, morphologically analysed and syntactically annotated, and all layers of annotation have either been manually post-corrected or fully manually annotated.¹⁷

As an offspring of the Old Church Slavonic portion of the PROIEL¹⁸ treebank, TOROT uses the PROIEL annotation scheme, which is modelled on dependency grammar (DG). Like DG, it uses overt elements in the sentences as nodes in a syntactic

¹⁶Tromsø Old Church Slavonic and Old Russian Treebanks (Eckhoff & Berdičevskis 2015). The version used in this thesis corresponds to release 20200116, which can be obtained from <http://torottreebank.github.io>.

¹⁷The Old Church Slavonic *Codex Zographensis* in TOROT was excluded from the dataset because it mostly overlaps in content with the *Codex Marianus*, which has more detailed annotation. The modern Russian portion of the TOROT treebank is also excluded.

¹⁸Pragmatic Resources in Old Indo-European Languages (Haug & Jøhndal 2008), <https://proiel.github.io>.

representation, rather than adding phrasal nodes as in phrase structure grammar. The PROIEL scheme, however, deviates from classic DG schemes (e.g. the one adopted by the Prague Dependency Treebank) in some important ways, most notably by allowing empty nodes (e.g. in cases such as asyndetic coordination and verb ellipsis) and by explicitly signalling secondary dependency (whenever there is a nonfinite construction whose subject is coreferent with an element of the matrix verb, as is the case with conjunct participles) (Haug, Jøhndal, Eckhoff, Welo, Hertenberg & Müth 2009).

The *Codex Marianus*, an Old Church Slavonic tetraevangelion from the PROIEL treebanks, also contains information-structural annotation, including givenness status (e.g. NEW, OLD, ACCESSIBLE) and anaphoric links from anaphors to their antecedents, as well as additional layers of semantic annotation, such as animacy of nominal referents and Aktionsart at the level of verb lemma. Furthermore, it is aligned to the Ancient Greek version of the New Testament in PROIEL¹⁹. Because of its alignment with the Ancient Greek version and its more detailed annotation, the analysis will be carried out separately on the *Codex Marianus* and on the rest of the TOROT treebank. I will refer to the former as a *deeply annotated* treebank, to distinguish it from the rest of the texts from the TOROT treebank, which contain only up to dependency annotation, namely the minimum requirement for a corpus to be classified as a treebank, and which I will refer to as *standard* treebanks.

Note that East Slavic is somewhat overrepresented in the standard treebanks, especially as far as non-translated or ‘original’ Early Slavic texts are concerned, but generally also in the time span covered. The East Slavic subcorpus covers the whole period from Old East Slavic (11th-late 14th century) to Early Middle Russian (15th century) and Late Middle Russian (16-late 17th century), with a total of over 200,000 tokens, whereas the South Slavic one (i.e. South Slavic texts excluding the Old Church Slavonic subcorpus and all classified under the general label of *Church Slavonic*) are represented by only around 13,000 tokens.

Although not meant to actually compensate for this imbalance, a smaller group of South Slavic texts is used in a brief case study at the end of Chapter 2. The case study summarizes the findings from Pedrazzini (2022), who looked at the extent to which corpora with considerably fewer and shallower levels of annotation can be exploited to investigate a discourse-driven syntactic phenomenon, particularly as a means of corroborating results emerging from treebanks with deeper, curated annotation. The small South Slavic corpus was experimentally fully automatically annotated, including

¹⁹The Greek New Testament version in PROIEL is von Tischendorf (1869–1872), which is not the source text of the *Codex Marianus*, as is clear when comparing some of the mismatches between the Ancient Greek and the Old Church Slavonic version. In the course of the relevant chapters, I will point out whether potential mismatches between the two languages according to the PROIEL version correspond to a match when compared to other Ancient Greek versions.

lemmatization, morphological analysis, and dependency parsing.²⁰ A breakdown of these texts can be found in Table 2. The case study from this dataset, which was not manually post-corrected and hence contains many errors, will be presented separately from both deeply-annotated and standard treebanks and will be referred to as *strategically annotated* treebanks.²¹

0.5.1 Identification of the constructions in the corpora

Absolute constructions and conjunct participles are not labelled as such in the annotated treebanks. To identify and extract potential occurrences, a combination of morphosyntactic tags and dependency relations were used. As already mentioned, deeply annotated and standard treebanks from TOROT use the PROIEL dependency tagset, which includes separate labels for clauses with an external argument (e.g. XADV and XCOMP), whereas the strategically annotated treebanks included in the dataset were syntactically annotated with OldSlavNet, which uses the Universal Dependency tagset.

From the TOROT treebanks, conjunct participles were identified by looking for participle forms in any tense-aspect, excluding resultative ones (i.e. *l*-participles), with an XADV relation. The referent of the subject of conjunct participle was identified via the ‘slash’ notation available for all clauses with an external argument to indicate coreference relations within one dependency tree. Most conjunct participles share the subject with their matrix clause, so the XADV node will also be connected to the matrix subject node, if the subject is overt, as in Figure 2a, or to the matrix verb node itself, if the subject is null, via the relation XSUB, to indicate argument sharing, as Figure 2b shows.

The annotation convention in PROIEL is to always analyze the shared arguments in control structures as dependents of the matrix clause. However, following the analysis of control in Greek and Latin participial adjuncts in Haug (2011, 2017), we can consider *some* conjunct participle configurations to be able to head a subject, specifically the leftmost FRAME or INDEPENDENT RHEME conjunct participle linearly preceding the matrix clause. Such conjunct participles can be analyzed as having a control relation of *equality* (rather than *subsumption*; cf. Dalrymple et al. 2019: 553-556) to the matrix clause, albeit with an additional *linearization* constraint on the overt realization of shared arguments. Namely, if overt, the shared argument is always found in the leftmost position relative

²⁰The simple script and dictionary used to lemmatize is available at <https://doi.org/10.6084/m9.figshare.24166254>. Part-of-speech and morphological tags were added using Scherrer, Rabus & Mocken’s (2018) pre-modern Slavic CLSTM tagger Scherrer & Rabus (2019), while for dependency parsing I used Pedrazzini & Eckhoff’s (2021) dependency parser, OldSlavNet.

²¹All strategically annotated texts were normalized by bringing down superscript letters, removing diacritics, substituting non-Unicode characters with their Unicode counterparts, among other preprocessing steps, all of which can be found in the preprocessing Python scripts in the project repository (<https://doi.org/10.6084/m9.figshare.24166254>). *Yagičev Zlatoust*, *Manasses Chronicles*, the *Sbornici* of Vladislav Grammarian, and the *Zografski Sbornik* are from the open-source digital editions of the *Cyrrillomethodiana* project (<https://histdict.uni-sofia.bg/>). The *Bdinski Sbornik* is from the *Obdurodon* project (<http://bdinski.obdurodon.org/>). *Hilandar Typikon* and *Karyes Typikon* are from the *Monumenta Serbica* project (<http://monumentaserbica.branatomic.com/>).

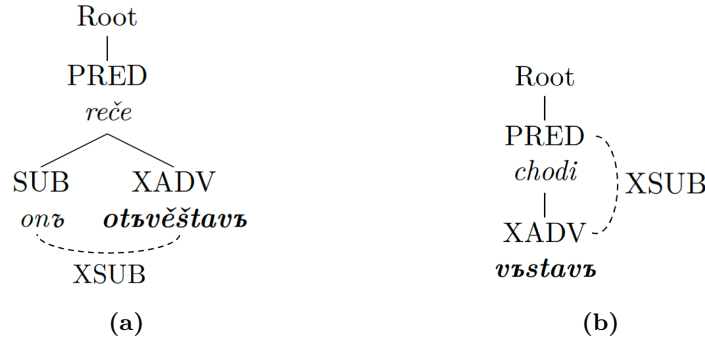


Figure 2: The slash notation in PROIEL/TOROT. (a) is an example of an overt-subject conjunct participle; (b) is an example of a null-subject conjunct participle.

to the matrix clause (‘in the leftmost legitimate host’; Haug 2017: 152). Throughout the thesis, therefore, only the leftmost pre-matrix conjunct participles are considered as potentially able to head their subject and, therefore, to be strictly referred to as *overt-* or *null-*subject. Those whose XSUB slash points to an overt argument (i.e. a node other than the matrix verb itself) will be considered as overt-subject only if the overt argument is found adjacent to or between material belonging to participle clause, which turned out to be the case in the vast majority of cases.²² In some instances, the dependence of the shared argument on the conjunct participle is evidenced by the fact that the shared argument is intermingled with, material that clearly belongs to the participle clause, as in (37), where *isъ* ‘Jesus’, the shared argument, occurs between the conjunct participle *prišedъ* ‘having come’ and another argument of the participle clause, i.e. *vъ domъ petrovъ*. Figures 3 show the difference between the conservative dependency analysis of (37) in PROIEL and its linearization analysis.

- (37) *i* [*prišedъ* *isъ* *vъ domъ* *petrovъ*]
 and come.PTCP.PFV.M.NOM.SG Jesus.NOM in house.ACC Peter’s.ACC
vidě *tōštъ* *ego* *ležęštъ* *ognemъ*
 see.AOR.3.SG mother-in-law.ACC 3.SG.M.GEN lie.PTCP.IPFV.F.ACC fire.INST
žegomъ
 burning.PTCP.IPFV.PAS.F.ACC
 ‘When Jesus came into Peter’s house, he saw Peter’s mother-in-law lying in bed
 with a fever’ (Matthew 8:14)

In some instances, both analyses may be possible, since the overtly-realized shared argument is adjacent to both the sentence-initial conjunct participle and the matrix verb, as in (38).

²²In only 10 cases, the shared overt subject is found after the matrix verb instead.

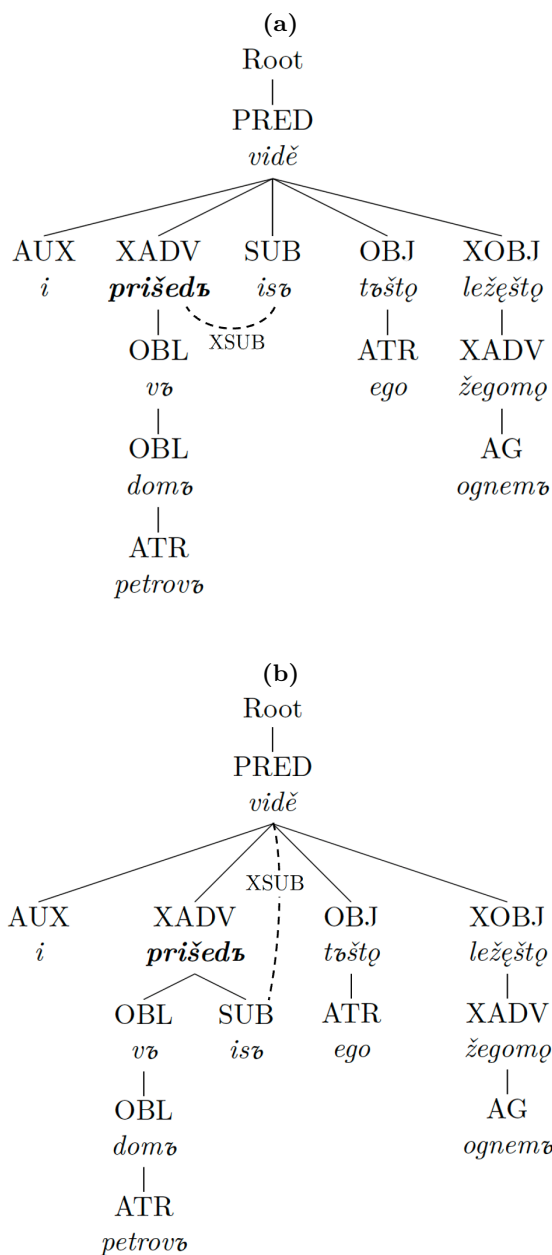


Figure 3: (a) PROIEL analysis of (37) (b) linearization analysis of (37), in which the subject of the conjunct participle, shared with the matrix clause, is analyzed as a dependent of the participle clause.

- (38) *i prišedъše oučenici ego vъzboudišę*
 and come.PTCP.PFV.NOM.PL disciple.NOM.PL 3.SG.M.GEN wake.AOR.3.PL
i
 3.SG.M.ACC

‘The disciples went and woke him’ (Matthew 8:25)

Throughout the analysis, cases like (38) will also be considered as overt-subject conjunct participles, with the caveat in mind that an alternative analysis may be possible.

Dative absolutes were identified by looking for participle forms in the dative and in any tense-aspect, again excluding resultative ones, with an ADV relation. The presence of an overt subject was not specified but identified, if present, by looking for tokens in the dative with a SUB relation and dependent on the participle itself, as in Figure 4.

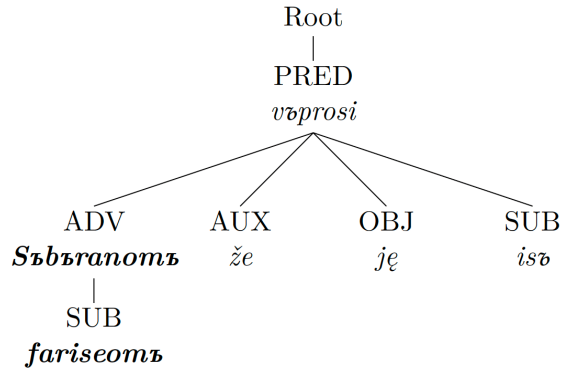


Figure 4: Dependency tree with a dative absolute and its subject

Note that, without specifying overt subjects as a requirement, this query also returns forms that meet the criteria but that are not dative absolutes. Given its relatively small size, the dataset of potential absolutes from deeply annotated treebanks (Codex Marianus) was fully manually corrected to filter out any wrongly extracted occurrences. From standard treebanks, most potential post-matrix null-subject dative participles were cases of nominalized participles, as in (39), and were therefore all excluded without a full manual check.

- (39) *i naprasnijemъ přeloženiја čudo*
 and suddenness.INST.SG transformation.GEN.SG miracle.ACC.SG
sъtvorilъ зъреštiimъ
 make.PTCP.RESULT.M.NOM.SG watch.PTCP.IPFV.M.DAT.PL
 ‘And by the suddenness of the transformation he has performed a miracle for the witnesses’ (*Encomion on the 40 Martyrs of Sebasteia*, Codex Suprasliensis 6: 412)

Potential null-subject dative absolutes preceding the matrix clause were instead manually checked and wrongly extracted occurrences (around 30% of potential pre-matrix null-subject dative absolutes) were filtered out. Only if the lemma was *byti* ‘be’, all occurrences were included regardless of position since a manual check indicated that these are mostly temporal expressions of the kind *pozdě byvōšu* ‘once it got late’, a very common usage of absolutes throughout the corpora.

If the syntactic subject of a participle clause was made of several coordinated constituents (e.g. ‘Jesus and the disciples’), then the one linearly occurring first (e.g. ‘Jesus’ from ‘Jesus and the disciples’) was extracted for the purpose of analysing the properties

of subjects.

The matrix verb of both conjunct participles and dative absolutes was considered to be the direct head of the participle itself if the head node was a verb, as in Figures 2-4. When the direct head of the participle is not a verb, it is normally due to one of two structures. If the direct head is a conjunction, generally multiple (typically two) coordinated participles are dominated by the same matrix verb, in which case PROIEL assigns the same relation tag as the participles themselves to the conjunction node (i.e. XADV or ADV), as in Figure 5. In this case, the head of the conjunction node typically corresponds to the matrix verb itself, bearing the relation tag PRED, in which case the PRED node is identified as the ‘true head’ of the participle clause.

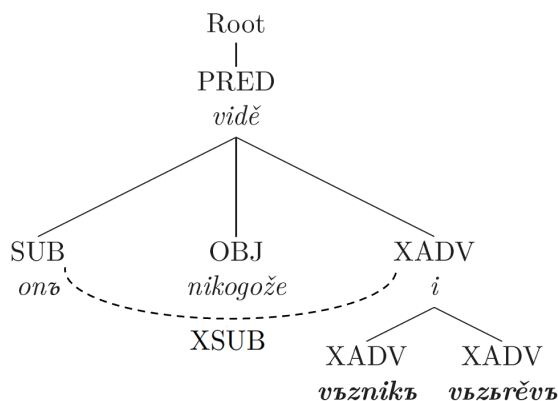


Figure 5: Dependency tree containing two coordinated conjunct participles

If the direct head is an empty node, it generally indicates syntactic independence of some kind, most often overt coordination to the finite clause, which falls into the set of ‘non-canonical’ participle constructions mentioned above. This is because, in order to avoid direct coordination of participle constructions with finite clauses, the annotation convention in PROIEL (for the sake of retrievability) is to treat such examples as elliptical constructions. During their annotation, an empty verbal node is therefore added above the participle clause to stand in for the matrix clause, as Figure 6 shows.

In such cases, for the purpose of the analysis, the matrix verb is identified with the empty node itself, which makes ‘non-canonical’ occurrences of participle constructions (i.e. conjunct participles and dative absolutes that are overtly coordinated to a following independent clause) more easily retrievable.

From OldSlavNet-annotated treebanks, dative absolutes were identified by looking for participles in the dative with an ADVCL dependency relation, conjunct participles by looking for participles in the nominative, also with an ADVCL dependency relation. Both these simple queries also retrieve several forms that are neither conjunct participles nor dative absolutes. In particular, adverbials with an external subject (XADV in PROIEL) and secondary dependency relations (e.g. XSUB in PROIEL) are not currently well-captured by state-of-the-art Early Slavic dependency parsers. This means that we are

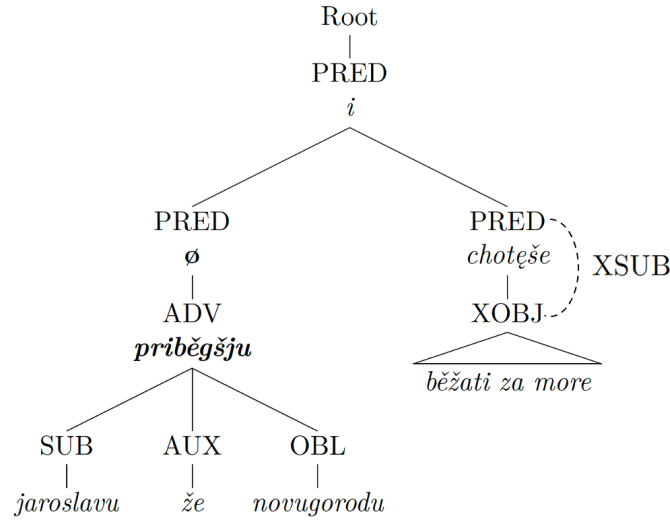


Figure 6: Dependency tree containing a dative absolute coordinated to the main clause

unable to extract the subjects of conjunct participles using fully automated methods, which naturally limits our analysis. To test whether fully automated methods can be used to formulate hypotheses or confirm previous results, the dataset extracted with these queries was not fully post-corrected. However, in addition to the quantitative analysis of strategically annotated treebanks, a case study on one particular strategically annotated text is carried out in Chapter 2, its occurrences checked and closely analysed, thus allowing us not only to evaluate the usefulness of the automatic annotation but also to compare the results of a distant-reading analysis to the observations made by close-reading the occurrences.

Jegda-clauses were first identified by looking for all occurrences of the lemma *jegda* and then extracting the verb dominating the subjunction to represent the *jegda*-clause itself. *Jegda*-clauses used as explicit relative clauses (as in ‘the time *when*...’), were excluded from the dataset by filtering out any *jegda*-clause bearing an ATR (attribute) or APOS (apposition) relation. The same method used to identify the ‘true matrix’ clause of a *jegda*-clause was used. As a rule of thumb, the node immediately dominating the *jegda*-clause verb was taken to be the matrix verb, as in Figure 7, unless it was a conjunction, in which case the node dominating the conjunction was extracted instead, as in Figure 8. As a first approximation, overt subjects were identified by looking for a SUB node headed by the *jegda*-clause verb, unless it was a conjunction, indicating multiple coordinated subjects, in which case the first conjunct was extracted as the subject for the purpose of the analysis in Chapter 3. This method, however, is only possible for *jegda*-clauses with a different subject than the one of the matrix clause. This is because, when co-referential, the subject is analysed as a daughter node of the matrix verb, without

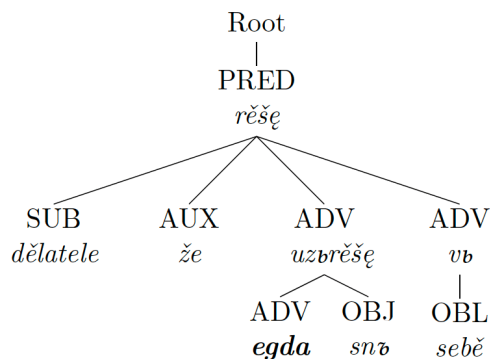


Figure 7: Dependency tree containing a *jegda*-clause.

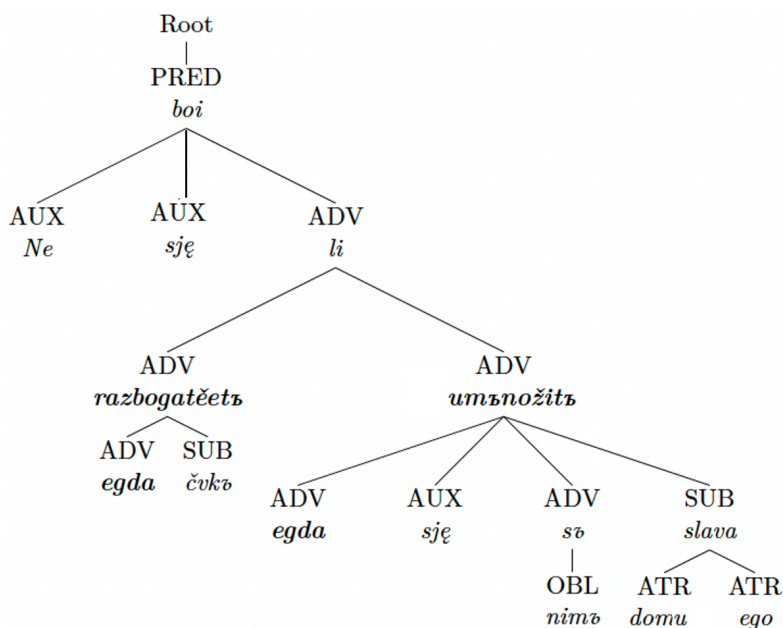


Figure 8: Dependency tree containing two coordinated *jegda*-clauses

indicating secondary dependency relations such as the XSUB one used with conjunct participles. As we will see in Chapter 3, in such cases, we are only able to establish such secondary dependency relation for deeply annotated treebanks, by leveraging the anaphoric link annotation in PROIEL, which also includes null constituents. Namely, if the overt subject of the matrix clause is anaphorically linked to the null subject of the *jegda*-clause, then we can identify the referent of the subject of the *jegda*-clause itself.

0.6 Thesis overview

The thesis is structured in two parts, differing in their scope and methodology.

Part I contains four chapters and looks into the distribution of participial adjuncts

variety	title	tokens
OCS	Codex Suprasliensis	99,150
	Codex Marianus	57,556
	Psalterium Sinaiticum	30,488
	Euchologium Sinaiticum	717
	Kiev Missal	352
CS	<i>Vita Constantini</i>	8,925
	<i>Vita Methodii</i>	3,014
	<i>Life of Petka Tarnovska</i> (Novaković)	859
	The Dobrejsho Gospel	366
OES	<i>Primary Chronicle</i> , Codex Laurentianus	55,088
	<i>Suzdal Chronicle</i> , Codex Laurentianus	23,343
	Uspenskij Sbornik	24,734
	<i>Novgorod First Chronicle</i> , Synodal Manuscript	17,419
	<i>Russkaja pravda</i>	3,903
	<i>Primary Chronicle</i> , Codex Hypatianus	3,512
	<i>The Tale of Igor's Campaign</i>	2,778
	<i>Zadonshchina</i>	2,343
	1229 Smolensk-Riga-Gotland Treaty	1,391
	The <i>Kiev Chronicle</i> , Codex Hypatianus	535
	Statute of Prince Vladimir	476
	Colophon to Mstislav's Gospel book	258
	Colophon to the Ostromir Codex	193
	Mstislav's letter	157
Varlaam's donation charter to Xutyn	140	
EMRus	<i>Life of Sergij of Radonezh</i>	19,966
	<i>The Tale of the Fall of Constantinople</i>	9,116
	Afanasij Nikitin's <i>Journey Beyond Three Seas</i>	6,498
	<i>The Tale of Dracula</i>	2,446
LMRus	Missive from Prince Ivan of Pskov	337
	<i>Domostroj</i>	22,622
	<i>The Life of Avvakum</i> , Pustozerskij Sbornik	22,168
	<i>The Taking of Pskov</i>	2,290
	Materials for the history of the schism	1,818
ONov	Vesti-Kuranty	1,114
	Correspondence of Peter the Great	95
	Birch bark letters	1,864

Table 1: Main corpus breakdown, with an indication of the language variety and the number of annotated tokens. OCS = Old Church Slavonic, OES = Old East Slavic, EMRus = Early Middle Russian, LMRus = Late Middle Russian, ONov = Old Novgorodian.

variety	title	tokens
Middle Bulgarian	<i>Yagičev Zlatoust</i>	64,225
	<i>Manasses Chronicles</i>	41,794
	<i>Sbornici</i> of Vladislav Grammarian	32,040
	Bdinski Sbornik	18,764
	Zografski Sbornik	10,315
Old Serbian	Saint Sava’s <i>Hilandar Typikon</i>	11,982
	Saint Sava’s <i>Karyes Typikon</i>	1,008

Table 2: Breakdown of strategically annotated texts

and *jegda*-clauses in Early Slavic, presenting a quantitative and descriptive corpus analysis and giving an account of their discourse functions and temporal semantics.

More specifically, **Chapters 1** and **2** focus on absolute constructions and conjunct participles and use corpus data to understand the extent to which we can apply previous accounts of the discourse functions of Ancient Greek conjunct participles to Early Slavic participle constructions, including dative absolutes. These first two chapters follow the same overall structure, with Chapter 1 looking at deeply annotated treebanks and Chapter 2 at standard treebanks. Both chapters show that, much like in Ancient Greek, adjunct participle clauses in Early Slavic largely function as ‘grammaticalized discourse relations’ (FRAMES, INDEPENDENT RHEMES, and ELABORATIONS; Bary & Haug 2011) and that their relation to the matrix clause can partly be inferred compositionally through a combination of aspect, position in the sentence, lexical variation, and information-structural properties. The specific treatment of absolute constructions as a subordination strategy in competition with conjunct participles led to the observation that they can be considered as specialized frame-setters and topic-shifters, unlike conjunct participles, whose interpretation is heavily dependent on other variables, such as tense-aspect and position in the sentence, whereby pre-matrix (especially perfective) participles are more likely to be INDEPENDENT RHEMES and post-matrix (especially imperfective) participles are more likely to be ELABORATIONS.

Chapter 3 focuses on *jegda*-clauses. The analysis indicates that there is overall more evident competition with absolutes (which are typically FRAMES) than with conjunct participles. This confirms the intuition that *jegda*, like English *when* in SDRT, can be considered as a trigger for the rhetorical relation *Background*, which is by and large compatible with Bary & Haug’s (2011) treatment of FRAMES. However, unlike participles, the aspect of *jegda*-clauses cannot always be straightforwardly established on the basis of morphology alone, due to the high frequency of present-tense forms, so that their temporal relations cannot be fully established compositionally in the same way as with participles under Bary & Haug’s (2011) framework. **Chapter 4**, therefore, provides a

tentative account of the typical functions of *jegda*-clauses and their temporal relation to the matrix clause by drawing from previous (broadly neo-Kleinian) accounts of the temporal semantics of temporal adverbials. It also formalizes Early Slavic examples containing both *jegda*- and participle clauses on the basis of the SDRT account of the rhetorical relation *Background* and *Narration*.

Part II contains two chapters and looks at adjunct participle clauses and finite temporal subordinates from a typological perspective.

Chapter 5 takes a temporary step back from Early Slavic and generates and analyses a probabilistic semantic map of WHEN-clauses using a massively parallel Bible corpus, to investigate whether the division of labour between participle clauses and finite temporal subordinates in Early Slavic and Ancient Greek corresponds to well-attested cross-linguistic patterns. It demonstrates that ‘null’ constructions (juxtaposed clauses such as participles, converbs, or independent clauses) cluster in particular regions of the semantic map cross-linguistically, clearly indicating that participle clauses are not equally viable as alternatives to any use of WHEN, but carry particular meanings that make them less suitable for some of its functions. It also shows that the distinction found in Ancient Greek between *hóte* and *hótan*, not found in Early Slavic (which only has *jegda*), corresponds to a widely attested cross-linguistic pattern, pointing to a general typological difference between the two languages. **Chapter 6** addresses the issue of understanding the functions of null constructions in the semantic map of WHEN, started in the previous chapter, not only at the level of their competition with *when*-counterparts but also with respect to the competition between different null constructions, which the semantic map of WHEN could not fully address. It provides a comparison between, on the one hand, conjunct participles and absolutes, and, on the other hand, cross-linguistic phenomena known to have similar functions to them, specifically clause chaining, bridging, insubordination, and switch reference.

The clear parallels to these phenomena not only provide independent evidence for the observations made in previous chapters on the division of labour between participle constructions, but they are also able to motivate several of the uses of ‘non-canonical’ participles constructions, particularly dative absolutes, which had otherwise often been written off as ‘aberrations’ by the earlier scholarship on Early Slavic.

Part I

The constructions in the corpus

Chapter 1

Participle constructions in deeply-annotated treebanks

1.1 Introduction

In this chapter and the next, I provide a descriptive and quantitative analysis of conjunct participles and dative absolutes as they appear in the Early Slavic corpus described in the Introduction.¹ I present statistics about their distribution and analyze the constructions across variables which can indirectly capture their discourse functions. The main framework for this analysis is Bary & Haug (2011) and the corpus evidence for their treatment of conjunct participles in Ancient Greek in Haug (2012). Similarly to the latter, the main variables I attempt to capture as a possible indication of the discourse functions of participle constructions are:

- a. the relative order of participle and matrix verb;
- b. aspect distribution;
- c. the properties of the subjects involved;
- d. the lexical variation among participles across different sentence configurations (i.e. pre- and post-matrix participle clauses, different positions of overt subjects relative to the participle).

As outlined in the Introduction, Bary & Haug's (2011) framework is mostly built on conjunct participles. In this chapter, however, I apply the same criteria to the analysis of both conjunct participles and absolute constructions, with the goal of gauging whether:

1. the functions of Early Slavic conjunct participles can be modelled in the same way as in Ancient Greek, namely within a main tripartite distinction between FRAMES, INDEPENDENT RHEMES and ELABORATIONS, by capturing these via indirect corpus evidence as defined above;

¹This chapter and the next expand on material published in Pedrazzini (2022).

2. the functions of absolute constructions can also be modelled as those of conjunct participles, namely as either FRAMES, INDEPENDENT RHEMES and ELABORATIONS, or whether their overall behaviour cannot be predicted based on the same criteria.

The analysis will be structured following the annotation depths available for each of the Early Slavic subcorpora and divided across the current and the next chapter. This chapter exploits data from a *deeply-annotated* treebank, the Old Church Slavonic Codex Marianus, currently the only text in the corpus aligned at the token level with the Greek source text and the only treebank for which information-structural annotation is also available, alongside several other annotation layers on large portions of the text, including Aktionsart and rhetorical relations. Comparison with the Greek parallel is made throughout the section by providing the Greek counterpart and commenting on differences, where relevant. The next chapter will deal with treebanks without information-structural or verb-semantic annotation, which were referred to as *standard treebanks* in the Introduction and which contain mixed data from Old Church Slavonic, different varieties of Church Slavonic, and Old East Slavic. It will also look at *strategically-annotated* treebanks, consisting of fully automatically annotated Middle Bulgarian and Old Serbian texts. This structure also serves, incidentally, as a proof of concept showing the advantages or disadvantages of building small, but deeply-annotated treebanks for historical languages with very high diatopic and diachronic variation (such as Early Slavic), as opposed to larger corpora with shallower annotation, as well as the extent to which methodology and results may differ when using historical treebanks with different annotation depths

Section 1.2 of this chapter comments on the relative order of participle constructions and matrix clause, and the distribution of aspect among participle constructions in different sentence configurations; Section 1.3 analyses the properties of the subjects in participle constructions; Section 1.4 looks into the lexical variation among participle constructions in different sentence configurations; Section 1.5 compares Old Church Slavonic to the Ancient Greek parallel, looking at correspondences and mismatches between the two version in the choice of construction; finally, Section 1.6 takes a look at the rhetorical relation annotation available for the Gospel of Luke.

1.2 Position in the sentence and aspect distribution

As the frequencies in Table 1.1 show, there is a very strong tendency for dative absolutes to occur to the left of the main verb, while that is somewhat less the case for conjunct participles.

The neat predominance of pre-matrix dative absolutes already suggests that the position of dative absolutes in the sentence is connected, in itself, to their functions and supports the widespread view that the unifying function of dative absolutes is a ‘backgrounding’ or ‘stage-setting’ one (Worth 1994: 30; Corin 1995: 259; Collins 2011: 113). As argued by

	pre-matrix	post-matrix	tot.
absolute	92.9% (171)	7.1% (13)	184
conjunct	66.6% (1044)	33.4% (523)	1567

Table 1.1: Relative order of participle and matrix clause in the Codex Marianus

Bary & Haug (2011) and Haug (2012), framing participles should always be found in the leftmost position since they set the stage and provide temporal anchoring for the whole sentence. FRAMES can, in fact, be treated as fronted adjuncts, that is, as information-structurally marked material in topicalized position (Haug 2012: 307). Like *frame setters* in Chafe (1976: 51) and Krifka (2007: 45-48), the concept of FRAME in Bary & Haug (2011) is tightly connected to that of aboutness (or sentence) topic. As Fabricius-Hansen & Haug (2012b) explain, frame setters and aboutness topics are similar in that ‘they both refer to entities [...] that are already established in, or easily inferable from, the preceding discourse’. FRAMES typically refer to a preceding event, as in (1), where the event providing the topic time for the whole sentence is Jesus casting out a demon. In other instances, they refer to bridging events linking to other events by means of ‘motion from one scene to another or perception of some previous action’ (Haug 2012: 299), as in (2), where the temporal anchor for the whole sentence is set by Jesus and the disciples moving – with the presupposition that they have set off from a previous setting.

- (1) a. *i izgnanu* *běsu* *progla*
and cast.out.PTCP.PFV.PASS.M.DAT.SG demon.M.DAT.SG speak.AOR.3.SG
němy
mute.M.NOM.SG
- b. *kai eklethentos* *tou* *daimoniou*
and cast.out.PTCP.PFV.PASS.M.GEN.SG the.M.GEN.SG demon.M.GEN.SG
elalēsen *ho* *kōphos*
speak.AOR.3.SG the.M.NOM.SG mute.M.NOM.SG
‘And when the demon was cast out, the mute spoke’ (Matthew 9.33, 38591, 15005)
- (2) a. *i sxhodeštem̃* *im̃* *s̃* *gory*
and come.down.PTCP.IPFV.DAT.PL he.DAT.PL from mountain.GEN.SG
zapovědě *im̃* *is̃* *gl̃e*
command.AOR.3.SG he.DAT.PL Jesus.NOM say.PTCP.IPFV.M.NOM.SG
- b. *kai katabainontōn* *autōn* *ek* *tou*
and come.down.PTCP.IPFV.GEN.PL he.GEN.PL from the.N.GEN.SG
orous *eneteilato* *autois* *ho* *Iēsous*
mountain.N.GEN.SG instruct.AOR.3.SG he.DAT.PL the Jesus.NOM

	imperfectives	perfectives
pre-matrix	59.1% (101)	40.9% (70)
post-matrix	53.8% (7)	46.2% (6)
tot.	58.7% (108)	41.3% (76)

Table 1.2: Dative absolutes in the Codex Marianus: aspect distribution by position relative to the matrix clause (row percentage)

	imperfectives	perfectives
pre-matrix	13.1% (121)	86.9% (806)
post-matrix	80.3% (216)	19.7% (53)
tot.	28.2% (337)	71.8% (859)

Table 1.3: Conjunct participles in the Codex Marianus: aspect distribution by position relative to the matrix clause (row percentage)

perfectives to occur in pre-matrix position ($p = 0.01$). This is, naturally, an incomplete picture based on few occurrences² and what really is noticeable is the difference in number between pre- and post-matrix occurrences.

As Table 1.3 shows, among conjunct participles, the association between aspect and position is statistically significant, with a strong effect size ($\phi = 0.62$, Cramér's V): $\chi^2(1) = 462.54$, $p < 0.01$, with much higher odds for perfectives to occur in pre-matrix than post-matrix position and for imperfectives to occur in post-matrix than pre-matrix position (odds ratio: 27.15). This is in line with Haug's (2012) Ancient Greek data, showing that participles to the right of the matrix are more likely to be ELABORATIONS while those to the left are more typically INDEPENDENT RHEMES. ELABORATIONS normally result in a 'complex rheme', where the event described by the elaborating participle adds granularity to the lexical meaning of the matrix verb and often receives a manner or means interpretation, as the example in (5). INDEPENDENT RHEMES are instead similar to independent clauses from the discourse perspective, as in (6).

- (5) a. *i vōstrōzaachō učēnici ego klasy.*
 and pluck.IMPF.3.PL disciple.NOM.PL 3.SG.GEN head.of.grain.ACC.PL
*i ěděachō **istirajōšte** rōkama*
 and eat.IMPF.3.PL rub.PTCP.IPFV.NOM.PL hand.INS.DU
- b. *kai etillon hoi mathētai autou tous*
 and pluck.IMPF.3.PL the disciple.NOM.PL 3.SG.GEN the.ACC.PL
*stachyas kai ēsthion **psōchontes***
 head.of.grain.ACC.PL and eat.IMPF.3.PL rub.PTCP.IPFV.NOM.PL

²The binomial test can most intuitively be seen as a test for the probability of successes. In the case of dative absolutes in this dataset, out of 171 attempts (i.e. the total number of occurrences), 101 were successful, 70 were not. However, the relatively small number of overall occurrences means that, in this case, only 5 occurrences would have changed the test result from significant to non-significant, hence why one should be wary of it for now.

tais chersin
the.F.DAT.PL hand.F.DAT.PL

‘And his disciples plucked the heads of grain and ate them, rubbing them in their hands’ (Luke 6:1)

- (6) a. *azъ prišedъ iscěljo i*
1.SG.NOM come.PTCP.PFV.M.NOM.SG heal.PRS.1.SG 3.SG.M.ACC
- b. *egō elthōn therapeusō auton*
1.SG.NOM come.PTCP.PFV.M.NOM.SG heal.FUT.1.SG 3.SG.M.ACC
- ‘I will come and heal him’ (Matthew 8:7)

As pointed out by Haug (2012: 311), while the perfective is generally the dominant aspect in narrative contexts, this is particularly true of the simple narrative style of the New Testament. The preponderance of perfective INDEPENDENT RHEMES is thus as unsurprising as that of perfective main verbs. Stacked perfective INDEPENDENT RHEMES always induce narrative progression (Bary & Haug 2011: 15-16), as clearly transpires from (7), for example. Although much less frequent, even when imperfective, they are equivalent to main verbs from the information-structural perspective, which is often reflected in their English translation, as shown in (8).

- (7) a. *i abie tekъ edinъ otъ*
and immediately run.PTCP.PFV.M.NOM.SG one.M.NOM.SG from
nichъ. i priemъ gobō.
3.PL.GEN and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG
isplъnъ ocъta. i vъznezъ
fill.PTCP.PFV.M.NOM.SG vinegar.GEN.SG and put.PTCP.PFV.M.NOM.SG
na trъstъ. napaěše i
on reed.ACC.SG give.to.drink.IMP.3.SG 3.SG.ACC
- b. *kai eutheōs dramōn heis ex auton*
and immediately run.PTCP.PFV.M.NOM.SG one.NOM.SG from 3.PL.GEN
kai labōn spongon plēsas
and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG fill.PTCP.PFV.M.NOM.SG
te oxous kai peritheis kalamō
with vinegar.GEN.SG and put.PTCP.PFV.M.NOM.SG reed.DAT.SG
epotizen auton
give.to.drink.IMP.3.SG 3.SG.ACC
- ‘Immediately one of them ran and took a sponge, filled it with sour wine and put it on a reed, and offered it to him to drink’ (Matthew 27:48)

- (8) a. *nъ i psi prichodešte oblizaachō*
but also dog.NOM.PL come.PTCP.IPFV.NOM.PL lick.IMP.3.PL
gnoi ego
wound.ACC.PL 3.SG.GEN

- b. *alla kai hoi kynes erchomenoi epeleichon*
 but also the dog.NOM.PL come.PTCP.IPFV.NOM.PL lick.IMPV.3.PL
ta helkē autou
 the.ACC.PL wound.ACC.PL 3.SG.GEN
 ‘Even the dogs would come and lick his sores’ (Luke 16:21)

Given the significant association among conjunct participles between the pre-matrix position and the perfective aspect, it seems likely that the *typical* function of pre-matrix conjunct participles is as INDEPENDENT RHEMES. But as 3 already showed, the sentence-initial position can be ambiguous between a FRAME and an INDEPENDENT RHEME interpretation, as Bary & Haug (2011) already argued for Ancient Greek conjunct participles. In (9), the ambiguity of the sentence-initial conjunct participle between a frame and an independent rheme function is, to some extent, also reflected in the different English translations.

- (9) a. *i postavšē jō po srědě. glašę*
 and place.PTCP.PFV.NOM.PL 3.SG.F.ACC in midst.DAT say.AOR.3.PL
emu.
 3.SG.M.DAT
- b. *kai stēsantes autēn en mesōi legousin*
 and set up.PTCP.PFV.NOM.PL 3.SG.F.ACC in midst.DAT say.PRS.3.PL
autōi
 3.SG.M.DAT
 ‘They made her stand before the group and said to Jesus’ (New International Version)
 ‘When they had set her in the midst, they say unto him’ (King James Bible)
 ‘Placing her in the midst, they said to him’ (English Standard Version)
 (John 8:3–4)

As we will see, subject realization and the position of overt subjects relative to the participle (specifically, the VS configuration) have some part in their interpretation as FRAMES. Overall, however, sentence-initial position conjunct participles will inevitably retain a certain degree of ambiguity, particularly with a null subject.

1.2.1 A remark about *bystz*-clauses

Before looking at other variables, it is worthwhile to comment briefly on a small, but coherent group of occurrences of dative absolutes in a construction introduced by *bystz* ‘it came to pass’, the usage of which in Old Church Slavonic can clearly be linked to a relatively widespread and well-studied construction attested in several pre-modern Indo-European languages, consisting, in its basic form, of

- (10) IT CAME TO PASS[---]_α[---]_β.

where α is filled by an adverbial clause and β by the main predication. Works on discourse marking (e.g. Traugott & Dasher 2002a: 52; Brinton 1996: 336) have treated ‘happen, come to pass’ in such contexts on a par with discourse connectives such as *in fact*, *well*, *then* or *so*, which hardly contribute to the truth-conditional meaning of a proposition, but rather mark the speaker’s stance towards the sequential relationship between discourse units. In particular, the discourse functions of Old English *gelamp*-clauses (the counterpart to Slavonic *bystǫ*) and the syntactic status of its constituents, which have been thoroughly analyzed by Brinton (1996: §5-6), strongly support the intuition that Old Church Slavonic *bystǫ*, like its Greek counterpart *egeneto* (and analogous forms in Gospel translations in other historical languages, such as Gothic *warþ* or Latin *factum est*), has a clear formulaic function. Brinton (1996: 134)³ argues that these are typically made of an episode-initiating, backgrounded clause which gives the time, place or circumstances for a (linearly) following discourse-foregrounded clause, which introduces the narrative proper. Brinton (1996: 142) states that ‘the discourse strategy [in *gelamp*-clauses] can be better understood as framing, as well as focusing, where the main clause [‘it came to pass’] and adverbial provide a frame which directs the reader’s focus to the events denoted in the complement clause’. Similar considerations are made by Hogeterp & Denaux (2018: 334-336) on *egeneto* formulas in New Testament Greek, where *egeneto* constructions are considered to be ‘important linguistic markers of narrative turns’, with the ‘structural function to introduce a new literary section or pericope’.

It is interesting to note that several of the mismatches between Old Church Slavonic and the Greek original (i.e. where Old Church Slavonic uses an absolute construction and Greek does not) occur in *bystǫ*-constructions. This suggests that the particular meaning of the α slot in this construction may correspond particularly well to the overall function of dative absolutes in Old Church Slavonic. (11)-(14) are some examples of dative absolutes in *bystǫ*-constructions.

- (11) a. *i bystǫ idǫstemъ imъ. ištistišę*
 and happen.AOR.3.SG go.PTCP.IPFV.DAT.PL 3.PL.DAT cleanse.AOR.3.PL
sę
 REFL

- b. *kai egeneto en tōi hupagein autous*
 and happen.AOR.3.SG in the.N.DAT.SG go.PRS.INF 3.PL.ACC
ekatharisthēsan
 cleanse.AOR.PASS.3.PL

‘And it came to pass that, as they went, they were cleansed’ (Luke 17:14)

³Note that Brinton (1996) adduces plenty of support from previous literature. In particular, Traugott’s extensive work on grammaticalization provides further solidity to Brinton’s claims. Traugott (e.g. 1982; [with König] 1991; [with Dasher] 2002a; 2003; [with Hopper] 2003; 2004; 2008) has dealt with a large variety of utterance-to-discourse-marker grammaticalization processes, which have helped tackle the status of ‘happen’ in Old English *gelamp*-constructions in Brinton (1996). For an in-depth survey of the most influential studies on discourse markers, including Traugott’s seminal works, see Maschler & Schiffrin (2015).

- (12) a. *bystō* *že* *běsu*
 happen.AOR.3.SG PTC demon.M.DAT.SG
izgъnanu *proglā* *němy*
 cast.out.PTCP.PFV.PASS.M.DAT.SG speak.AOR.3.SG mute.M.NOM.SG
- b. *egeneto* *de* *tou* *daimoniou*
 happen.AOR.3.SG PTC the.M.GEN.SG demon.M.GEN.SG
exelthontos *elalēsen* *ho*
 cast.out.PTCP.PFV.PASS.M.GEN.SG speak.AOR.3.SG the.M.NOM.SG
kōphos
 mute.M.NOM.SG
 ‘It came to pass that, as the demon had been cast out, the mute man spoke’
 (Luke 11:14)
- (13) a. *bystō* *že* *chodeštemъ* *imō* *i*
 happen.AOR.3.SG PTC walk.PTCP.IPFV.DAT.PL he.DAT.PL and
tō *vōnide* *isō* *vō* *vesō*
 that.M.NOM.SG enter.AOR.3.SG Jesus.NOM in village.F.ACC.SG
edīnō
 certain.F.ACC.SG
- b. *egeneto* *de* *en tōi* *poreuesthai* *autous*
 happen.AOR.3.SG PTC in the.N.DAT.SG proceed.PRS.INF he.ACC.PL
kai autos eisēlthen eis kōmēn tina
 and he.NOM.SG enter.AOR.3.SG into village.F.ACC.SG certain.F.ACC.SG
 ‘It came to pass that, as they were walking, Jesus entered a certain village’
 (Luke 10:38)
- (14) a. *i* *bystō* *idoštju* *emu* *vō*
 and happen.AOR.3.SG go.PTCP.IPFV.M.DAT.SG he.DAT.SG to
imō. *i* *tō* *prochoždaaše* *meždju*
 Jerusalem.ACC and that.M.NOM.SG walk.through.IMP.3.SG between
samariejō *i* *galilējō*
 Samaria.F.INS.SG and Galilee.F.INS.SG
- b. *kai egeneto* *en tōi* *poreuesthai* *eis*
 and happen.AOR.3.SG in the.N.DAT.SG proceed.PRS.INF to
Hierousalēm kai autos diērchetō dia
 Jerusalem.ACC and he.NOM.SG walk.through.IMP.3.SG through
meson Samarias kai Galilaias.
 midst.N.ACC.SG Samaria.GEN.SG and Galilee.GEN.SG
 ‘And it came to pass, as he went to Jerusalem, that he passed through the
 midst of Samaria and Galilee’ (Luke 17:11)

While (12a) translates a Greek genitive absolute (12b), the dative absolutes in (11a) and (13a)-(14a) correspond to a nominalized accusative with infinitive introduced by *en* ‘in, during’. The second conjunction in (13)-(14), apparently introducing the main clause,

can in fact be seen as a calque from the Greek, which, in turn, is likely a calque from Hebrew: the construction (*kai*) *egeneto* (*de*) ‘(and) it came to pass that’ itself, opening a narrative section, has been recognized as a syntactic Semitism characteristic of Luke’s Gospel (Hogeterp & Denaux 2018: §6.1; Janse 2007: 657). Hogeterp & Denaux (2018: 326) specifically single out the configuration (*kai*) *egeneto* *de* + temporal expression + *kai* + finite verb (i.e. with a second conjunction seemingly introducing the main clause) as the one standing out the most for its ‘Hebraistic quality’ among the possible *egeneto* formulas, where the second *kai* would reflect the temporal use of Biblical Hebrew *waw* (–ו)⁴ in constructions of the type ‘when ... then’ (Waltke & O’Connor 1990: 436). This is claimed by the authors by comparing the occurrences of the same construction in Septuagint Greek with their counterparts in the Masoretic Text, as well as with the usage of *en toi* + infinitive followed by *kai* in Ancient Greek, which appears to be largely at odds with the one in Koine Greek⁵.

Whichever syntactic analyses for (11)-(14) one decides to endorse, the function of the dative absolute in *bystō*-clauses is thus consistent throughout – namely, to provide the setting, or frame, for the eventuality expressed by the following finite clause.

1.2.2 Summary

In this section, we observed that the vast majority of dative absolutes occur before the matrix clause, suggesting that the position of absolutes in the sentence is, in itself, likely to be associated with their primary function. This can be identified as a stage-setting or backgrounding one, as previous studies on Early Slavic dative absolutes had already observed, or, under Bary & Haug’s (2011) terminology, as FRAME participles. Conjunct participles, similarly to their Ancient Greek counterparts (as per the corpus evidence in Haug 2012), showed a significant association between the position of the participle relative to the matrix clause and the aspect of the participle. Pre-matrix conjunct participles are much more likely to be perfective, whereas post-matrix conjunct participles are much more likely to be imperfective. This is in line with Bary & Haug’s (2011) analysis, whereby conjunct participles as INDEPENDENT RHEMES are typically pre-matrix and perfective, whereas as ELABORATIONS they are typically imperfective and generally occur after the matrix verb. We have also seen that, sentence-initially, conjunct participles, even when perfective, can also function as FRAMES, either clearly so, as in (3), or as a possible interpretation in ambiguous cases such as (9), where they could also be interpreted as INDEPENDENT RHEMES.

Finally, I made a digression on *bystō*-clauses (‘it came to pass that...’), a construction

⁴Biblical Hebrew *waw* has been described as an ‘all-purpose connector’ (Steiner 2000), which ‘places propositions or clauses one after another, without indicating the hierarchical relation between them’ and ‘frequently joins logically subordinate clauses to a main clause’ (Waltke & O’Connor 1990: 649).

⁵Hogeterp & Denaux (2018: 436) find that ‘in the relatively few cases where ἐν τῷ with infinitive has a temporal significance in the voluminous literary Greek works of Thucydides, Polybius, Diodorus Siculus, and Longus, it is never syndetically followed by an apodotic καὶ in their works’.

occurring rather frequently in the Gospels which has widely studied counterparts in other pre-modern Indo-European languages. The construction is made of functionally relatively fixed slots, and we have seen that dative absolutes often occur in what has been analysed as the episode-initiating, backgrounded predication which precedes a following foregrounded, main eventuality. This observation independently supports the analysis of dative absolutes as prototypical FRAMES, particularly since their usage in *bysto*-constructions occurs independently of the Greek sources.

1.3 Subjects

Examples like (15) and (16) are discursively very similar. The main difference appears to be syntactic: the subject of the conjunct participle in (16) is co-indexed with that of the matrix clause, whereas the subject of the dative absolutes in (15) is not.

- (15) a. *i vŕšedŕšju emu vŕ domŕ.*
 and enter.PTCP.PFV.M.DAT.SG 3.SG.M.DAT in house.ACC.SG
uĉenici ego vŕprašachŕ i edinogo.
 disciple.NOM.PL 3.SG.M.GEN ask.IMPF.3.PL 3.SG.M.ACC one.M.GEN.SG
ĉko
 that

- b. *kai eiselthontos autou eis oikon hoi mathētai autou kat' idian epērōtōn auton*
 and enter.PTCP.PFV.M.GEN.SG 3.SG.M.GEN in house.ACC.SG
 the.M.NOM.PL disciple.M.NOM.PL 3.SG.M.GEN in private.ACC
 ask.IMPF.3.PL 3.SG.M.ACC
 ‘And when he had entered the house, his disciples began to ask him privately’ (Mark 9:28)

- (16) a. *i vŕšedŕ vŕ crkvŕ. naĉētŕ*
 and enter.PTCP.PFV.M.NOM.SG in temple.ACC.SG start.AOR.3.SG
izgoniti prodajōštejē vŕ nei
 drive out.INF sell.PTCP.IPFV.ACC.PL in 3.SG.F.LOC
- b. *Kai eiselthōn eis to hieron ērksato ekballein tous pōlountas*
 and enter.PTCP.PFV.M.NOM.SG in the.ACC.SG temple.ACC.SG
 rule.AOR.3.SG drive out.INF.PRS the.M.ACC.PL sell.PTCP.IPFV.ACC.PL
 ‘When Jesus entered the temple, he began to drive out those who were selling there’ (Luke 19:45)

It has often been observed, however, that in both Early Slavic (Collins 2004, 2011) and Ancient Greek (Fuller 2006; Haug 2011) subject co-indexation between absolutes and matrix clauses is, in fact, attested in the Gospels (17), as well as in Classical

Greek (18) and Early Slavic texts (19), despite traditional descriptive grammars treating them as ‘exceptions’.

- (17) *i bystō idōštemъ imъ. ištistišę*
 and happen.AOR.3.SG go.PTCP.IPFV.DAT.PL he.DAT.PL cleanse.AOR.3.PL
sę
 REFL

‘And it came to pass that, as they went, they were cleansed’ (Luke 17:14)

- (18) *Kroise anartēmenou*
 Croesus.M.VOC.SG to be prepared.PTCP.PRF.PASS.M.GEN.SG
seu andros basileos khrēsta
 yourself.2.M.GEN.SG man.M.GEN.SG king.M.GEN.SG useful.PL.N.ACC
erga kai epea poieein aiteo
 work.PL.N.ACC and word.PL.N.ACC make.PRS.INF ask.2.SG.PRS.IMP.MID
dosin hēntina bouleai toi
 giving.SG.F.ACC which.SG.F.ACC wish.2.SG.PRS.MID you.2.SG.DAT
genesthai parautika
 happen.PFV.PST.INF.MID directly

‘Croesus, now that you, a king, are determined to act and to speak with integrity, ask me directly for whatever favor you like.’ (Herodotus, *Histories* 1.90.1; translation by Godley 1920.)

- (19) *vamъ zloslavnom suštem ot kudu imate*
 you.DAT.PL heretic.DAT.PL be.PTCP.IPFV.DAT.PL from where will.PRS.2.PL
pokazati jako sergie onъ pravoslavnъ
 show.INF that Sergius.NOM that.NOM.M.SG orthodox.NOM.M.SG
bě i cha radi postrada
 be.IMP.2.SG and Christ.GEN for suffer.AOR.3.SG

‘Since you are heretics, how will you show that Sergius was an Orthodox and suffered for Christ?’ (Euthymius of Tarnovo, *Life of Hilarion of Meglin*, Sbornici of Vladislav Grammarian 171r)

Several partial explanations have been given as to why the authors or translators would not use an agreeing participle where syntactically possible: a change in the subject’s referent’s semantic role between adverbial and main clause (e.g. Collins 2004, 2011), or of the ‘underlying subject’ even if not of the ‘formal subject’ (Levinsohn 2000); the possibility that an absolute ‘brings out the participial member with more prominence and force’ than a conjunct participle (Winer 2001 [1882]); the goal of making the participial clause more prominent (Goodwin 1997 [1930]); or even the scribal late realization that they could have used an agreeing participle instead (Winer 2001 [1882]). While the syntactic factor plays a role, there are broader information-structural and discourse factors involved, which neither morphosyntactic nor semantic-role explanations can single-handedly capture. It may therefore be useful to look into the properties of the subjects of dative absolutes and conjunct participles beyond subject co-referentiality.

1.3.1 Realization and position

As discussed in the Introduction, the annotation convention in PROIEL is to always analyze the argument shared by conjunct participles and their matrix as dependents of the matrix clause. As Haug (2011, 2017) argues, however, conjunct participles in some configurations can be considered heads of the shared argument, specifically the leftmost FRAME or INDEPENDENT RHEME conjunct participle linearly preceding the matrix clause. Only the leftmost pre-matrix conjunct participles, therefore, were here considered as potentially able to head their subject and, therefore, to be strictly referred to as *overt*- or *null*-subject. Dative absolutes, on the other hand, have their own internal subject, regardless of configuration. A fair comparison between dative absolutes and conjunct participles, then, may be made between the constructions in sentence-initial position, which is also where the clearest functional overlap occurs, since they can both work as FRAMES. Table 1.4 shows the frequency of overt and potential null subjects in sentence-initial conjunct participles and dative absolutes.

	overt	null
conjunct (sentence-initial)	60.5% (455)	39.5% (297)
absolute (sentence-initial)	85.4% (123)	14.6% (21)

Table 1.4: Frequency of overt and null subjects among sentence-initial conjunct participles and dative absolutes in the Codex Marianus

The vast majority of sentence-initial dative absolutes have an overt subject. All the 21 dative absolute occurrences classified as ‘null’ in Table 1.4 are, in fact, expressions of the type *pozdě byvøšu* ‘when it got late’, *byvøšju že dñni* ‘when the day came’ or *večrou byvøšju* ‘when the evening came’. In PROIEL/TOROT, these are all analyzed as impersonal temporal constructions consisting of a copula and a predicative adverb or noun, the latter agreeing with the participle. Although one could argue for a finer-grained distinction between personal and impersonal temporal constructions among these occurrences,⁶ from the discourse perspective, these can all still be analyzed as generic frame-setters.

Among sentence-initial conjunct participles, there are as many as 39.5% potential null subjects. Intuitively, this reflects the fact that conjunct participles in sentence-initial position (which can be either FRAMES or INDEPENDENT RHEMES) are functionally more diverse than dative absolutes (which are instead typically FRAMES).

As Haug (2012: 316) observes, ‘sentences that shift the subject are more likely to have a frame to anchor it to the previous discourse, whereas this is not as necessary in

⁶There is arguably a difference between, for example, *pozdě byvøšu* ‘when it got late’ and *byvøšju že dñni* ‘when the day came’. In the latter, *dñni* ‘day’ should probably be analysed as the subject of the absolute. Still, manually disambiguating personal from impersonal constructions would not change the overall observation that overt subjects are much more frequent than null subjects among sentence-initial dative absolutes.

continuous sequences of events with the same agent'. FRAMES are particularly useful to reinstate previously mentioned referents that had no longer been in focus in the previous discourse, making them very likely to overtly realize their subject but also, specifically, to have the subject following the participle itself (Haug 2012: 316-320). Following Haug (2012), then, overt-subject conjunct participles in the VS configuration should be more likely to function as FRAMES than INDEPENDENT RHEMES. It is not difficult to find examples where this is clearly the case, as in the VS and SV conjunct participles in (20) and (21), functioning as a FRAME and an INDEPENDENT RHEME, respectively (as the English translation also suggests).

- (20) a. *Slyšavъ* *že junoša slovo*
 hear.PTCP.PFV.M.NOM.SG PTC young man.NOM.SG word.ACC.SG
otide *skrōbe*
 go away.AOR.3.SG mourn.PTCP.IPFV.M.NOM.SG
- b. *akousas* *de ho neaniskos*
 hear.PTCP.PFV.M.NOM.SG PTC the.M.NOM.SG young man.M.NOM.SG
apēlthen *lupoumenos*
 depart.AOR.3.SG offend.PTCP.IPFV.PAS.M.NOM.SG
 But when the young man heard this, he went away sad' (Matthew 19:22)
- (21) a. *i se prokažēnъ pristoppъ*
 and PTC leper.M.NOM.SG move-forward.PTCP.PFV.M.NOM.SG
klaněše se emu glē
 worship.IMP.F.3.SG REFL him.DAT say.PTCP.IPFV.M.NOM.SG
- b. *kai idou lepros proselthōn*
 and PTC leper.M.NOM.SG move-forward.PTCP.PFV.M.NOM.SG
prosekunei autōi legōn
 worship.IMP.F.3.SG him.DAT say.PTCP.IPFV.M.NOM.SG
 'A man with leprosy came and knelt before him and said' (Matthew 8:2)

As Table 1.5 shows, the VS configuration is much more frequent than the SV one among sentence-initial dative absolutes, once again supporting their interpretation as typical FRAMES. Among sentence-initial conjunct participles, the split between the two configurations is instead more even.

	SV	VS
conjunct (sentence-initial)	45.7% (208)	54.3% (247)
absolute (sentence-initial)	21.9% (27)	78.1% (96)

Table 1.5: Position of overt subjects relative to sentence-initial conjunct participles and dative absolutes in the Codex Marianus

If the VS configuration is associated with the reinstatement of older referents (which framing participles help contextualize in the new discourse), the expectation is that,

more than other functions, FRAMES should involve subject referents whose immediately previous mention in the discourse is generally more distant in the preceding discourse. We can look at the anaphoric distance between the (overt or null) subject of a participle construction and its antecedent using the anaphoric-link annotation from nominal referents to their antecedents in the Greek New Testament in PROIEL. This level of annotation also includes links between null arguments, so that we can check when the last mention of a referent was even if either the anaphor or the antecedent is prodropped.

As Figure 1.1 shows, sentence-initial conjunct participles with an overt subject following the participle itself have, on average, a significantly more distant antecedent (in number of tokens) than those whose subject precedes the participle.

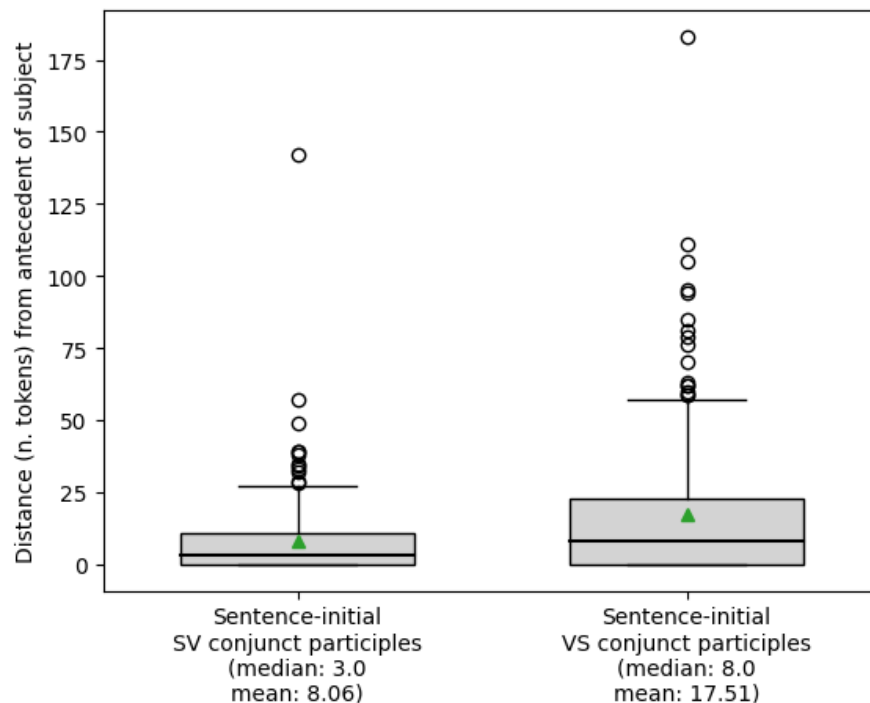


Figure 1.1: Average anaphoric distance between the subject and immediate antecedent in sentence-initial SV and VS conjunct participle. Test results: Welch’s t -test: -5.06 , p -value < 0.01 ; one-tailed Mann–Whitney U -test: 17526.5 , p -value < 0.01 .

This matches the observation that sentence-initial conjunct participles with their subject following the participle itself are more likely to function as FRAMES than those with their subject preceding the participle, since the VS configuration is associated with the reinstatement of old or inactive referents.

When we compare sentence-initial dative absolutes and conjunct participles, including both null- and overt-subject, we find that the difference is not significant (despite dative absolutes having a higher mean), as Figure 1.2 shows.

This result is somewhat surprising, since sentence-initial conjunct participles often have a null subject, whereas dative absolutes do not, and null subjects may be expected

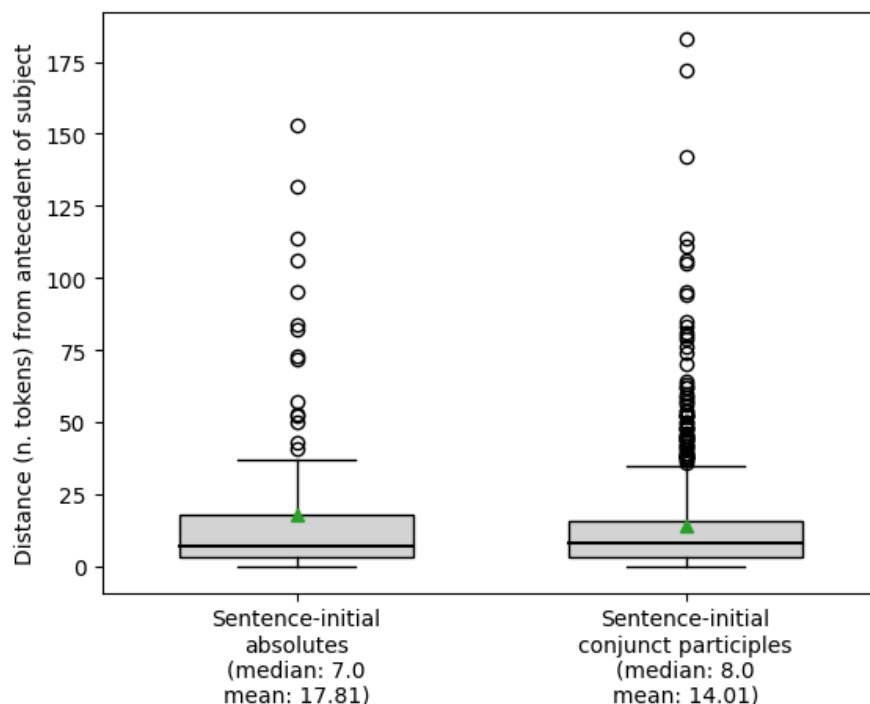


Figure 1.2: Average anaphoric distance between the subject and immediate antecedent in sentence-initial dative absolutes and conjunct participle, including both null and overt subjects. Test results: Welch’s t -test: 1.44, p -value = 0.07; one-tailed Mann–Whitney U -test: 42747, p -value = 0.49.

to occur, overall, with highly salient referents (e.g. ‘Jesus’ in the New Testament) and therefore to be able to have longer and more distant anaphoric relations without incurring in ambiguity in the resolution of the anaphora. This is what we find if we compare the average distance of the antecedent of overt-subject and null-subject conjunct participles in sentence-initial position (Figure 1.3), with null subjects having, on average, a more distant antecedent than overt subjects.

This overall property of null subjects would be expected to skew the result ‘in favour’ of the subjects of conjunct participles in Figure 1.2, but that was not the case. In itself, this suggests that the reinstatement of older referents as topics for a new discourse is a consistent and more evident function of dative absolutes, which, as we have seen, typically work as FRAMES. As further confirmation, if we compare only overt subjects in sentence-initial dative absolutes and conjunct participles, the former have, on average, significantly more distant antecedents than the latter, as Figure 1.4 shows.

Yet, if we look at the pick-up *rates* of the subject referents of sentence-initial conjunct participles and dative absolutes in the previous discourse, we see that there is no significant difference between the two constructions, whether we consider both null and overt subjects or overt subjects only. Namely, without considering the *distance* of

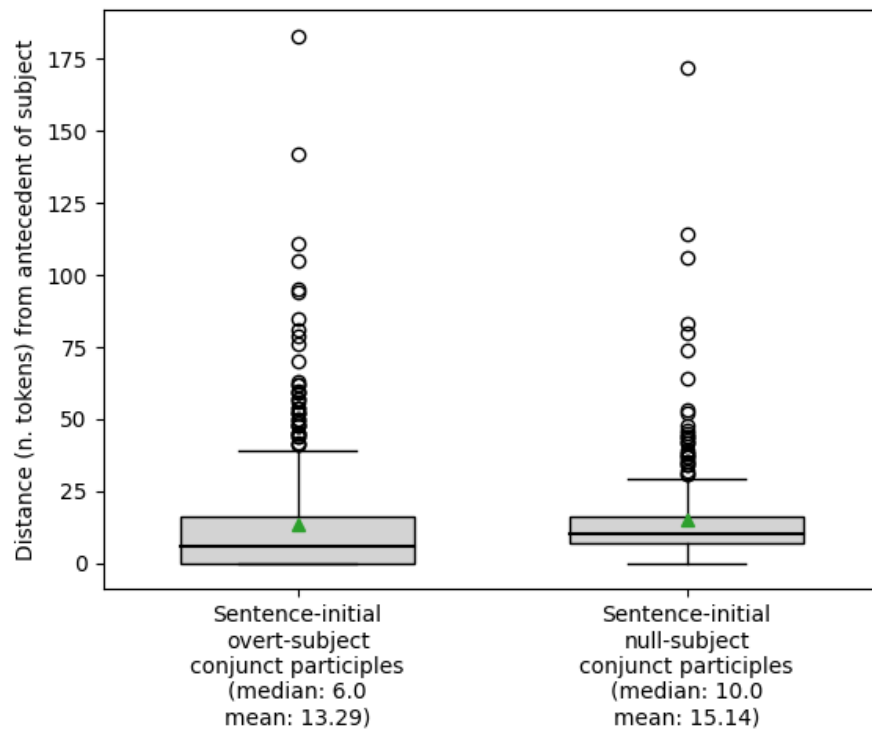


Figure 1.3: Average anaphoric distance between the subject and immediate antecedent in overt-subject and null-subject conjunct participles in sentence-initial position. Test results: Welch’s t -test: -1.26 , p -value = 0.2 ; one-tailed Mann–Whitney U -test: 43138.5 , p -value < 0.01 .

previous sentences (in number of tokens), the two constructions are very similar when it comes to the number of times the referent of their subject is mentioned in the previous discourse. Figures 1.5a-1.5d compare the average pick-up rate of the subject referent of sentence-initial dative absolutes and conjunct participles (both overt and null subjects) at a different number of preceding sentences (1, 5, 30 and 60). Figures 1.6a-1.6d compare the average pick-up rate of the subject referent of sentence-initial dative absolutes and conjunct participles, but only considering overt subjects.

Conjunct participles and dative absolutes, then, seem to be quite similar when it comes to the overall saliency of their subjects, considered only as the number of mentions in the previous discourse. When crossed with these results, however, the frequency of different parts of speech among their subjects suggests a somewhat more complex picture.

1.3.2 Parts of speech

As Table 1.6 shows, most of the subjects of dative absolutes belong to categories typically encoding old or accessible referents.

Only indefinite pronouns and common nouns can potentially introduce new referents: these together represent only 24.5% of all subjects, without, however, considering that

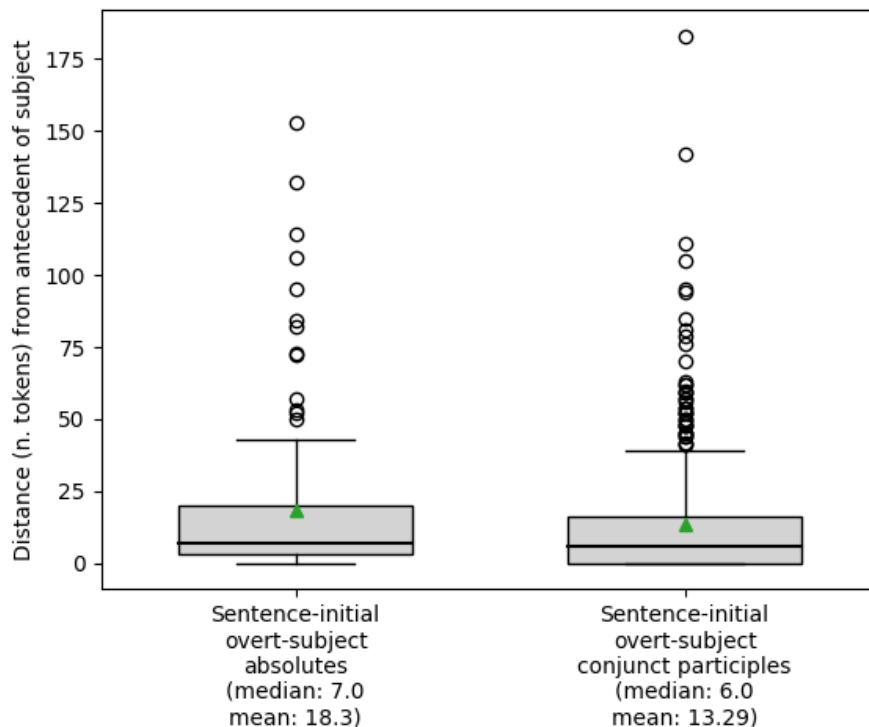


Figure 1.4: Average anaphoric distance between the subject and immediate antecedent in sentence-initial dative absolutes and conjunct participle, overt subjects only. Test results: Welch’s t -test: 1.8, p -value = 0.03; one-tailed Mann–Whitney U -test: 28299.5, p -value = 0.01.

part of speech	frequency
personal pronouns	53.3% (98)
common nouns	22.3% (41)
proper nouns	5.9% (11)
demonstrative pronouns	2.2% (4)
indefinite pronouns	2.2% (4)
<i>null</i>	14.1% (26)

Table 1.6: Dative absolutes in the Codex Marianus: subject parts of speech

several common nouns will also likely have old or accessible referents (e.g. *narodъ* ‘crowd, multitude’, *běsъ* ‘demon’, *gospodъ* ‘Lord’, *slъnъce* ‘sun’).

As Table 1.7 shows, as many as 51.1% of all subjects are a third-person pronoun **i*.⁷

Common nouns are instead the most frequent part of speech among the subjects of conjunct participles (Table 1.8), followed not too far behind by proper nouns.

Demonstratives are also very frequent among the subjects of conjunct participles. Note that, although there is no etymological form attested for the nominative third-person

⁷The asterisk (here and below) is due to the fact the third-person personal pronoun is not attested in Old Church Slavonic in the nominative singular. Instead, demonstratives are generally used in the nominative (*tъ* ‘that one’, *onъ* ‘that one there’ and *sъ* ‘this one’).

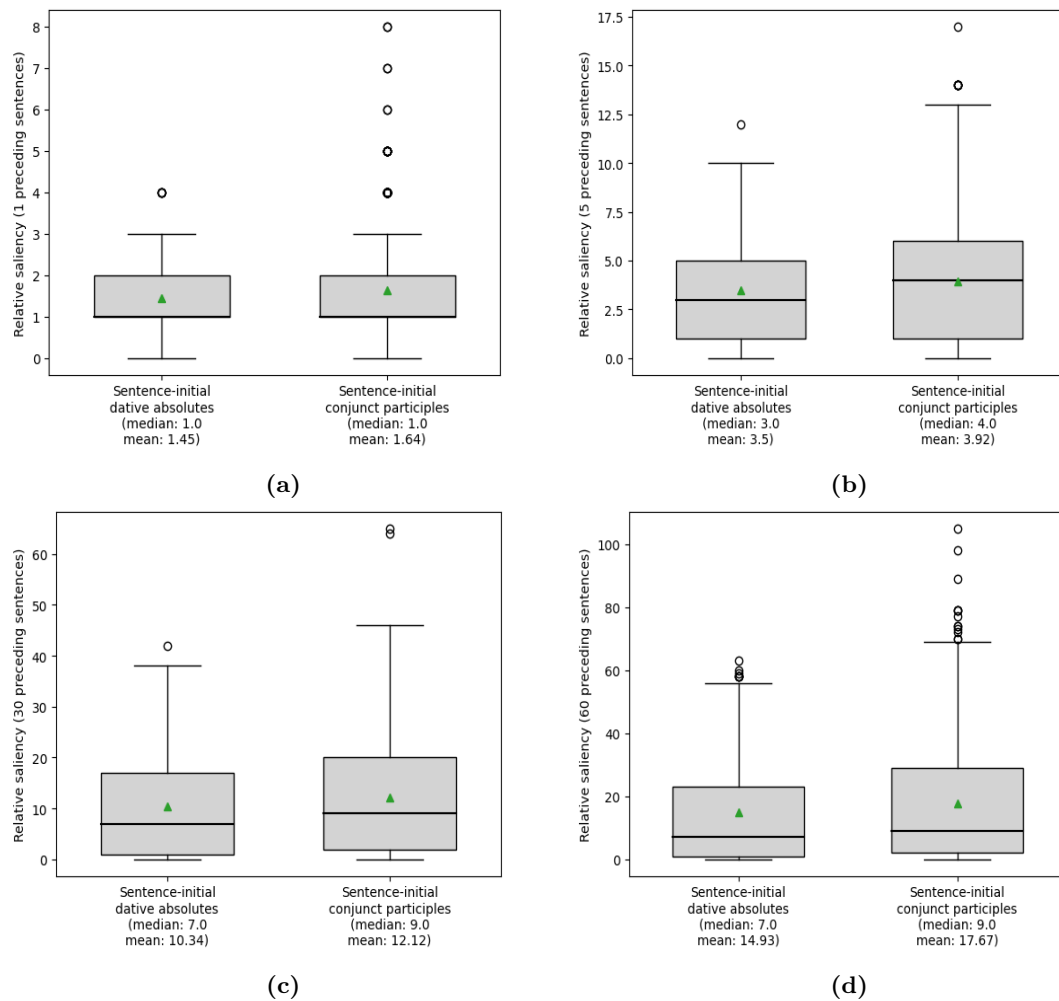


Figure 1.5: Pick-up rates of subject referents in the previous discourse: sentence-initial dative absolutes and conjunct participles (both overt and null subjects), based on the preceding 1, 5, 30, and 60 sentences. The green triangle represents the mean. Test results: (a) Welch’s t -test: -1.82 , p -value = 0.07 ; one-tailed Mann–Whitney U -test: 40503.5 , p -value = 0.19 . (b) Welch’s t -test: -1.42 , p = 0.15 ; one-tailed Mann–Whitney U -test: 39996 , p = 0.14 . (c) Welch’s t -test: -1.69 , p = 0.09 ; one-tailed Mann–Whitney U -test: 39996 , p = 0.11 . (d) Welch’s t -test: -1.5 , p = 0.13 ; one-tailed Mann–Whitney U -test: 39910 , p = 0.13 .

personal pronoun ($*i$), demonstrative forms such as $on\bar{o}$ ‘that one (there)’ and $t\bar{o}$ ‘that one’, particularly the former, are often used instead. Demonstrative pronouns and personal pronouns, then, should perhaps be counted together in Table 1.8 for a fair comparison with dative absolutes (where the third person personal pronoun occurs in the dative).

After $isus\bar{o}$ ‘Jesus’, the demonstrative $on\bar{o}$ ‘that one (there)’ is the most common lemma among the subjects of conjunct participles. As Table 1.10 shows, among the top 10 common nouns occurring as the subject of conjunct participles, there are only human referents. Among the subjects of dative absolutes, it is instead relative frequent to find inanimate referents, such as $sl\bar{o}n\bar{o}bce$ ‘sun’, $pe\check{c}alb$ ‘suffering, grief’, $\check{c}as\bar{o}$ ‘time, moment,

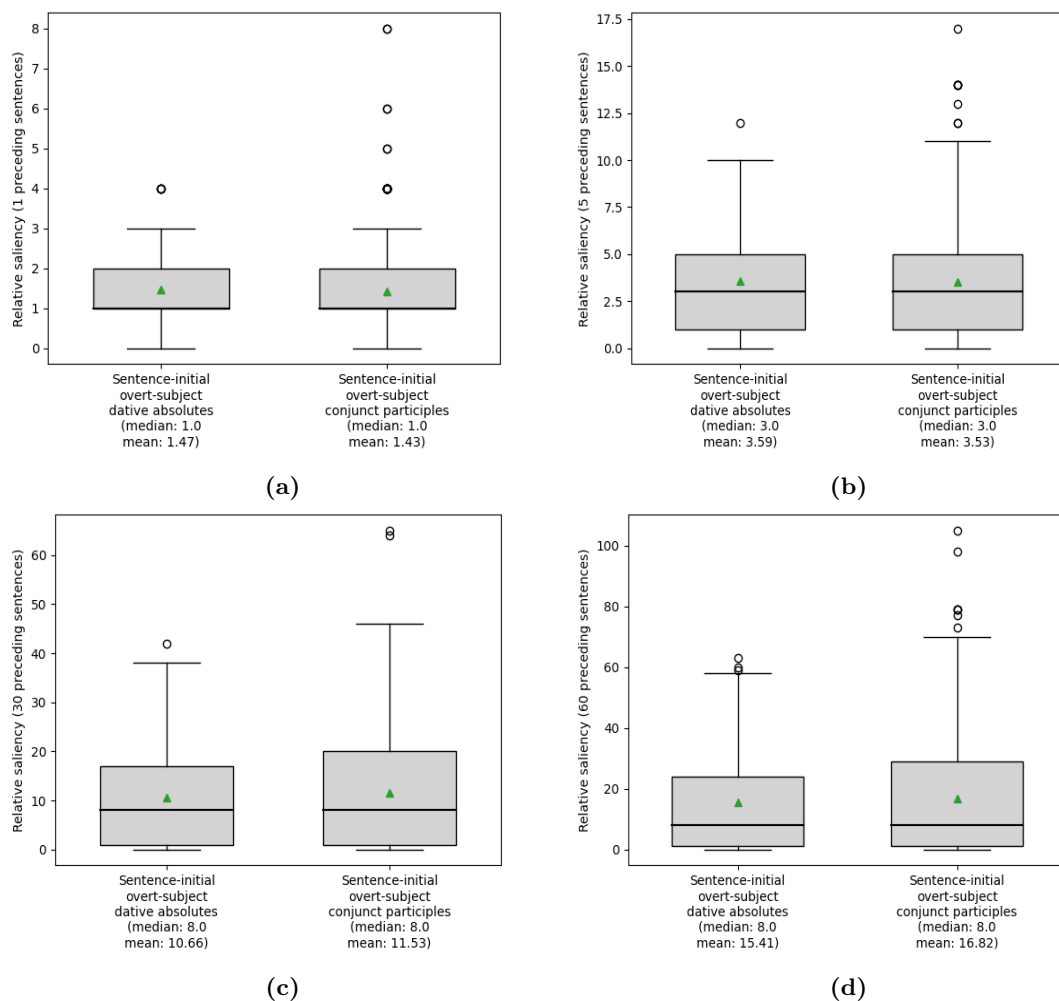


Figure 1.6: Pick-up rates of subject referents in the previous discourse: sentence-initial dative absolutes and conjunct participles (overt subjects only), based on the preceding 1, 5, 30, and 60 sentences. The green triangle represents the mean. Test results: (a) Welch’s t -test: 0.33, p -value = 0.73; one-tailed Mann–Whitney U -test: 26498, p -value = 0.42. (b) Welch’s t -test: 0.18, p = 0.85; one-tailed Mann–Whitney U -test: 25941.5, p = 0.68. (c) Welch’s t -test: -0.77, p = 0.44; one-tailed Mann–Whitney U -test: 25091, p = 0.87. (d) Welch’s t -test: -0.72, p = 0.47; one-tailed Mann–Whitney U -test: 25184, p = 0.92.

hour’, *dvrb* ‘door’, and *větrō* ‘wind’.

As typically framing adverbials, this is not surprising, since inanimate referents are often part of the stage in which a new discourse is set (e.g. ‘when the door opened’, ‘as the wind blow’, ‘when the sun rose’, etc.).

If sentence-initial conjunct participles in the VS configuration are more likely to be interpreted as FRAMES than those in the SV one, which may be more likely to function as INDEPENDENT RHEMES, we may also expect some differences in the parts of speech most commonly occurring among the subjects in the two configurations. Tables 1.11 and 1.12 show the frequency of different parts of speech among sentence-initial conjunct

lemma	frequency
<i>*i</i> ‘he, they’	51.1% (94)
<i>narodŭ</i> ‘crowd, multitude’	3.3% (6)
<i>isusŭ</i> ‘Jesus’	2.7% (5)
<i>slŭnŭce</i> ‘sun’	2.2% (4)
<i>vŭsb</i> ‘all’	1.6% (3)
<i>tŭ</i> ‘that (one)’	1.6% (3)
<i>běsŭ</i> ‘demon’	1.1% (2)
<i>gospodŭ</i> ‘Lord’	1.1% (2)
<i>ljudije</i> ‘people’	1.1% (2)
<i>člověkŭ</i> ‘person’	1.1% (2)

Table 1.7: Dative absolutes in the Codex Marianus: ten most-frequent subject lemmas

part of speech	frequency
common nouns	20.1% (315)
proper nouns	17.2% (270)
demonstrative pronouns	8.1% (127)
indefinite pronouns	2.2% (35)
adjectives	1.7% (26)
verbs	1.7% (26)
personal pronouns	1.4% (22)
cardinal numerals	1.2% (19)
relative pronouns	1% (15)
ordinal numerals	0.4% (6)
interrogative pronouns	0.2% (4)
<i>null</i>	44.8% (703)

Table 1.8: Conjunct participles in the Codex Marianus: subject parts of speech

lemma	frequency
<i>isusŭ</i> ‘Jesus’	12.1% (190)
<i>onŭ</i> ‘that one (there)’	6.8% (106)
<i>učeníkŭ</i> ‘disciple’	2.48% (39)
<i>jedinŭ</i> ‘someone’	2.1% (33)
<i>žena</i> ‘woman, wife’	1.9% (30)
<i>farisei</i> ‘Pharisee’	1.46% (23)
<i>petrŭ</i> ‘Peter’	1.08% (17)
<i>narodŭ</i> ‘crowd, multitude’	1.08% (17)
<i>archierei</i> ‘bishop’	1.02% (16)
<i>člověkŭ</i> ‘person’	0.89% (14)

Table 1.9: Conjunct participles in the Codex Marianus: ten most-frequent subject lemmas

participles in the VS and in the SV configuration, respectively.

We observe, in fact, neat differences between the two configurations, most evidently as

construction	most frequent common nouns
dative absolutes	<i>narodŭ</i> ‘crowd, multitude’, <i>slŏnbce</i> ‘sun’, <i>gospodŭ</i> ‘Lord’, <i>dŏšti</i> ‘daughter’, <i>ljudiŭje</i> ‘people’, <i>člověkŭ</i> ‘person’, <i>časŭ</i> ‘time, moment, hour’, <i>pečalb</i> ‘suffering, grief’, <i>běsŭ</i> ‘demon’, <i>dvvrŭ</i> ‘door’, <i>větrŭ</i> ‘wind’
conjunct participles	<i>učeníkŭ</i> ‘disciple’, <i>žena</i> ‘woman, wife’, <i>farisei</i> ‘Pharisee’, <i>narodŭ</i> ‘crowd, multitude’, <i>archierei</i> ‘bishop’, <i>člověkŭ</i> , <i>sŏtŏnikŭ</i> ‘centurion’, <i>rabŭ</i> ‘servant, slave’, <i>angelŭ</i> ‘angel’, <i>cěsarŭ</i> ‘emperor, king’

Table 1.10: Dative absolutes and conjunct participles in the Codex Marianus: most frequent common nouns among subjects lemmas

part of speech	frequency
proper nouns	50.5% (138)
common nouns	35.5% (97)
verbs	4% (11)
demonstrative pronouns	2.9% (8)
indefinite pronouns	2.6% (7)
cardinal numerals	2.2% (6)
personal pronouns	0.7% (2)
adjectives	0.7% (2)
relative pronouns	0.4% (1)
ordinal numerals	0.4% (1)

Table 1.11: Pre-matrix VS conjunct participles in the Codex Marianus: subject parts of speech

part of speech	frequency
demonstrative pronouns	30.1% (90)
common nouns	28.4% (85)
proper nouns	20.7% (62)
personal pronouns	5% (15)
adjectives	3.7% (11)
indefinite pronouns	3.7% (11)
relative pronouns	3.3% (10)
cardinal numerals	2% (6)
interrogative pronouns	1.3% (4)
verbs	1% (3)
ordinal numerals	0.7% (2)

Table 1.12: Pre-matrix SV conjunct participles in the Codex Marianus: subject parts of speech

far as the relative frequency of proper nouns and demonstrative pronouns is concerned. Demonstratives are much more frequent in the SV than in the VS configuration (90 occurrences as opposed to 8, respectively), whereas proper nouns are more common in

the latter than in the former (138 in the VS as opposed to 62 in the SV configuration). Most of the personal pronouns are also found in the SV one (15, as opposed to 2 in the VS one), which seems to support what we just observed above, namely that there is a case to be made for demonstrative pronouns and personal pronouns to be counted together in the analysis. These figures, overall, support what we have already observed, namely that the VS configuration in conjunct participles is much more likely to indicate a framing function than the SV one, as Haug (2012) had already argued regarding conjunct participles in Ancient Greek. In order to resolve the anaphora in a clause with a pronoun or a demonstrative, the antecedent should not, for ease of processing, be too far in the surrounding discourse from the anaphor itself, except with very salient referents. This is reflected both in the smaller average distance of the antecedent in sentence-initial conjunct participles, as we saw above, and in the inherently anaphoric parts of speech (demonstratives and personal pronouns), which are much more common in the SV than in the VS configuration. Among sentence-initial VS conjunct participles, which are more likely to be used as FRAMES, on the other hand, proper nouns (which are naturally referentially more explicit than demonstratives and personal pronouns) are much more common than in the SV configuration. This also agrees with the observation that the average distance of the immediate antecedent of sentence-initial SV conjunct participles is significantly greater than that in the VS configuration.

These observations, however, seem at odds with the fact that the subjects of dative absolutes in the Codex Marianus have, on average, similarly distant antecedents as sentence-initial conjunct participles but *also* have inherently anaphoric parts of speech occurring overwhelmingly more frequently among their subjects than more explicit ones (e.g. proper nouns, as with VS conjunct participles).

We may expect the average distance to the antecedent of the subjects of dative absolutes to be at least shorter than those of sentence-initial VS conjunct participles, among which, as we have seen, referentially more explicit parts of speech are instead more common. As Figure 1.7 shows, however, that is not the case. Just like the other configurations (cf. Figures 1.2 and 1.4), there is no significant difference between sentence-initial dative absolutes and sentence-initial VS conjunct participles in terms of average distance from their subject to its immediate antecedent.

This difference between the subjects of dative absolutes and the subjects of VS conjunct participles, which are otherwise functionally very similar to dative absolutes, suggests that dative absolutes occur overwhelmingly more frequently with highly discourse-prominent referents (in the sense of Himmelmann & Primus 2015 and Heusinger & Schumacher 2019, among others), which a measure of saliency as mere number of previous mentions in the discourse may not be enough to capture. As we will see in more detail in the next chapter with a much larger and more diverse dataset, dative absolutes generally seem to attract fewer, but particularly prominent, referents than conjunct participles

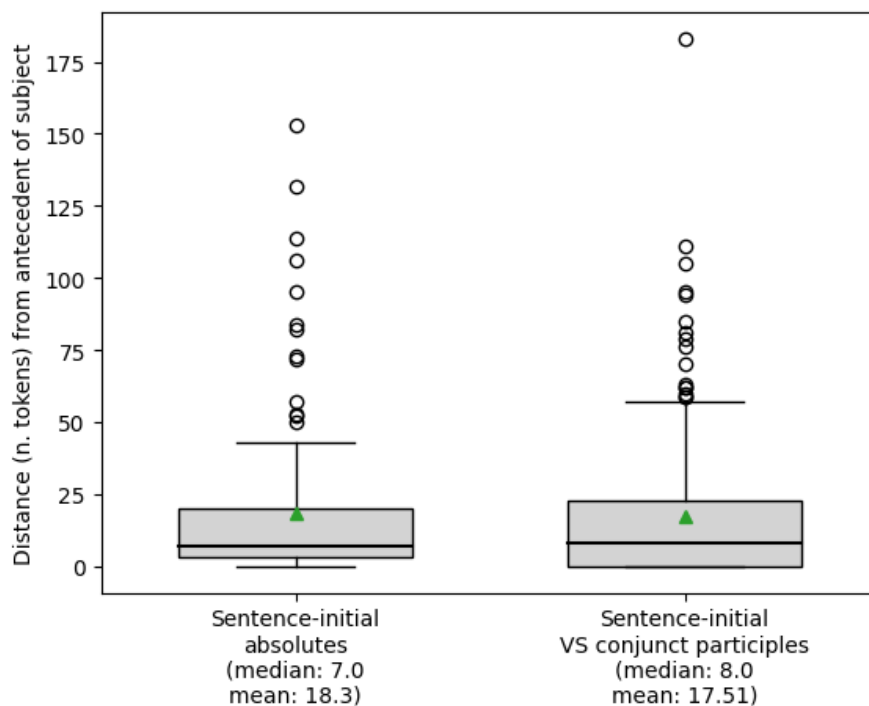


Figure 1.7: Average anaphoric distance between subject and immediate antecedent in sentence-initial dative absolutes and VS conjunct participle, overt subjects only. Test results: Welch’s *t*-test: 0.25, *p*-value = 0.39; one-tailed Mann–Whitney *U*-test: 14003.5, *p*-value = 0.46.

as a whole. In the Gospels, the majority of third-person singular personal pronouns are likely to refer to ‘Jesus’ and the majority of third-person plurals to the ‘Apostles’, which are highly salient and prominent referents throughout the New Testament. As Heusinger & Schumacher (2019: 123) argues, more prominent discourse referents ‘show a higher forward-looking potential in that they are referred to more often and they more frequently constitute the topic of the next utterance’, providing ‘a better accessibility for subsequent anaphoric expressions’, which may explain why, despite the longer and more distant anaphoric chains observed in the Gospels, inherently anaphoric parts of speech are overwhelmingly more frequent than others among the subjects of dative absolutes.

1.3.3 Summary

This section analyzed some of the properties of the subjects of conjunct participles and dative absolutes in sentence-initial position, which is where both constructions can function as FRAMES and where conjunct participles can be ambiguous between a FRAME and an INDEPENDENT RHEME interpretation. The vast majority of sentence-initial dative absolutes, as expected, were found to have an overt subject, whereas sentence-initial conjunct participles showed a more even split, although overt subjects were found to be

somewhat more frequent (60.5% overt subjects versus 39.5% null subjects).

Zooming in on overt-subject conjunct participles and dative absolutes, we saw that the position of the subject relative to the participles is predominantly post-verbal in dative absolutes, whereas in conjunct participles the SV and VS configurations are found with relatively even frequency (45.7% SV and 54.3% VS).

Looking at the anaphoric properties of overt subjects in sentence-initial conjunct participles, we observed a significant association between the position of the subject relative to the participle and the average anaphoric distance of the subject to its immediate antecedent, whereby subjects in the VS configuration have, on average, significantly more distant antecedents than subjects in the SV one. This had already been observed by Haug (2012) regarding Ancient Greek conjunct participles and can be associated with the reinstatement of old or inactive referents. This is a function that can be generally ascribed to FRAMES, which are adjuncts in topicalized position and, as such, are more likely to be used to explicitly topicalize eventualities (and their participants) when there is a shift of some kind in the discourse (of subject, agent, place, time, etc.).

A comparison between sentence-initial dative absolutes and sentence-initial conjunct participles indicated that the subjects of the two constructions have overall similarly distant antecedents and that they are similarly activated in the previous discourse. Relative *saliency* was here simply calculated as the number of mentions (pick-up rates) of a subject referent at different sentence windows (1, 5, 30, and 60 preceding sentences). While no significant differences between the relative saliency of subjects in sentence-initial dative absolutes and conjunct participles were detected, the frequency of different parts of speech among their subjects suggested that there may be differences between their subjects which a simple measure of saliency is not able to capture. Among VS conjunct participles, which are more likely to work as FRAMES, referentially more explicit parts of speech (e.g. proper nouns) are the most frequent. Because VS participle constructions help reinstate older referents, the predominance of such parts of speech among VS conjunct participles was explained with the fact that they may help facilitate anaphora resolution. However, the vast majority of subjects among dative absolutes are personal pronouns (i.e. inherently anaphoric parts of speech), despite also working typically as FRAMES and having their subject following the participles. A possibility was advanced that absolute constructions generally attract few, but highly prominent subject referents (e.g. Jesus and the Apostles, as far as the New Testament is concerned). However, only a larger and more diverse dataset (Chapter 2) may verify whether this is the case.

1.4 Lexical variation among participles

Lexical variation among participles in different syntactic configurations can also indirectly point to important functional differences. Table 1.13 shows the 10 most frequent lemmas among dative absolutes and conjunct participles.

subcorpus	absolutes	conjuncts
total	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>vɔniti</i> ‘go in’, <i>iti</i> ‘go’, <i>isɔchoditi</i> ‘exit, go out’, <i>vɔzlezati</i> ‘lay down’, <i>jasti</i> ‘eat’, <i>priti</i> ‘come, arrive’, <i>iziti</i> ‘go down’, <i>iměti</i> ‘have’	<i>viděti</i> ‘see’, <i>priti</i> ‘come, arrive’, <i>slyšati</i> ‘hear’, <i>iti</i> ‘go’, <i>pristopiti</i> ‘ascend, approach’, <i>iziti</i> ‘go down’, <i>prijęti</i> ‘take, receive’, <i>vɔstati</i> ‘stand up’, <i>byti</i> ‘be’, <i>vɔzbręti</i> ‘look up at’
left	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>vɔniti</i> ‘go in’, <i>iti</i> ‘go’, <i>isɔchoditi</i> ‘exit, go out’, <i>vɔzlezati</i> ‘lay down’, <i>jasti</i> ‘eat’, <i>iziti</i> ‘go down’, <i>priti</i> ‘come, arrive’, <i>iměti</i> ‘have’	<i>viděti</i> ‘see’, <i>priti</i> ‘come, arrive’, <i>slyšati</i> ‘hear’, <i>iti</i> ‘go’, <i>pristopiti</i> ‘ascend, approach’, <i>iziti</i> ‘go down’, <i>prijęti</i> ‘take, receive’, <i>vɔstati</i> ‘stand up’, <i>vɔzbręti</i> ‘look up at’, <i>byti</i> ‘be’
right	<i>byti</i> ‘be’, <i>pristopiti</i> ‘ascend, approach’, <i>vɔsijati</i> ‘begin to shine’, <i>sɔpati</i> ‘sleep’, <i>vlasti</i> ‘rule’, <i>vɔniti</i> ‘go in’, <i>zatvoriti</i> ‘close’, <i>vɔzdati</i> ‘give (in homage)’, <i>pospěšstvovati</i> ‘assist, promote’, <i>utvrždati</i> ‘establish, build’	<i>učiti</i> ‘learn, teach’, <i>byti</i> ‘be’, <i>iměti</i> ‘have’, <i>iskušati</i> ‘tempt, try’, <i>viděti</i> ‘see’, <i>iskati</i> ‘seek’, <i>rešti</i> ‘say’, <i>věděti</i> ‘know’, <i>moliti se</i> ‘pray’, <i>propovědati</i> ‘preach’
SV	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>zatvoriti</i> ‘close’, <i>isɔchoditi</i> ‘exit, go out’, <i>četvrvtovlaststvovati</i> ‘serve as a tetrarch’, <i>vɔsijati</i> ‘begin to shine’, <i>čuditi</i> ‘marvel’, <i>plęsati</i> ‘dance’, <i>vɔzlezati</i> ‘lay down’, <i>ugoditi</i> ‘please, satisfy’	<i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>učiti</i> ‘learn, teach’, <i>iměti</i> ‘have’, <i>iti</i> ‘go’, <i>věděti</i> ‘know’, <i>prijęti</i> ‘take, receive’, <i>iskušati</i> ‘tempt, try’
VS	<i>byti</i> ‘be’, <i>vɔniti</i> ‘go in’, <i>glagolati</i> ‘speak, say’, <i>iti</i> ‘go’, <i>iziti</i> ‘go down’, <i>priti</i> ‘come, arrive’, <i>vɔzlezati</i> ‘lay down’, <i>isɔchoditi</i> ‘exit, go out’, <i>jasti</i> ‘eat’, <i>sɔchoditi</i> ‘go down’	<i>viděti</i> ‘see’, <i>iti</i> ‘go’, <i>priti</i> ‘come, arrive’, <i>pristopiti</i> ‘ascend, approach’, <i>slyšati</i> ‘hear’, <i>iziti</i> ‘go down’, <i>vɔstati</i> ‘stand up’, <i>prijęti</i> ‘take, receive’, <i>vɔniti</i> ‘go in’, <i>vɔzbręti</i> ‘look up at’

Table 1.13: Dative absolutes and conjunct participles in the Codex Marianus: ten most frequent lemmas, overall, by position of the participle relative to the matrix clause, and by position of the subject relative to the participle

Motion verbs, such as *priti* ‘come, arrive’, *vɔniti* ‘go in’, and simply *iti* ‘go’ are common among both conjunct participles and dative absolutes across all configurations. *Byti* ‘be’ is the top lemma among dative absolutes in virtually all configurations, except the SV one, which, however, is an infrequent configuration among absolutes in the first place. Among dative absolutes we find several verbs denoting activities, such as *jasti* ‘eat’, *sɔpati* ‘sleep’, and *plęsati* ‘dance’, which one can easily envision as framing situations

within which a main eventuality occurs, as in (22)-(23).

- (22) a. ***sypřeštemъ*** *že* *člvkomъ.* *pride*
 sleep.PTCP.IPFV.M.DAT.SG PTC person.PL.M.DAT come.AOR.3.SG
vragъ *ego.* *i* *vbsě* *plěvelъ*
 enemy.M.NOM.SG 3.SG.M.GEN and sow.AOR.3.SG weed.M.ACC.SG
po *srědě* *pšenice* *i* *otide.*
 among wheat.SG.F.DAT and.SG.F.GEN and leave.AOR.3.SG
- b. *en de tōi* ***katheudein*** *tous* *anthrōpous*
 in PTC the.N.DAT.SG sleep.PRS.INF the.PL.M.ACC man.PL.M.ACC
ēlthen *autou* *ho* *ekhthros* *kai*
 come.AOR.3.SG 3.SG.M.GEN the.M.NOM.SG enemy.M.NOM.SG and
epespeiren *zizania* *ana meson* *tou*
 sow.AOR.3.SG weed.PL.N.ACC in among.N.ACC.SG the.M.GEN.SG
sitou *kai apēlthen*
 wheat.M.GEN.SG and depart.AOR.3.SG
 ‘But while his men were sleeping, his enemy came and sowed weeds among the wheat, and left’ (Matthew 13:25)
- (23) a. ***Ēdōštemъ*** *že* *imъ* *priimъ*
 eat.PTCP.IPFV.M.DAT PTC 3.PL.M.DAT take.PTCP.PFV.M.NOM.SG
is *chlěbъ* *i* *blgsštъ*
 Jesus.M.NOM.SG bread.M.ACC.SG and bless.PTCP.PFV.M.NOM.SG
prělo mi
 break.AOR.3.SG
- b. ***Esthiontōn*** *de* *autōn* *labōn*
 eat.PTCP.IPFV.M.GEN PTC him.M.GEN.PL take.PTCP.PFV.M.NOM.SG
ho *Iēsous* *arton* *kai*
 the.M.NOM.SG Jesus.M.NOM.SG bread.M.ACC.SG and
eulogēsas *eklasen*
 bless.PTCP.PFV.M.NOM.SG break.AOR.3.SG
 ‘And as they were eating, Jesus took bread, blessed and broke it’ (Matthew 26:26)

Among conjunct participles, besides the pool of frequent lemmas in common with dative absolutes, we find several telic verbs, such as *vъzeti* ‘take’, *prijeti* ‘take, receive’, *slyšati* ‘hear’ and *vъzvrěti* ‘look up at’. We can more easily envision these as part of serial events, namely as INDEPENDENT RHEMES, as in (24)-(25) (see also *priimъ* in (23)).

- (24) a. *tъ* ***vъzemъ*** *daždъ*
 that.M.ACC.SG take.PTCP.PFV.M.NOM.SG give.2.SG.PRS.IMP
imъ *za me* *i* *za se.*
 him.M.DAT.PL for 1.SG.ACC and for self.ACC.SG

- b. *ekeinon labōn dos*
 that.M.ACC.SG take.PTCP.PFV.M.NOM.SG give.IMP.PFV.PST.2.SG
autois anti emou kai sou
 3.PL.M.DAT for 1.SG.GEN and yourself.GEN.SG
 ‘Take that and give it to them for me and you’ (Matthew 17:27)

- (25) a. *is že vzbřěvš na nb*
 Jesus.M.NOM.SG PTC look at.PTPC.PFV.M.NOM.SG on 3.SG.M.ACC
vzljubi i
 love.AOR.3.SG 3.SG.M.ACC

- b. *ho de Iēsous emblepsas*
 the.M.NOM.SG PTC Jesus.M.NOM.SG look at.PTCP.PFV.M.NOM.SG
autōi ēgapēsen auton
 3.SG.M.DAT love.AOR.3.SG 3.SG.M.ACC
 ‘Jesus looked at him and loved him’ (Mark 10:21)

A large portion of the Greek New Testament in PROIEL includes annotation on the lexical aspect (Aktionsart) of verbs, including the categories ACTIVITY, SEMELFACTIVE, STATE, and TELIC. This level of annotation is experimental (Haug & Jøhndal 2008: 41) and seems to broadly follow Vendler’s (1967) terminology and Comrie’s (1976) categorization into three higher-level Aktionsarten, namely TELIC, ATELIC and STATIVE, with the first two further divided into two subcategories. ACCOMPLISHMENTS and ACHIEVEMENTS fall into the TELIC category, while SEMELFACTIVES and ACTIVITIES fall into the ATELIC one. The PROIEL annotation thus only attempted to further distinguish ATELIC verbs and kept ACCOMPLISHMENTS and ACHIEVEMENTS subsumed under TELIC.

We can cross the observations just made about the type of lemmas occurring among the constructions with the Aktionsart annotation in PROIEL on the Greek token corresponding to the participles. Since the main ambiguity is between pre-matrix constructions, we shall narrow down the extraction of tags to pre-matrix occurrences only. Tables 1.14 and 1.15 show the Aktionsart tags corresponding to pre-matrix conjunct participles and pre-matrix dative absolutes, respectively.⁸

		matrix				tot.
		ACT	SEM	ST	TEL	
conjunct	ACT	0	0	0	3.58% (21)	3.58% (21)
	SEM	0	0	0	0.68% (4)	0.68% (4)
	ST	0.51% (3)	0	1.7% (10)	12.61% (74)	14.82% (87)
	TEL	6.64% (39)	1.53% (9)	2.39% (14)	70.36% (413)	80.92% (475)

Table 1.14: Aktionsarten of pre-matrix conjunct participles and their matrix clause in the Codex Marianus. ACT = activity, SEM = semelfactive, ST = state, TEL = telic.

⁸Note that, despite the wide coverage, the annotation is not complete, so some occurrences do not figure here.

		matrix				tot.
		ACT	SEM	ST	TEL	
absolute	ACT	0	0.58% (1)	1.17% (2)	18.71% (32)	20.47% (35)
	SEM	0	0	0	0.58% (1)	0.58% (1)
	ST	0	0	0	15.2% (26)	15.2% (26)
	TEL	5.26% (9)	0	5.26% (9)	53.22% (91)	63.74% (109)

Table 1.15: Aktionsarten of pre-matrix dative absolutes and their matrix clause in the Codex Marianus. ACT = activity, SEM = semelfactive, ST = state, TEL = telic.

Telic verbs are the most frequent among both constructions, although more prominently so among conjunct participles. As anticipated, what is most striking is the high frequency of activity verbs among dative absolutes. Their total raw frequency is, in fact, higher than the one of activity verbs among conjunct participles, despite the difference in overall frequency between the two constructions.

As Haug (2012: 320) argues on the basis of Ancient Greek, FRAMES are assumed to be more ‘predictable’, since they are ‘presupposed (anaphoric or accommodated) and not explicitly asserted’, so that we expect to find less lexical variation among sentence-initial participles than in other positions. ELABORATIONS and INDEPENDENT RHEMES are instead expected to present more variation as they typically introduce new information. Since the post-matrix position is likely to be predominantly occupied by ELABORATIONS, we can expect post-matrix conjunct participles to show a very high degree of lexical variation. Pre-matrix conjunct participles are instead often ambiguous between FRAMES and INDEPENDENT RHEMES, so we expect a lower degree of lexical variation than among post-matrix participles. To measure the overall lexical richness among participle lemmas in each participle construction, I use two metrics: the percentage of participle occurrences belonging to the 10 most frequent lemmas for a given configuration or subcorpus and a moving-average type-token ratio (MATTR). Standard type-token ratio (TTR) is a simple measure of lexical richness consisting of a text’s vocabulary size divided by the text size and can be used to look at the proportion of total number of tokens and unique lemmas. Given the difference in dataset sizes between conjunct participles and dative absolutes, I use a variation of TTR (MATTR; Covington & McFall 2010) consisting of the average between the TTRs of a moving window of n occurrences, so that the final score is not affected by sample size. MATTR is, of course, normally calculated over a window of adjacent tokens in running text. I instead treat subsequent occurrences of each participle construction as the adjacent tokens and calculate MATTR over the resulting list. For example, assuming a text has 10 occurrences of dative absolutes, we take the lemma of each occurrence and place them in an ordered list as follows:

```
[`byti', `byti', `priiti', `byti', `chotěti',
`chotěti', `glagolati', `priiti', `byti', `chotěti']
```

We then treat the ordered list as running text and calculate MATTR over the predefined window size. For example, if our window was of 2 tokens, we would calculate the TTR of the list above for the first two occurrences, then for the second and third occurrence, then for the third and fourth, and so on. We then average all the TTRs obtained in this way to get the MATTR score. The closer the MATTR to 1, the higher the degree of lexical richness, thus, in our case, variation among participle lemmas. A window of 40 occurrences was chosen after manually evaluating between windows of {20,40,60} occurrences and observing that 40 was the smallest number of occurrences that allowed us to look at the lexical variation in most subsamples of the corpus without losing information about differences in variation. Too great a window size would result in not being able to look at variation in potentially meaningful subsamples (e.g. the SV configuration of dative absolutes) where the number of occurrences is below the window size, while too small a window size would result in not capturing variation at all.

In Table 1.16, the first column for each of the constructions indicates how many occurrences belong to the 10 most-frequent lemmas in that configuration (i.e. whether we are dealing with few high-frequency lemmas or several low-frequency lemmas), the second the MATTR score.⁹

Subsample	Absolutes		Conjuncts	
	10MFL	MATTR	10MFL	MATTR
total	50.54%	0.64	36.79%	0.72
SV	NA	NA	47.34	0.70
VS	45.83	0.64	64.71	0.50
left	51.46%	0.64	44.12%	0.67
right	NA	NA	30.11%	0.81

Table 1.16: Lexical variation among participles, overall, by position relative to the matrix clause, and by position of the subject relative to the participle. *10MFL* = 10 most frequent lemmas; *MATTR* = moving-average type-token ratio; *left* and *right* = pre-matrix and post matrix, respectively. Note that the closer the MATTR is to 1 (and the lower the percentage for the 10MFL), the greater the lexical variation.

As before, post-matrix dative absolutes, as well as sentence-initial SV occurrences, are too few to make a comparison, which will have to await the next, much larger dataset. However, as predicted, dative absolutes show less total lexical variation than conjunct participles, except for the VS configuration, where it is lower among conjunct participles. This also supports the observation that VS conjunct participles are more likely to function as FRAMES than SV ones, which are much more likely INDEPENDENT RHEMES. Much

⁹The numbers in Table 1.16 do not include *glagolati* ‘speak, say’ (normally imperfective: *gl[agol]je*) for post-matrix conjunct participles and *otvověštati* ‘answer’ (normally perfective: *otvověštav*) for pre-matrix conjunct participles, since these would likely skew the frequencies. As noted by Haug (2012: 288) on the respective Greek forms (*legōn* ‘say’ and *apokritheis* ‘answer’), these work roughly like quotative particles, a usage sometimes considered a Semiticism (ibid.). *Glagolati* ‘talk, say’ in absolute constructions should arguably not be considered on a par with the same lemma in conjunct participles, since in dative absolutes it is generally used as an intransitive verb (‘talking’ rather than ‘saying’).

like in Greek (see the data in Haug 2012), Old Church Slavonic conjunct participles to the right of the main verb show an overall higher degree of lexical variation than those to the left, as the difference in MATTR and 10MFL in the table indicate. Only a much larger dataset, as the one used in the next chapter, however, may confirm whether this is, in fact, a general pattern.

1.5 Greek-Old Church Slavonic mismatches

It is hard, from a philological standpoint, to judge the extent to which the different participle clauses should be considered ‘native’ Slavonic constructions. The evident pervasiveness of the phenomena intuitively suggests that both dative absolutes and conjunct participles may have, at most, expanded their scope under the influence of Greek, rather than to represent mechanical translation devices. This has already been claimed in the literature (e.g. Birnbaum, 1958; Růžička, 1958; MacRobert, 1986; Corin, 1995) and finds, in fact, a certain degree of evidence in the treebank data. Old Church Slavonic does follow the Greek to a large degree, translating most genitive absolutes into dative absolutes and most conjunct participles into conjunct participles. However, it also displays limited but consistent signs of independence in the use of the constructions as either FRAMES, INDEPENDENT RHEMES or ELABORATIONS.

1.5.1 Conjunct participles

Tables 1.17 and 1.18 show the number of matches and mismatches between conjunct participles in Old Church Slavonic and Greek.

Conjunct participle	1424
Finite main clause	46
Dative absolute	4
Finite subordinate	2
Infinitive	2
Other/NA	33

Table 1.17: Old Church Slavonic constructions corresponding to a conjunct participle in Ancient Greek

Conjunct participle	1424
Finite main clause	34
Genitive absolute	1
Other/NA	112

Table 1.18: Ancient Greek constructions corresponding to a conjunct participle in Old Church Slavonic

Given the much greater frequency of conjunct participles compared to dative absolutes, and since a close-reading comparison between each occurrence of the construction

Dative absolute	153
Finite subordinate	3
Conjunct participle	3
Finite main clause	2
Prepositional phrase	1

Table 1.19: Old Church Slavonic constructions corresponding to a genitive absolute in Ancient Greek

10 genitive absolutes do not appear to be translated into a dative absolute. Of these, the two main clauses in Old Church Slavonic (Luke 23.45, Matthew 17.26) are contradicted by versions of the Greek New Testament other than Tischendorf's also having a main clause, whereas one of the conjunct participles (John 21.11) has the dubious mixed-case form *tolikou sqšt̃b* '(there) being so many', which however corresponds to a regular dative absolute in the *Codex Zographensis* (*tolikou sqšt̃ju*).

Of the potentially genuine discrepancies, three are Greek genitive absolutes corresponding to Old Church Slavonic finite *egda*-clauses (28-29), two of which occur in the same sentence; one is a Greek absolute rendered with a prepositional phrase in Old Church Slavonic, perhaps due to the difficulty in finding an exact match to Greek *mesoō* 'to be in the middle' (30):

- (28) a. *i egda pride ṽb crkṽb pristopišę k̃b*
 and when come.AOR.3.SG in temple.ACC approach.AOR.3.PL towards
nemu učęšt̃ju archierei i
 3.SG.DAT teach.PTCP.IPFV.M.DAT.SG chief.priest.NOM.PL and
star̃vci ljud̃bstii gl̃šte
 elder.NOM.PL people.ADJ.NOM.PL say.PTCP.IPFV.M.NOM.PL
- b. *Kai elthontos autou eis to hieron*
 and come.PTCP.AOR.M.GEN.SG he.GEN in the temple.ACC
prosēlthon autōi didaskonti hoi
 approach.AOR.3.PL 3.SG.DAT teach.PTCP.IPFV.M.DAT.SG the
archiereis kai hoi presbyteroi tou laou
 chief.priest.NOM.PL and the elder.NOM.PL the.GEN.SG people.GEN.SG
legontes
 say.PTCP.IPFV.M.NOM.PL
 'And when he came into the temple, the chief priests and the elders of the
 people confronted him as he was teaching, saying' (Matthew 21:23)
- (29) a. *da ne egda položit̃b osnovaniě. i ne*
 so-that NEG when lay.PRS.3.SG foundation.ACC and NEG
možet̃b sṽvr̃šiti. vsi vidęštei
 be-able.PRS.3.SG finish.INF all.NOM.PL see.PTCP.IPFV.NOM.PL

načřnřtř *řřgati* *sř* *emu* *řko*
begin.PRS.3.PL mock.INF REFL he.DAT that

- b. *hina mēpote thentos* *autou themelion kai*
thus not-ever lay.PTCP.AOR.M.GEN.SG he.GEN.SG foundation.ACC and
mē ischyontos *ektelesai pantes hoi*
NEG be-able.PTCP.IPFV.M.GEN.SG finish.AOR.INF all.NOM.PL the
heōrountes *arxōntai autōi empaizein*
see.PTCP.IPFV.M.NOM.PL begin.AOR.SBJV.MID he.DAT mock.PRS.INF
legontes *hoti*
say.PTCP.IPFV.NOM.PL that

‘... lest, after he has laid the foundation, and is not able to finish, all who see it begin to mock him’ (Luke 14:29-30)

- (30) a. *Abie že vř přřpolovlenie prazdnřnika vřzide is*
now PTC in middle.ACC feast.GEN go-up.AOR.3.SG Jesus.NOM
vř crkř *i uřaaře*
in temple.ACC and teach.IMP.3.SG
- b. *Ēdē de tēs heortēs mesousēs*
now PTC the feast.GEN be-in-middle.PTCP.IPFV.F.GEN.SG
anebē Iēsous eis to hieron kai edidaske
go-up.AOR.3.SG Jesus.NOM in the temple.ACC and teach.IMP.3.SG
‘Now about the midst of the feast Jesus went up into the temple, and taught’
(John 7:14)

Two coordinated genitive absolutes correspond to two conjunct participles in the dative, agreeing with an oblique argument of the matrix (31).

- (31) a. *i vy podobnři člkomř* *čajřřtemř*
and 2.PL.NOM similar.M.NOM.PL man.DAT.PL wait.PTCP.IPFV.DAT.PL
ga svoego kogda vřzratitř sř otř
master.M.GEN.SG own.M.GEN.SG when come back.PRS.3.SG REFL from
brakř. *da přřsedřřju* *i*
wedding feast.GEN.PL so that come.PTCP.PFV.M.DAT.SG and
tlřknřvřřju *abie otvřzřtř sř*
knock.PTCP.PFV.M.DAT.SG immediately open.PRS.3.PL REFL
emu
3.SG.M.DAT
- b. *kai humeis homoioi anthrōpōis prosdekhomenois*
and 2.PL.NOM same.M.NOM.PL man.DAT.PL receive.PTCP.IPFV.DAT.PL
ton kurion heautōn pote
the.M.ACC.SG master.M.ACC.SG self.3.PL.M.GEN when
analusēi ek tōn gamōn
return.SBJV.AOR.3.SG from the.M.GEN.PL wedding feast.M.GEN.PL

hina elthontos kai krousantos
 so that come.PTCP.PFV.M.GEN.SG and strike.PTCP.PFV.M.GEN.SG
eutheōs anoiksōsin autōi
 immediately open.SBJV.AOR.3.PL 3.SG.M.DAT
 ‘and be like men who are waiting for their master to come home from the
 wedding feast, so that they may open the door to him at once *when he comes*
and knocks [lit. to him coming and knocking].’ (Luke 12:36)

More interesting are the mismatches in the opposite direction:

Genitive absolute	153
Accusative with infinitive	16
‘Agreeing absolutes’	9
Finite subordinate	2
Conjunct participle	1
Finite main clause	1
<i>Other/NA</i>	2

Table 1.20: Ancient Greek constructions corresponding to a dative absolute in Old Church Slavonic

One dative absolute (Mark 1.42) does not have a Greek parallel at all in Tischendorf’s edition (other versions however have a genitive absolute), whereas one (Matthew 28.1) corresponds to a Greek participle in the dative, which can be read (in both languages) as an impersonal temporal expression (Old Church Slavonic *svitajōšti*, Greek *tēi epiphōskousēi*, ‘being dawn’).

As many as 16 dative absolutes render Greek accusatives with infinitive. Most accusatives with infinitive translated into dative absolutes are found in nominalized *en tōi* + infinitive constructions (14 occurrences), as exemplified in (32). The majority of these occur in *egeneto*-clauses (9 examples), the discourse structure of which was discussed in Section 1.2.1, arguing that the temporal adverbial clause regularly following *egeneto/bystv* (and similar formulas in other languages) can be generally considered as a framing expression introducing a narrative turn.

- (32) a. *i ubošę že sę. vředřema ima*
 and fear.AOR.3.PL PTC REFL enter.PTCP.PFV.DAT.DU 3.DU.DAT
oněma vř oblakv
 that.DAT.DU in cloud.ACC.SG
- b. *ephobēthēsan de en tōi eiselthein autous eis*
 fear.AOR.3.PL PTC in the.N.DAT.SG enter.AOR.INF 3.PL.ACC in
tēn nephelēn
 the.F.ACC.SG cloud.F.ACC.SG
- ‘And they were afraid, as they entered the cloud’ (Luke 9:34)

Of the remaining accusatives with infinitive, two are subject complement clauses of *egeneto/ginetai* (33-34), which are rendered as *bysto*-clauses syntactically and functionally equivalent to the ones translating the *en tōi* + accusative with infinitive construction.

- (33) a. *i bystō v̄zležestju emu v̄*
 and happen.AOR.3.SG dine.PTCP.IPFV.M.DAT.SG 3.SG.DAT in
domu ego. i m̄nodzi mytare i
 house.LOC.SG 3.SG.GEN and many.NOM.PL tax.collector.NOM.PL and
gr̄š̄n̄nici v̄zležachō s̄v̄ ism̄v̄ i s̄v̄ uč̄eniky
 sinner.NOM.PL dine.IMPF.3.PL with Jesus.INS and with disciple.INS.PL
ego
 3.SG.GEN

- b. *kai ginetai katakeisthai auton en tēi oikia*
 and happen.PRS.3.SG dine.PRS.INF 3.SG.ACC= in the house.DAT.SG
autou, kai polloi telōnai kai hamartōloi
 3.SG.GEN and many tax.collector.NOM.PL and sinner.NOM.PL
synanekeinto tōi Iēsou kai tois mathētais
 dine.IMPF.3.PL the.DAT.SG Jesus.DAT and the.DAT.PL disciple.DAT.PL
autou
 3.SG.GEN

‘And it came to pass, as he was dining in his house, that many tax collectors and sinners also sat together with Jesus and his disciples’ (Mark 2:15)

- (34) a. *i bystō mimo chodeštju emu*
 and happen.AOR.3.SG close-by walk.PTCP.IPFV.M.DAT.SG 3.SG.DAT
isu. v̄ soboty skvozě s̄niě. i nač̄eš̄e
 Jesus.DAT on Sabbath through cornfield.GEN.PL and start.AOR.3.PL
uč̄enicy ego p̄ot̄v̄ tvoriti
 disciple.NOM.PL 3.SG.GEN way.ACC.SG make.INF
v̄str̄ogajōšte klasy
 pluck.PTCP.IPFV.NOM.PL corn.ACC.PL

- b. *Kai egeneto auton en tois sabbasin*
 and happen.AOR.3.SG 3.SG.ACC in the.DAT.PL Sabbath.DAT.PL
paraporeuesthai dia tōn sporimōn, kai
 walk.PRS.INF through the.GEN.PL cornfield.GEN.PL and
hoi mathētai autou ērxanto hodon
 the.NOM.PL disciple.NOM.PL 3.SG.GEN start.AOR.3.PL way.ACC.SG
poiein tillontes tous stachuas
 make.PRS.INF pluck.PTCP.IPFV.NOM.PL the.ACC.PL corn.ACC.PL

‘And it came to pass that he went through the cornfields on the Sabbath day and his disciples began to pluck the ears of corn as they went’ (Mark 2:23)

One dative absolute translates a nominalized *dia to* + accusative with infinitive ‘because of the *x*-ing of *y*’ (where *x* is the infinitive and *y* the agent), whose translation into

Old Church Slavonic is coincidentally also one of the very few examples of potential null-subject dative absolutes (35).

- (35) a. *se izide sějei da sěet̃. i*
 behold go.out.AOR.3.SG sow.PTCP.IPFV.M.NOM.SG to sow.SUP and
sějoštumu. ova ubo padq pri
 sow.PTCP.IPFV.M.DAT.SG some.NOM.PL then fall.AOR.3.PL along
poti.
 way.LOC.SG
- b. *idou exēlthen ho speirōn tou*
 behold go.out.AOR.3.SG the sow.PTCP.IPFV.M.NOM.SG the.N.GEN.SG
speirein. kai en tōi speirein auton
 sow.PRS.INF and in the.N.DAT.SG sow.PRS.INF 3.SG.ACC
ha men epesen para tēn
 some.N.NOM.PL indeed fall.AOR.3.PL along the.F.ACC.SG
hodon
 road.F.ACC.SG
- ‘Behold, a sower went out to sow. And as he sowed, some seed fell by the wayside; and the birds came and devoured them’ (Matthew 13:4)

The choice of dative absolute in this example may also have to do with the clause-bridging function of the adverbial in this sentence, a function which, as we will see in the next chapters (when looking at standard Early Slavic treebanks), is typical of dative absolutes outside of the Codex Marianus. The bridging function is clear from the repetition of the predicate in the immediately preceding sentence (‘sow’), with a change from foreground (a sower [went out to sow]_{foreground}) to background for a new foregrounded clause ([as he sowed]_{background} [...] _{foreground}), involving a shift in perspective or setting, in this case a change in event participants of the foregrounded event. Outside the Codex Marianus, as we will see, it is not uncommon to see dative absolutes used to bridge two discourse segments where shifts of different kinds occur (e.g. shift in location or timeframe).¹¹

9 dative absolutes correspond to a pre-matrix participle construction with an agreeing subject in the dative, as in (36-37), or accusative, as in (38) (‘agreeing absolutes’ in Table 1.20). In PROIEL, these are analysed as absolute constructions, even though the clause agrees in case with an argument of the matrix clause, because the argument is found inside the participle clause and is then repeated in the matrix clause. Of the dative absolutes translating this type of participle construction, one (Matthew 14.6) matches a

¹¹Old Church Slavonic adjectives (attributive and nominalized) and participles are inflected in either the ‘long’ or the ‘short’ form (sometimes referred to as ‘weak’ and ‘strong’ respectively, from the Germanic tradition). It is interesting that the dative absolute in this example appears in the ‘long form’, which is notably rare among dative absolutes (Lindberg 2013: 37). The difference is generally one of definiteness: old or inferable nominal referents are marked on the adjective through the long form, whereas adjectives whose referent cannot be retrieved via context or world knowledge normally appear in the short form. In (35), the expected short form would have been **sějoštu*. The choice of the long form here is likely, in fact, to be connected with the pro-dropping of the subject.

genitive absolute in New Testament versions other than Tischendorf's. The remaining 8 examples do not stand out with respect to the typical properties of dative absolutes observed so far: all except one (37) involve a verb of movement, very common among dative absolutes; they all have a third-person pronoun (6 examples) or *Jesus* (1 example) as their subject, which always follows the participle; they are all sentence-initial. They can thus be identified as typical FRAMES, bridging two events by means of motion from one scene to another.

- (36) a. *poslěďbъ že vьzležeštemъ imъ. edinuemu-na-desete*
 after PTC sit.PTCP.IPFV.DAT.PL 3.PL.DAT eleven.DAT
ěvi se
 appear.AOR.3.SG REFL
- b. *hysteron anakeimenois autois tois hendeka*
 after sit.PTCP.IPFV.DAT.PL 3.PL.DAT the.DAT.PL eleven.DAT.PL
ephanerōthē
 appear.AOR.3.SG
 'Afterward he appeared unto the eleven as they sat at meat' (Mark 16:14)
- (37) a. *i vьlězъšu emu vъ korabъ. po*
 and enter.PTCP.PFV.M.DAT.SG 3.SG.DAT in boat.ACC.SG after
nemъ idō učenici ego
 3.SG.LOC go.AOR.3.PL disciple.NOM.PL 3.SG.GEN
- b. *kai embanti autōi eis to ploion*
 and go.PTCP.IPFV.DAT.SG 3.SG.DAT in the boat.ACC.SG
ēkolouthēsan autōi hoi mathētai autou
 follow.AOR.3.PL 3.SG.DAT the disciple.NOM.PL 3.SG.GEN
 'When he got into a boat, his disciples followed him' (Matthew 8:23)
- (38) a. *išъdъšu že emu vъ vrata.*
 go.out.PTCP.PFV.M.DAT.SG PTC 3.SG.DAT in porch.ACC.PL
uzvrě drugaě. i gla imъ
 see.AOR.3.SG another.F.NOM.SG and say.AOR.3.SG 3.SG.DAT
- b. *exelthonta de auton eis ton pulōna,*
 go.out.PTCP.PFV.M.ACC.SG PTC 3.SG.ACC in the porch.ACC.SG
eiden auton allē kai legei
 see.AOR.3.SG 3.SG.ACC another.F.NOM.SG and say.AOR.3.SG
tois ekei
 that.DAT.PL there
 'When he had gone out into the porch, another [woman] saw him and told them' (Matthew 26:71)

Of the remaining dative absolutes from Table 1.20, one of the occurrences seemingly translating a conjunct participle is possibly scribal error (39), since both Greek and the *Codex Zographensis* have the 'expected' conjunct participle (*legontes* and *gl[ago]l[jošte]*).

- (39) a. *vъprosiše že i gl[agol]jōštju. učitelju.*
 ask.IMPF.3.PL PTC 3.SG.ACC say.PTCP.IPFV.M.DAT.SG master.VOC.SG
kogda ubo si bōdōtъ.
 when then this.NOM.PL be.FUT.3.PL
- b. *epērōtēsan de auton legontes · didaskale,*
 ask.IMPF.3.PL PTC 3.SG.ACC say.PTCP.IPFV.NOM.PL master.VOC.SG
pote oun tauta estai;
 when then this.NOM.PL be.FUT.3.PL
 ‘And they asked him, saying, Master, but when shall these things be?’ (Luke 21:7)

However, note that this specific usage of dative absolutes, namely as introducers of direct speech, is relatively common outside of the Codex Marianus, regardless of referential identity between the subjects of the matrix and of the participle. As we will see in the final chapter, it is, in fact, a rather common usage of switch-reference markers cross-linguistically.

The remaining apparent mismatches (two corresponding to finite temporal subordinates, Mark 4.6 and John 2.3, and one to a finite main clause, Mark 6.21) are contradicted by genitive absolutes in New Testament editions other than Tischendorf’s.

1.6 Remarks on the discourse-relation annotation on the Gospel of Luke

Another experimental layer of annotation in the PROIEL treebanks regards discourse relations combining three main frameworks for the annotation of rhetorical relations, namely the Penn Discourse Treebank 2.0 (Prasad, Dinesh, Lee, Miltsakaki, Robaldo, Joshi & Webber 2008), the RST Discourse Corpus (Carlson & Marcu 2001), and the DISCOR project (Reese, Hunter, Asher, Denis & Baldrige 2007).¹² The annotation is currently available for the whole Gospel of Luke from the Greek New Testament. A rhetorical relation holds between two discourse units, identified as those roughly corresponding to any verbal clause except those with non-predicative dependency relations such as subject, object, and nominal arguments. As in Prasad et al. (2008), the PROIEL annotation has a hierarchical structure, with five higher-level relations, each having two possible further levels of granularity describing the type of discourse relation holding between any two given discourse units.

Table 1.21 and 1.22 shows the frequency of discourse relations introduced by conjunct participles and dative absolutes.

We can immediately see that the top discourse relation introduced by conjunct participles in the Gospel of Luke is *TEMPORAL_Narration_same_story*, which most typically holds

¹²Details on this experiment were kindly provided by Hanne Eckhoff in a personal communication. The annotation is available in the official releases of the Greek New Testament in the PROIEL treebanks.

discourse relation tag	<i>n</i>	%
TEMPORAL_Narration_same story	133	31.29
TEMPORAL_Synchronous_circumstance	71	16.70
TEMPORAL_Asynchronous_precedence	67	15.76
TEMPORAL_Coincidence_means	44	10.35
TEMPORAL_Continuation_	16	3.76
TEMPORAL_Synchronous_overlap	16	3.76
CONTINGENCY_Cause_reason	14	3.29
TEMPORAL_Coincidence_manner	10	2.35
ATTRIBUTION_Content_speech content	8	1.88
COMPARISON_Concession_	7	1.64
EXPANSION_Elaboration_characterization	6	1.41
CONTINGENCY_Condition_hypothetical	5	1.17
COMPARISON_Juxtaposition_contrast	5	1.17
TEMPORAL_Narration_new story	5	1.17
CONTINGENCY_Purpose_	3	0.70
CONTINGENCY_Result_real result	3	0.70
CONTINGENCY_Cause_reason	2	0.47
COMPARISON_Juxtaposition_parallel	2	0.47
EXPANSION_Restatement_specification	2	0.47
ATTRIBUTION_Content_	1	0.23
ATTRIBUTION_Content_Non-verbal content	1	0.23
CONTINGENCY_Condition_general	1	0.23
EXPANSION_Alternative_chosen alternative	1	0.23
CONTINGENCY_Result_pragmatic result	1	0.23
COMPARISON_Sameness_	1	0.23

Table 1.21: Frequency of discourse relations tags among conjunct participles in the Codex Marianus

discourse relation tag	<i>n</i>	%
TEMPORAL_Synchronous_overlap	33	55.93
TEMPORAL_Asynchronous_precedence	15	25.42
TEMPORAL_Continuation_	5	8.47
CONTINGENCY_Cause_reason	2	3.38
TEMPORAL_Synchronous_circumstance	1	1.69
TEMPORAL_Narration_new story	1	1.69
TEMPORAL_Narration_same story	1	1.69
TEMPORAL_Coincidence_means	1	1.69

Table 1.22: Frequency of discourse relations tags among dative absolutes in the Codex Marianus

between two main clauses belonging to the same line of narration or subject matter, but also in cases such as (40). The next most frequent relation is *TEMPORAL_Synchronous_circumstance*, which is introduced by clauses such as *rubbing* in *plucked the heads of grain and ate them, rubbing them* in (5) above, repeated here as (41).

- (40) a. *onō* *že* ***prošedŕ*** *po srědě*
 that.M.NOM.SG PTC walk through.PTPC.PFV.M.NOM.SG in middle.DAT
ichŕ *iděaše*
 3.PL.M.GEN go.IMP.F.3.SG
- b. *autos* *de* ***dielthōn*** *dia mesou*
 3.SG.M.NOM PTC go through.PTCP.PFV.M.NOM.SG in middle.N.GEN.SG
autōn *eporeueto*
 3.PL.M.GEN go.IMP.F.MID.3.SG
 ‘But he walked right through the crowd and went on his way’ (Luke 4:30)
- (41) a. *i* *vŕstrŕzaachŕ* *učenici* *ego* *klasŕ.*
 and pluck.IMP.F.3.PL disciple.NOM.PL 3.SG.GEN head.of.grain.ACC.PL
i *ěděachŕ* ***istiraŕjŕše*** *rŕkama*
 and eat.IMP.F.3.PL rub.PTCP.IPFV.NOM.PL hand.INS.DU
- b. *kai etillon* *hoi mathētai* *autou* *tous*
 and pluck.IMP.F.3.PL the disciple.NOM.PL 3.SG.GEN the.ACC.PL
stachyas *kai ēsthion* ***psōchontes***
 head.of.grain.ACC.PL and eat.IMP.F.3.PL rub.PTCP.IPFV.NOM.PL
tais *chersin*
 the.F.DAT.PL hand.F.DAT.PL
 ‘And his disciples plucked the heads of grain and ate them, rubbing them in their hands’ (Luke 6:1)

These most clearly correspond to ELABORATION participles, whereas the former are most likely INDEPENDENT RHEMES.

Focussing on the tags occurring more than ten times, next among the discourse relations introduced by conjunct participles are *TEMPORAL-Asynchronous-precedence*, *TEMPORAL-Coincidence-means*, *TEMPORAL-Continuation-*, *TEMPORAL-Synchronous-overlap*, *CONTINGENCY-Cause-reason*, *TEMPORAL-Coincidence-manner*.

TEMPORAL-Asynchronous-precedence and *TEMPORAL-Synchronous-overlap* can most likely be expected to be introduced by framing expressions, the former when referring back to some time before the event time of the situation being described by the current topic time (i.e. roughly equivalent to those introduced by ‘after’-clauses or ‘when’-clauses), the latter when the time of the framing eventuality overlaps (and generally *includes*) with that of the event time of the main eventuality (i.e. roughly equivalent to those introduced by ‘while’-clauses), as in (42) and (43), respectively.

- (42) a. *i* ***vŕzvraštŕše*** *se* *apli* *povědaše*
 and return.PTCP.PFV.M.NOM.PL REFL apostle.PL.M.NOM tell.AOR.3.PL
emou *eliko* *sŕtvariše*
 3.SG.M.DAT everything make.AOR.3.PL

- b. *kai hupostrepsantes hoi apostoloi*
 and turn back.PTCP.PFV.M.NOM.PL the.PL.M.NOM apostle.PL.M.NOM
diēgēsanto autōi hosa epoiēsan
 tell.AOR.3.PL.MID 3.SG.M.DAT everything.N.ACC.PL make.AOR.3.PL
 ‘And the apostles, when they had returned, told Him all that they had done’
 (Luke 9:10)

- (43) a. *szvrzšenz že vsěkz bđetz*
 prepare.PTCP.PFV.PASS.M.NOM.SG PTC anyone.M.NOM.SG be.FUT.3.SG
ěkože oučitelz ego
 like teacher.M.NOM.SG 3.SG.M.GEN

- b. *katērtismenos de pas*
 prepare.PTCP.PRF.PASS.M.NOM.SG PTC everyone.M.NOM.SG
estai hōs ho didaskalos autou
 be.FUT.3.SG.MID like the.M.NOM.SG teacher.M.NOM.SG 3.SG.M.GEN
 ‘But anyone, when they have been fully trained, will be like their teacher’
 (Luke 6:40)

TEMPORAL_Coincidence_means and *TEMPORAL_Coincidence_manner* are other two typical functions of ELABORATIONS, which can often be read as manner or means adverbs, as in (44).

- (44) a. *mariě že szbljudaaše vsę gly*
 Mary.NOM PTC treasure.IMPF.3.SG all.M.ACC.PL word.M.ACC.PL
siję szlagajšti vč srdci
 this.ACC.PL ponder.PTCP.IPFV.F.NOM.SG in heart.N.LOC.SG
svoemz
 own.LOC.SG
- b. *hē de Maria panta sunetērei*
 the.SG.F.NOM PTC Mary.SG.F.NOM all.PL.N.ACC protect.IMPF.3.SG
ta rhemata tauta
 the.PL.N.ACC thing.PL.N.ACC this.PL.N.ACC
sunballousa en tēi kardiai
 consider.PTCP.IPFV.F.NOM.SG in the.SG.F.DAT heart.SG.F.DAT
autēs
 him.3.SG.F.GEN
 ‘But Mary treasured all these things, pondering them in her heart’ (Luke 2:19)

CONTINGENCY_Cause_reason is a discourse function which we can equally expect to be compatible with ELABORATIONS and FRAMES, though it is perhaps more typical for a participle to receive a causal reading when it follows the matrix clause, where ELABORATIONS are more common. (45) and (46) are examples of this tag on pre- and post-matrix conjunct participles respectively.

- (45) a. *i ne obrěťšę kđđ vřnesti*
 and NEG find.PTCP.PFV.M.NOM.PL where bring in.PRS.INF
i narodomř. vřzlěřřę na
 him.3.M.ACC.SG crowd.M.INST.SG climb.PTCP.PFV.M.NOM.PL on
chramř. skvozě skđelř nizřvřsiřę
 building.M.ACC.SG through tiles.SG.F.ACC tiling.AOR.3.PL
i.
 let down.3.M.ACC.SG
- b. *kai mē heurontes poias*
 and not find.PTCP.PFV.M.NOM.PL what way.SG.F.GEN
eisenegkōsin auton dia ton
 bring in .AOR.3.PL.SBJV him.3.M.ACC.SG because of the.M.ACC.SG
okhlon anabantes epi to
 crowd.M.ACC.SG go up.PTCP.PFV.M.NOM.PL on the.N.ACC.SG
dōma dia tōn keramōn
 house.N.ACC.SG through the.PL.M.GEN tiling.PL.M.GEN
kathēkan auton
 let down.AOR.3.PL him.3.M.ACC.SG
 ‘And when they could not find by what way they might bring him in because of the crowd, they went upon the housetop, and let him down through the tiling’ (Luke 5:19)
- (46) a. *i rōgaachđ se emu věđřšte ěko*
 and laugh.IMPV.3.PL REFL 3.SG.M.DAT know.PTPC.IPFV.M.NOM that
umřřřř
 die.AOR.3.SG
- b. *kai kategelōn autou eidotes hoti*
 and laugh at.IMPV.3.PL 3.SG.M.GEN know.PTCP.PRF.M.NOM.PL that
apethanen
 die.AOR.3.SG
 ‘And they laughed at Him, knowing that she was dead’ (Luke 8:53)

TEMPORAL.Continuation is likely to mostly occur between two elaborating participles headed by the same matrix clause. If some equivalence be made between this tag and the SDRT description of the discourse relation *Continuation*, then that seems to be clearly the case. According to SDRT, *Continuation* is a discourse relation

whose sole semantic content is to mark that its terms bear the same discourse relation to a dominating constituent. This implies that coordinated constituents of a sub-structure must behave in a homogeneous fashion with respect to a dominating constituent. (Asher & Vieu 2005: 595)

Most of the examples are, in fact, similar to (47), where *TEMPORAL.Continuation* holds between the coordinated elaborating participles (bracketed in the example) that follow the matrix clause.

- (47) a. *Pride* *sn̄* *člvčsky.* [*ědy*
 come.AOR.3.SG son.NOM.SG man's.NOM.M.SG eat.PTCP.IPFV.M.NOM.SG
i *p̄ije]*
 and drink.PTCP.IPFV.M.NOM.SG
- b. *elēluthen* *ho* *huios* *tou*
 come.AOR.3.SG the.NOM.M.SG son.NOM.M.SG the.GEN.M.SG
anthrōpou [*esthiōn* *kai* *p̄inōn]*
 man.GEN.M.SG eat.PTCP.IPFV.M.NOM.SG and drink.PTCP.IPFV.M.NOM.SG
 ‘The Son of Man came eating and drinking’ (Luke 7:34)

If we look at the frequency of relations among pre- (Table 1.23) and post-matrix (Table 1.24) conjunct participles, we can see that our intuitions are largely confirmed.

discourse relation tag	<i>n</i>	%
TEMPORAL_Narration_same story	129	45.58
TEMPORAL_Asynchronous_precedence	66	23.32
TEMPORAL_Coincidence_means	11	3.88
TEMPORAL_Synchronous_overlap	10	3.53
CONTINGENCY_Cause_reason	9	3.18
ATTRIBUTION_Content_speech content	8	2.82
TEMPORAL_Synchronous_circumstance	8	2.82
COMPARISON_Concession_	5	1.76
TEMPORAL_Narration_new story	5	1.76
TEMPORAL_Coincidence_manner	5	1.76
CONTINGENCY_Condition_hypothetical	5	1.76
COMPARISON_Juxtaposition_contrast	4	1.41
CONTINGENCY_Result_real result	3	1.06
TEMPORAL_Continuation_	3	1.06
CONTINGENCY_Cause_reason	2	0.70
COMPARISON_Juxtaposition_parallel	2	0.70
EXPANSION_Elaboration_characterization	2	0.70
ATTRIBUTION_Content_	1	0.35
CONTINGENCY_Purpose_	1	0.35
CONTINGENCY_Condition_general	1	0.35
EXPANSION_Alternative_chosen alternative	1	0.35
CONTINGENCY_Result_pragmatic result	1	0.35
COMPARISON_Sameness_	1	0.35

Table 1.23: Frequency of discourse relation tags among pre-matrix conjunct participles in the Codex Marianus

We can see that the vast majority of the tag *TEMPORAL_Narration_same story* occurs in pre-matrix position, where we expect most INDEPENDENT RHEMES. The same is true for *TEMPORAL_Asynchronous_precedence* and, to a lesser degree, for *TEMPORAL_Synchronous_overlap*. These we can expect to occur more frequently with pre-matrix conjunct participles functioning as FRAMES, particularly when the participle is translatable as

discourse relation tag	<i>n</i>	%
TEMPORAL_Synchronous_circumstance	63	44.36
TEMPORAL_Coincidence_means	33	23.23
TEMPORAL_Continuation_	13	9.15
TEMPORAL_Synchronous_overlap	6	4.22
TEMPORAL_Coincidence_manner	5	3.52
CONTINGENCY_Cause_reason	5	3.52
TEMPORAL_Narration_same_story	4	2.81
EXPANSION_Elaboration_characterization	4	2.81
COMPARISON_Concession_	2	1.40
CONTINGENCY_Purpose_	2	1.40
EXPANSION_Restatement_specification	2	1.40
ATTRIBUTION_Content_Non-verbal_content	1	0.70
TEMPORAL_Asynchronous_precedence	1	0.70
COMPARISON_Juxtaposition_contrast	1	0.70

Table 1.24: Frequency of discourse relation tags among post-matrix conjunct participles in the Codex Marianus

an *after*-clause (namely, when it introduces the *TEMPORAL_Asynchronous_precedence* relation). Among post-matrix conjunct participles, on the other hand, *TEMPORAL_Synchronous_circumstance* is by far the most common and overwhelmingly more frequent than in pre-matrix position. *TEMPORAL_Coincidence_means* and *TEMPORAL_Continuation_*, other two relations which were strongly expected to occur most frequently with ELABORATIONS are also much more common among post-matrix participles.

Finally, looking back at Table 1.22, we can see that the relations *TEMPORAL_Synchronous_overlap* and *TEMPORAL_Asynchronous_precedence* are by far the most common among dative absolutes, which fully agrees with their typical functions as FRAMES.

1.7 Summary

Dative absolutes in the Codex Marianus were found to align strongly with the properties of Bary & Haug’s (2011) FRAMES regardless of syntactic configuration, which supports the observations made in previous literature (e.g. by Worth 1994: 30; Corin 1995: 259; Collins 2011: 113) that the unifying function of dative absolutes is a ‘backgrounding’ or ‘stage-setting’ one. The typical framing function of absolutes was immediately supported by their neat predominance in pre-matrix position, indicating that their placement in the sentence is in itself connected to their discourse function. Following Bary & Haug’s (2011) account of framing participles, and similarly of Chafe’s (1976) and Krifka’s (2007) frame-setters, FRAMES are, in fact, typically found in the leftmost position in the sentences, since they can be modeled as fronted adjuncts in topicalized position and they set the stage for the whole sentence.

Conjunct participles, on the other hand, were found to clearly fulfil different discourse functions depending on syntactic configuration, aspect of the participle, and position of the subject (when overt). Similarly to Bary & Haug's (2011) and Haug's (2012) analysis of Ancient Greek conjunct participles, a highly significant association was found between the perfective aspect and pre-matrix position imperfective aspect and post-matrix position. This was taken as indirect evidence that pre-matrix conjunct participles in Early Slavic are more likely to function as INDEPENDENT RHEMES, which are similar to independent clauses from the discourse perspective, whereas post-matrix conjunct participles are more likely to function as ELABORATIONS, which are temporally dependent on the matrix verb (i.e. they never introduce new temporal referents and they never induce narrative progression) and typically express manner or means/instrument.

In sentence-initial position, even when perfectives, conjunct participles can be ambiguous between a FRAME and an INDEPENDENT RHEME interpretation. We examined the anaphoric properties of sentence-initial conjunct participles and observed a significant connection between the subject's position relative to the participle and the average distance between the subject and its immediate antecedent. Subjects in the VS configuration tend to have, on average, antecedents that are significantly farther away compared to subjects in the SV configuration, which agrees with findings from Haug (2012) on Ancient Greek data. The VS configuration in sentence-initial conjunct participles was attributed to the reintroduction of old or inactive referents, a function which can be associated with the use of FRAMES to explicitly signal a shift in the event participants or their information status from the previous discourse.

The subjects of dative absolutes, albeit very similar to sentence-initial VS conjunct participles in terms of average distance of the immediate antecedents and overall referential activation in the previous discourse, were found to be represented predominantly by inherently anaphoric parts of speech, whereas among VS conjunct participles more explicit ones (e.g. proper nouns) were found to be the most common. A possibility was advanced that dative absolutes predominantly attract few, very prominent discourse referents (e.g. Jesus or the Apostles, in the case of the Gospels), whereas framing conjunct participles may be frequently used with any salient referent.

Annotation on Aktionsart on the Greek New Testament also allowed us to check which Aktionsarten more commonly occur among the parallels to dative absolutes and conjunct participles. Telic verbs are the most common among both constructions, although more prominently so among conjunct participles, whereas activity verbs are markedly more common among dative absolutes. This is also in line with the typical functions of the constructions observed through other variables: activity verbs are intuitively more easily used in framing constructions, whereas telic verbs more readily so for series of events with narrative progression.

The analysis of discourse-relation annotation on the Gospel of Luke largely confirmed

these patterns, showing that rhetorical relations intuitively more compatible with INDEPENDENT RHEMES occur more frequently among pre-matrix conjunct participles, those more compatible with FRAMES equally so among dative absolutes and pre-matrix conjunct participles, while those more compatible with ELABORATIONS are most frequent among post-matrix conjunct participles.

Finally, Old Church Slavonic-Greek parallel data allowed us to single out potential mismatches between the two languages and to corroborate or refine observations made on the basis of largely overlapping patterns. Mismatches between absolute constructions have highlighted two main groups of independently used dative absolutes, those rendering nominalized accusatives with infinitive and those translating ‘agreeing absolutes’ (i.e. pre-matrix participles constructions with a separate subject, like absolutes, but with the whole close agreeing with a matrix argument, like conjunct participles), in both cases clearly functioning as framing participles. Among conjunct participles, two main groups of mismatches can be most clearly identified, namely Greek main verbs which get translated into conjunct participles (of which those corresponding to a Greek imperative are a particularly clear sign of independent Slavic usage) and free rendition of Greek phrases other than participles into a (particularly elaborating) conjunct participle.

The next chapter replicates the analysis of dative absolutes and conjunct participles carried out so far on a much larger and diverse dataset, leveraging these findings as a ‘blueprint’ to obtain further insights into the constructions’ functions across different texts and varieties.

Chapter 2

Participle constructions in standard treebanks

2.1 Introduction

This chapter looks into the properties of conjunct participles and dative absolutes in a much larger dataset containing morphosyntactic and dependency tagging, but no information-structural annotation, such as givenness status and anaphoric links from nominal referents to their antecedents. It is a relatively diverse subcorpus, in terms of geographic and historical varieties represented. The first and main part of the analysis uses all the treebanks from the overview in the Introduction, except for those referred to as *strategically annotated* (i.e. fully automatically annotated) and the Codex Marianus, the latter used in the previous chapter. The second part of the analysis looks more briefly at the constructions in strategically annotated treebanks, which contain texts written in South Slavic varieties not well represented in the first part of the analysis.

The analysis in this chapter follows the same method adopted in Chapter 1 but for the parts concerning information status. The overarching goal is to compare the overall behaviour of the constructions in the Codex Marianus, representing one of the oldest and most deeply annotated Old Church Slavonic text—with aligned Greek parallels—, with that observed in texts of different Early Slavic varieties and genres, including non-religious or non-liturgical texts (most notably, original Slavic texts).

2.2 Dative absolutes and conjunct participles in standard treebanks

Following the extraction criteria outlined in Section 0.5.1 of the Introduction, the final dataset resulted in 1,534 dative absolutes and 8,232 conjunct participles.¹ As in Chapter 1, the analysis is divided into groups of variables expected to capture the relevant properties of the constructions. Section 2.2.1 looks at the position of participle

¹Query performed in December 2022.

constructions relative to the matrix clause and at the distribution of imperfective and perfective occurrences, Section 2.2.2 at the properties of realized subjects, and Section 2.2.3 at the lexical variation among the constructions in different subcorpora.²

2.2.1 Position and aspect distribution

As Table 2.1 shows, dative absolutes are more frequent in pre-matrix position (67.4%), which is what we would expect given the clear framing function that emerged from the analysis in Chapter 1. However, as many as 13.9% of the occurrences follow the matrix. This indicates that, outside the limited narrative style represented by the Gospels, the matrix-dative absolute configuration is not as uncommon as the first dataset suggested.

The overall position of conjunct participles, as in the Gospels, does not in itself serve as a strong indication of particular functions. The pre-matrix position is somewhat more frequent (59.9%), which is also to be expected, given the double function of conjunct participles in that position as *FRAMES* and *INDEPENDENT RHEMES*. As many as 40% are instead found in post-matrix position. Based on the previous Chapter, these can be expected to mainly be *ELABORATIONS*, which, as we will see, seems to normally be the case.

No clear matrix clause was identified for 18.7% of dative absolutes and for 0.1% of conjunct participles (*NA* in Table 2.1) using the extraction method outlined in Section 0.5.1 of the Introduction.

	pre-matrix	post-matrix	NA
absolute	67.4% (1034)	13.9% (213)	18.7% (287)
conjunct	59.9% (4932)	40% (3293)	0.1% (7)

Table 2.1: Relative order of participle and matrix clause in standard treebanks

Most of the occurrences of dative absolutes counted under *NA* are connected to the matrix clause by a coordinating conjunction (*i* ‘and’ or *a* ‘but, whereas’), in which case, as explained in the Introduction, the annotation convention in PROIEL/TOROT is to add an empty verb node to stand in for the matrix clause. In such cases, purely for the purpose of the analysis in this chapter, the matrix verb is identified with the empty node itself, in order to make ‘non-canonical’ occurrences of participle constructions more easily retrievable.³ The remainder of the occurrences for which no information on the main verb was immediately available are absolutes lacking a clearly identifiable matrix clause, to which they could be attached in the same sentence. In many of these examples, they introduce reported speech, a usage examined by Collins (2011: 113-122) in his broader analysis of ‘syntactically independent’ dative absolutes in Early Slavic. As Collins notes,

²Note that ‘(Old) Church Slavonic’ will be used to refer together to Old Church Slavonic and later Church Slavonic texts from this dataset (i.e. excluding the *Codex Marianus*) unless otherwise specified.

³That is, these occurrences are by no means necessarily to be analysed as ‘true’ elliptic constructions.

- (2) *to že uslyšavъ*
 that.M.ACC.SG PTC listen.PTCP.PFV.M.NOM.SG
bodъchnovenyi ocb našb
 filled.with.the.Holy.Spirit.M.NOM.SG father.M.NOM.SG our.NOM.M
feodosii osklabivъ sja licъmъ i malo
 Feodosij.NOM.M smile.PTCP.PFV.M.NOM.SG REFL face.INST.SG and little
proslъzivъ sja <spověda> tomu
 start.crying.PTCP.PFV.M.NOM.SG REFL <tell.AOR.3.SG> that.DAT.M.SG
 ‘Then after listening, our father Feodosij, filled with the Holy Spirit, smiling with his face and tearing up a little <told> him’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 46b)

One occurrence is instead an overtly subordinated participle in a predicative complement clause introduced by *jako* ‘that’ (3).

- (3) *tgda že ocb našb feodosii*
 then PTC father.M.NOM.SG our.1.M.NOM.SG Feodosij.M.NOM.SG
napълnivъ se stgo dúcha
 fill.PTCP.PFV.M.NOM.SG REFL holy.SG.M.GEN spirit.SG.M.GEN
načatъ togo obličati jako nepravedъno
 begin.AOR.3.SG that.SG.M.GEN reproach.INF that unfairly.SG.N.ACC.STRONG
stvoriвъša i ne po zakonu
 act.PTCP.PFV.M.GEN.SG and NEG according to law.M.DAT.SG
sědъša na stolě tomъ i jako
 sit.PTCP.PFV.M.GEN.SG on throne.SG.M.LOC that.SG.M.LOC and that
ocę si i brat
 father.SG.M.GEN this.3.M.DAT.SG and brother.SG.M.GEN
starěišago progъnavъša
 oldest.SG.M.GEN.COMP banish.PTCP.PFV.M.GEN.SG
 ‘At the same time, our father Feodosij, filled with the Holy Spirit, began to reproach the prince that he had acted unfairly, that he had not sat on the throne lawfully, and that he had expelled his father and older brother’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 58b)

The usage of a conjunct participle introduced by *jako* in contexts such as the one in (3) is not really at odds with the function of conjunct participles in their typical configuration as INDEPENDENT RHEMES or FRAMES (pre-matrix, juxtaposed, mostly perfective). The participles in (3) describe a series of events occurring one after the other in a chain-like fashion and presuppose that they collectively lead to some main situation (the current one in which Feodosij reproaches the prince).

A closer look at potential post-matrix dative absolutes reveals that a few of these occurrences are either pre-matrix from the discourse perspective (e.g. *bystъ*-clauses in (4); cf. Section 1.2.1 of Chapter 1) or augmented with overt subordinating conjunctions, such as *jako* ‘as, since’ or *ašte li* ‘if’,⁴ which, as the syntactic head of the participle,

⁴The occurrence of absolute constructions introduced by a connective is obviously not unheard of (see in particular Stump 1985, and König 1991 on English *with*-augmented absolutes).

were extracted by the query as the potential matrix of the construction (thus resulting as post-matrix from the raw dataset). Some absolutes introduced by a subordinating conjunction do indeed follow the matrix (5), whereas others precede it (6):

- (4) *i byst̃ bes̃duj̃ošt̃ema ima i*
 and happen.AOR.3.SG converse.PTCP.IPFV.DAT.DU 3.DU.M.DAT and
s̃ṽz̃praš̃aj̃ošt̃ema sẽ čto byṽš̃ee i čto
 discuss.PTCP.IPFV.DAT.DU REFL what be.PTCP.PFV.N.ACC.SG and what
b̃ođošt̃eje i sam̃o sp̃s̃o
 be.PTCP.FUT.N.ACC.SG and self.NOM saviour.NOM
p̃ribliž̃ sẽ id̃ěše s nima
 approach.PTCP.PFV.M.NOM.SG REFL go.IMP.F.3.SG with 3.DU.INST
 ‘And it came to pass that, as they were conversing and discussing what had been and what would be, the Saviour himself approached them and started walking with them’ (John Chrysostom, *Homily on Christ’s resurrection*, Codex Suprasliensis f. 103v)

- (5) *i plakati sẽ nač̃a popad̃bja jako mertvu*
 and cry.INF REFL begin.AOR.3.SG priest’s-wife.NOM as dead.DAT
sušt̃ju onomu
 be.PTCP.IPFV.M.DAT.SG that.M.DAT
 ‘And the priest’s wife started to weep, as if he were dead’ (*Primary Chronicle*, Codex Laurentianus f. 88b)

- (6) *ašte li sice. izvolšu bgu. ... da budet*
 if Q thus desire.PTCP.PFV.M.DAT.SG god.DAT ... VOL be.FUT.3.SG
volę gnę.
 will.F.NOM.SG lord.ADJ.F.NOM.SG
 ‘It thus God desires ... may Lord’s will be done’ (*Life of Sergij of Radonezh* f. 59v)

Overall, however, *byst̃*-clauses and augmented absolutes are very few compared to the total number of occurrences (15 *byst̃*-clauses and 8 augmented absolutes), so that the vast majority of post-matrix absolutes in Table 2.1 should be considered genuine occurrences following their matrix clause.

As in the Gospels, by crossing aspect with position in the sentence, we obtain clearer functional patterns, as Table 2.3 shows.

	imperfectives	perfectives
pre-matrix	45.8% (474)	54.2% (560)
post-matrix	76.1% (162)	23.9% (51)
<i>NA</i>	50.9% (146)	49.1% (141)
tot.	51% (782)	49% (752)

Table 2.2: Dative absolutes in standard treebanks: aspect distribution by position relative to the matrix clause (row percentage)

	imperfectives	perfectives
pre-matrix	27% (1334)	73% (3598)
post-matrix	76.5% (2518)	23.5% (775)
NA	57.1% (4)	42.9% (3)
tot.	46.8% (3856)	53.2% (4376)

Table 2.3: Conjunct participles in standard treebanks: aspect distribution by position relative to the matrix clause (row percentage).

Just like in the Gospels, the association between aspect and position is statistically significant among conjunct participles, with moderate to strong effect size ($\phi = 0.48$, Cramér's V): $\chi^2(1) = 1924.52$, $p < 0.01$, with much higher odds for perfectives to occur in pre-matrix than post-matrix position and for imperfectives to occur in post-matrix than pre-matrix position (odds ratio: 8.69). This indicates that their usage as INDEPENDENT RHEMES, as in (7), and ELABORATIONS, as in (8), is overall very similar to that found in the Codex Marianus.

- (7) *drakula* *že* *vostavъ* *vzem*
 Dracula.M.NOM.SG PTC stand.PTCP.PFV.M.NOM.SG take.PTCP.PFV.M.NOM.SG
mečb *svoi* *iskoči* *s* *polaty*
 sword.SG.M.ACC his.ACC.SG jump out.AOR.3.SG from palace.F.GEN.SG
 ‘Then Dracula stood up, took his sword and jumped out of the palace’ (*The Tale of Dracula*, f. 215v)

- (8) *ekove* *brate* *ebi* *ležę*
 Jakov.VOC brother.VOC.SG screw.IMP.2.SG lie down.PTCP.IPFV.M.NOM.SG
ebechoto aesovo
 horny ballser.VOC.SG
 ‘Brother Jakov, do your screwing lying down, you horny ballser!’ (Birchbark letters St. R. 35, translation by Schaeken 2018a)

Among dative absolutes, perfectives and imperfectives are overall almost equally frequent (49% and 51% respectively). Among pre-matrix occurrences, perfectives are somewhat more frequent, whereas among post-matrix ones, imperfectives predominate. Also in this case there is a statistically significant association between aspect and position in the sentence, but with a much lower effect size ($\phi = 0.25$, Cramér's V): $\chi^2(1) = 76.88$, $p < 0.01$, with higher odds for perfectives to occur in pre-matrix than post-matrix position and for imperfectives to occur in post-matrix than pre-matrix position (odds ratio: 4.07). However, because of the very different numbers of pre- and post-matrix absolutes (1034 versus 213), the raw number of imperfective absolutes is still higher among pre-matrix than post-matrix absolutes, which means that a test of independence is perhaps not the most suitable in this case. Among dative absolutes to the left of the matrix, we can

see that the proportion of the two aspects is relatively even, in line with the pattern found in the Gospels. This was explained in Section 0.2.2 of the Introduction, following Haug (2012: 311-312), with the fact that the dominant aspect of framing participles is expected to vary with genre – unlike conjunct participles, whose aspect was shown to strongly correlate with their position. The difference in number between imperfectives and perfectives in pre-matrix position is, in fact, not statistically significant in itself.⁵ Among dative absolutes to the right of the matrix, however, imperfectives are clearly more frequent. Recall that for the Gospels no generalizations about post-matrix absolutes were possible, because of the very low number of occurrences in that position compared to pre-matrix absolutes. This time we observe a significantly higher number of imperfective than perfective post-matrix dative absolutes.⁶ *Semantically*, it appears that three main groups of post-matrix dative absolutes could be identified, exemplified by (9)-(11).

- (9) *on* *že* *šed* *sěde* *na*
 that.M.NOM.SG PTC go.PTCP.PFV.M.NOM.SG sit.AOR.3.SG on
stolě *černigově* *jaroslavu*
 throne.SG.M.LOC Chernigov.SG.M.LOC Yaroslav.M.DAT.SG
suštju *nověgorodě* *togda*
 be.PTCP.IPFV.M.DAT.SG Novgorod.SG.M.LOC then
 ‘He thus departed thence and established himself upon the throne of Chernigov,
 while Yaroslav was at Novgorod.’ (*Primary Chronicle*, Codex Laurentianus f. 50b)
- (10) *i* *chodi* *po* *manastyrem*. *i* *ne* *vzljubi*
 and go.AOR.3.SG about monastery.M.DAT.PL and NEG love.AOR.3.SG
bu *ni* *choteštju*
 God.M.DAT.SG NEG want.PTCP.IPFV.M.DAT.SG
 ‘He went about the monasteries and liked none of them, since God did not so will’
 (*Primary Chronicle*, Codex Laurentianus f. 53a)
- (11) *i* *oněm* *že* *stavšem* *na* *nošbněm*
 and that.M.DAT.PL PTC stay.PTCP.PFV.M.DAT.PL in night.SG.N.LOC
stanovišti *blaženyi* *že* *ne* *doida*
 camp.SG.N.LOC blessed.M.NOM.SG PTC NEG arrive.PTCP.IPFV.M.NOM.SG
jako *i* *zvrěimo* *ich* *tu* *že* *opočivaaše*
 as PTC distance.SG.N.ACC them.GEN here PTC rest.IMP.3.SG
jedinomu *bogu* *sbljudajuštju* *i*
 alone.M.DAT.SG God.M.DAT.SG watch over.PTCP.IPFV.M.DAT.SG 3.SG.M.ACC
 ‘And when they stayed overnight, the blessed, stopping to see them from afar,
 spent the night here, and God alone guarded him.’ (*Life of Feodosij Pečerskij*,
 Uspenskij sbornik f. 31a)

The first, and most common, exemplified by (9), involves absolutes receiving a temporal interpretation and adding the circumstances in which the main eventuality occurs,

⁵Binomial test, $p = 0.99$

⁶Binomial test, $p < 0.01$.

specifically with respect to an event participant outside of the foreground. As we will see, temporally, these should still be considered on a par with what we have called FRAMES so far. The second group, exemplified by (10) involves those receiving a causal interpretation. *Temporally*, these are also FRAME-like. The reason why neither of these examples can be considered an ELABORATION is that ELABORATIONS are temporally dependent on the matrix clause, whose runtime they inherit, and enforce a reading whereby the runtime of the matrix may *include* or, more often, is equivalent to, that of the participle. Both FRAMES and ELABORATIONS, when imperfective, indicate that the participle event goes on for the whole runtime of the matrix event, but the opposite, namely that the main event holds for the whole runtime of the participle event, is only true for ELABORATIONS (Haug 2012: 296). Note the difference between (12a), a FRAME, and (12b), an ELABORATION.

- (12) a. *Pondering over what had happened*, he finally realized what went wrong
 b. He walked through the park *pondering over what had happened*

The states introduced by the predicates ‘be in Novgorod’ in (9) and ‘(not) will’ in (10) do not necessarily only last as long as the main eventuality does. In both cases, in fact, a reading in which the event time of the matrix clause is *included* in that of the absolute seems more likely.

The third example, however, is quite different from the other two and is more compatible with ELABORATIONS. *Sъbljudajuštju* ‘guarding’ in (11) clearly uses the matrix event time as its reference time. The event time of *spent the night here* lasts for the whole event time of *God alone guarding him*, but also, crucially, the whole event time of *God alone guarding him* lasts for the whole event time of *spent the night here*. Intuitively, the event time of *God alone guarding him* does not necessarily cover the preceding discourse within the same sentence, namely *stopping to see from afar*, but rather forms a complex predicate only with the matrix clause.

In some examples, the ELABORATION function of the dative absolute is unmistakable, as in (13).

- (13) *potomъ* *vъzmо̄štajetъ* *sę* *i* *vъskypitъ* *zъbomъ*
 afterwards shake.PRS.3.SG REFL and boil.PRS.3.SG tooth.M.DAT.PL
zybъjо̄štemъ *sę* *žilamъ*
 become loose.PTCP.IPFV.M.DAT.PL.STRONG REFL vein.PL.F.DAT
sъgrъždajо̄štemъ *sę* *i* *vъsemu*
 become contracted.PTCP.IPFV.F.DAT.PL REFL and whole.SG.N.DAT
tělu *nevolejо̄* *sъlęcajо̄štu* *sę*
 body.SG.N.DAT discomfort.F.INST.SG twist.PTCP.IPFV.N.DAT.SG REFL
 ‘Afterwards, he was shaking and boiling up, the teeth going loose, the veins becoming contracted, the whole body twisting in discomfort’ (*Encomion on the 40 Martyrs of Sebasteia*, Codex Suprasliensis 6)

In this example, the absolutes clearly add granularity to the main verbs, with the temporal effect that the running time of the matrix and of the participles is identical.

A small group of post-matrix imperfective absolutes may receive a contrastive or adversative reading—a semantic effect which strikes a clear chord, cross-linguistically, with the double function of temporal connectives (expressing simultaneity) as adversative conjunctions (cf. English *while*, as in *he went out with friends, while I stayed at home*), particularly in postponed position (cf. Mauri & Ramat 2012 and Musi 2016). (14) and (15) are two such examples.

- (14) *i povelě prinesti svěšte*
and command.AOR.3.SG bring.INF candle.PL.F.ACC
gorěšte. i přižagati lice
burn.PTCP.IPFV.F.ACC.PL and burn.INF face.SG.N.ACC
paule. propovědniku vřpíjoštu i
Paul's.SG.N.ACC.STRONG preacher.M.DAT.SG cry out.PTCP.IPFV.M.DAT.SG and
glagološtu
say.PTCP.IPFV.M.DAT.SG
'And he commanded to bring burning candles and burn Paul's face, while the preacher was crying out, saying' (*Vita of Paul and Juliana*, Codex Suprasliensis f. 7r)

- (15) *togo že styi priimō*
that.SG.M.GEN PTC saint.M.NOM.SG take.PTCP.PFV.M.NOM.SG
prěpodobnyima svoima rōkama pojemle
holy.F.INST.DU his.3.F.INST.DU hand.F.INST.DU take.PTCP.IPFV.M.NOM.SG
maslo otō kandila mazaše jemu
oil.SG.N.ACC from vigil lamp.SG.N.GEN smear.IMP.3.SG 3.SG.M.DAT
vse tělo kažuštu
whole.SG.N.ACC body.SG.N.ACC show.PTCP.IPFV.M.DAT.SG
boleštuumu i naznamenujoštu vō
be sick.PTCP.IPFV.M.DAT.SG and point.PTCP.IPFV.M.DAT.SG in
koemō městě plzasaše dchō oky
which.SG.N.LOC place.SG.N.LOC creep.IMP.3.SG soul.M.NOM.SG like
zmija vgněždaē se v nēmō
serpent.F.NOM.SG nest.PTCP.IPFV.M.NOM.SG REFL in 3.SG.M.LOC
'After taking him with his holy hands, the saint took oil from the vigil lamp and started smearing his whole body, while the sick was showing and pointing in which place the soul was creeping like a serpent nesting itself in him' (*Vita of St. Aninas the Wonderworker*, Codex Suprasliensis f. 150r)

Similarly, some examples receive a conditional interpretation, as in (16).

- (16) *nō i tako ne priimōtō blagoděti židove*
but even so NEG receive.PRS.3.PL grace.GEN.SG Jew.M.NOM.PL.
mōně ne soštu tu
1.SG.M.DAT NEG be.PTCP.IPFV.M.DAT.SG here

‘Jews will not receive grace while I am not here’ (John Chrysostom, *Homily on the resurrection of Lazarus after four days*, Codex Suprasliensis f. 23v)

Both cases such as (9), and those such as (14)-(15) and (16) would be considered by Fabricius-Hansen & Haug (2012b) as temporally and semantically equivalent to fronted, frame-setting adjuncts, even though they occur in post-matrix position. As discussed in Section 0.2.3 of the Introduction, they can be interpreted as *restrictive* adverbial adjuncts, since, unlike typical INDEPENDENT RHEMES and ELABORATIONS, they restrict the domain of the matrix VP by introducing a *state* within which the main eventuality holds. If we consider the similar semantic contribution of (mildly or strongly contrastive) post-matrix restrictive adjuncts and frame-setters, we can see how dative absolutes may have a fairly consistent function regardless of position in the sentence and semantic effects (such as contrastivity and causality).

The interpretation of imperfective dative absolutes as ELABORATIONS in post-matrix position is, however, clearly possible (as (13) showed), and may to some extent be compositionally driven, given the correlation between imperfective aspect and post-matrix position. Table 2.4 shows the number of imperfective post-matrix dative absolutes which *could* be interpreted as ELABORATIONS on the basis of temporal relations between the matrix and the absolutes, versus those that cannot. Those with *byti* in the aorist as the matrix verb are excluded from the count, as they are all *bystv*-clauses, where the absolute is pre-matrix from the discourse perspective. Non-ELABORATIONS largely have a FRAME-like temporal semantics (i.e. they do not depend on the matrix clause temporally and the event time they introduce includes that of the matrix clause) and can be mostly interpreted as clause-final restrictive adjuncts. These, as already discussed, can be semantically equivalent to fronted adjuncts, so they are referred to as FRAMES, even if they are not fronted.⁷

function	<i>n</i>
ELABORATION	16.2% (23)
FRAME	83.8% (119)

Table 2.4: ELABORATIONS versus FRAME imperfective absolutes in post-matrix position

Overall, post-matrix dative absolutes functioning as FRAMES are much more common than those serving as clear ELABORATIONS. Many ELABORATION dative absolutes involve verbs such as *pomagati* ‘help’, *pospěšbestvovati* ‘contribute, help’, *sobljudati* ‘guard, watch over’, *ukrěpljati* ‘give strength’, with their subject being invariably *bogv* ‘God’. These are rather clear examples of ELABORATIONS receiving a *manner*, *comitative* or *instrument* reading (e.g. ‘with the help of God’), as in (17), which is quite typical of elaborating adverbial clauses as a whole (cf. Fabricius-Hansen & Haug 2012b).

⁷The numbers are based on my own annotations, without the opinion of a second annotator, as is standard practice when annotating a corpus, particularly of a dead language, so one should be cautious when evaluating them. The dataset with the contexts tagged as ELABORATION or FRAME can be found in the repository: <https://doi.org/10.6084/m9.figshare.24166254>.

- (17) *i to rekъ paki po pervoe*
 and that.N.ACC.SG say.PTCP.PFV.M.NOM.SG more on righteous.N.ACC.SG
držžaše se dobroe ustroenie. bogu
 stay.IMPF.3.SG REFL good.N.ACC.SG behaviour.SG.N.ACC god.SG.M.DAT
pomagajuštu emu na blagoe proizvolenie
 help.PTCP.IPFV.M.DAT.SG 3.SG.M.DAT on good.N.ACC.SG intention.N.ACC.SG
 ‘After saying this, he stayed on the righteous path more than before, with God
 helping him in his good intentions’ (*Life of Sergij of Radonezh* f. 46v)

As Table 2.5 shows, examples of dative absolutes as ELABORATIONS are relatively evenly distributed across varieties if we consider the proportions of FRAME-like and ELABORATION dative absolutes. In raw terms, *post-matrix* absolutes may seem to be more frequent in non-translated texts and East Slavic varieties, but these are, as already mentioned, over-represented in this dataset to start with.

	ELABORATION	FRAME
East Slavic	15	73
South Slavic	8	46
Old East Slavic	8	64
Early Middle Russian	6	9
Late Middle Russian	1	0
Old Church Slavonic	8	46
Later Church Slavonic	0	0
translations	8	46
originals	15	73
Codex Suprasliensis	8	45
<i>Domostroj</i>	1	0
<i>Life of Sergij of Radonezh</i>	6	9
<i>Psalterium Sinaiticum</i>	0	1
<i>First Novgorod Chronicle</i> (Synodal)	0	3
<i>Primary Chronicle</i> (Hypthianus)	1	0
<i>Primary Chronicle</i> (Laurentianus)	1	25
<i>Suzdal Chronicle</i> (Laurentianus)	4	14
Uspenskij Sbornik	2	22

Table 2.5: Frequency of ELABORATIONS and FRAME imperfective absolutes by variety, text (or manuscript), and by translated versus original Early Slavic texts. Only manuscripts in which at least one ELABORATION is found are indicated. Varieties are listed regardless

Summary

As in deeply annotated treebanks, the analysis of conjunct participles indicated a strong association between their position relative to their matrix clause and their aspect, pointing to their overall typical functions (already established in the previous chapter) as INDEPENDENT RHEMES in pre-matrix position, where they are much more likely to be perfective, and as ELABORATIONS in post-matrix position, where they are much more likely to be imperfective. Post-matrix dative absolutes were found to be less frequent

(13.9%) than those in pre-matrix position, but not as uncommon as post-matrix absolutes in the Codex Marianus (where they only constituted 7.1% of the occurrences). Given the higher number of post-matrix absolutes than those extracted from the first dataset, we were able to look for potential associations between position in the sentence and other variables. We found some (significant, but not as strong as with conjunct participles) association between the perfective aspect and the pre-matrix position, and between the imperfective and the post-matrix position. We therefore looked more closely into post-matrix occurrences and observed that three main groups of post-matrix absolutes could be distinguished semantically, each with potential differences in their most likely temporal interpretation.

The first and most common group is made of absolutes receiving a primary temporal interpretation and clearly functioning as FRAMES despite being found in post-matrix position (e.g. *He thus departed thence and established himself upon the throne of Chernihiv, **while Yaroslav was at Novgorod***). Several of these occurrences may also receive a (mildly or strongly) contrastive or conditional reading (e.g. *And he commanded to bring burning candles and burn Paul's face, **while the preacher was crying out**, saying or *The Jews will not receive Grace **if I'm not here****). These can be interpreted as clause-final restrictive adjuncts, which, as Fabricius-Hansen & Haug (2012b) explain and as we mentioned in Section 0.2.3 of the Introduction, often receive contrastive readings and correspond semantically to frame-setting fronted adjuncts, in that they restrict the domain of the matrix clause by introducing a *state* within which the main eventuality holds.

The second group are absolutes receiving a causal interpretation (e.g. *He went about the monasteries and liked none of them, **since God did not so will***), which can also be considered similar to FRAMES from the *temporal* perspective, since their runtime may *include* the runtime of the matrix clause, which may be considered incompatible with an ELABORATION interpretation, since with ELABORATIONS the runtime of the matrix clause and of the participle clause are identical.

The third and final main group of post-matrix absolutes are those that can actually be interpreted as ELABORATIONS. In such cases, we may get the same temporal identity we observe with elaborating conjunct participles (e.g. *Afterwards, he was shaking and boiling up, the teeth **going** loose, the veins **becoming** contracted, the whole body **twisting** in discomfort*). A manual annotation of post-matrix occurrences indicated that about 16% of post-matrix absolutes *can* receive an ELABORATION interpretation, but as many as 83.8% of them are still compatible with the temporal semantics of FRAMES despite being in post-matrix position. Furthermore, several elaborating dative absolutes are represented by almost-fixed expressions of the type ‘with God helping PERSON’ and occurring with verbs such as *szbljudati* ‘guard, watch over’, *pomagati* ‘help’, *ukrępljati* ‘give strength’, with their subject being invariably *bogŭ* ‘God’ (e.g. *And when they stayed overnight, the blessed, stopping to see them from afar, spent the night here, with **God alone guarding him***).

2.2.2 Properties of subjects

Table 2.6 shows the frequency of overt and null subjects among sentence-initial conjunct participles and dative absolutes.

	overt	null
conjunct (sentence-initial)	50.8% (1564)	49.2% (1512)
absolute (sentence-initial)	94.4% (656)	5.6% (39)

Table 2.6: Overt and null subjects in sentence-initial conjunct participles and dative absolutes in standard treebanks (row percentage)

As in the previous dataset, the vast majority of sentence-initial dative absolutes have an overt subject, while among sentence-initial conjunct participles overt and null subjects are virtually equally frequent.

When we look at the position of the overt subject relatively sentence-initial conjunct participles and dative absolutes (Table 2.7), we see that the VS configuration is not predominant as in the Codex Marianus: in fact, it is slightly less frequent than the SV one.

	SV	VS
conjunct (sentence-initial)	65.2% (1020)	34.8% (544)
absolute (sentence-initial)	52.4% (344)	47.6% (312)

Table 2.7: Position of overt subjects relative to sentence-initial conjunct participles and dative absolutes in standard treebanks (row percentage)

As Table 2.8 shows, over half of the subjects of dative absolutes are represented by parts of speech which are either inherently anaphoric (personal and demonstrative pronouns) or likely to encode old or accessible referents (proper nouns).

part of speech	frequency (%)
<i>null</i>	<4.4
common nouns	38.5
personal pronouns	25
proper nouns	15.1
demonstrative pronouns	9.9
adjectives	3.1
indefinite pronouns	2.1
verbal nouns	0.8
cardinal numerals	0.6
ordinal numerals	0.2
relative pronouns	0.1
quantifier	0.1
interrogative pronouns	<0.1
possessive pronouns	<0.1

Table 2.8: Dative absolutes in standard treebanks: subject parts of speech

In comparison to the Gospels, subject parts of speech are much more varied (13 parts of speech, against 5 in the Gospels). However, many of these occur relatively rarely, and the only four standing out are common nouns, personal pronouns, proper nouns, and demonstratives. The relative order of the most frequent parts of speech in the two syntactic positions, namely pre- (Table 2.9) and post-matrix (Table 2.10), is the same as for dative absolutes as a whole. Common nouns have a higher frequency among post-matrix absolutes than among pre-matrix ones, while the opposite is true for demonstratives. This could be due to the fact that pre-matrix participles are more likely to have their subject referring back to the immediately preceding sentence, whereas in post-matrix position, more explicit reference to the subject may be used to avoid anaphoric ambiguity (e.g. with the matrix subject).

part of speech	frequency (%)
<i>null</i>	4.9
common nouns	36.0
personal pronouns	26.0
proper nouns	15.7
demonstrative pronouns	10.4
adjectives	3.5
indefinite pronouns	1.6
verbal nouns	0.8
cardinal numerals	0.3
relative pronouns	0.2
quantifier	0.2
ordinal numerals	0.2
interrogative pronouns	0.1
possessive pronouns	0.1

Table 2.9: Pre-matrix dative absolutes in standard treebanks: subject parts of speech

part of speech	frequency (%)
<i>null</i>	2.8
common nouns	51.6
personal pronouns	19.7
proper nouns	11.7
demonstrative pronouns	5.2
indefinite pronouns	4.2
cardinal numerals	1.9
adjectives	1.9
verbal nouns	1

Table 2.10: Post-matrix dative absolutes in standard treebanks: subject parts of speech

Among the conjunct participles which were considered to be able to their own subject (i.e. the leftmost conjunct participle preceding the matrix clause), around half (50.7%)

were found to be null, which is not surprising considering the high percentage of null-subject conjunct participles that had been found in the previous dataset (44.8%). Subject parts of speech among overt-subject conjunct participles are even more varied than among dative absolutes (18 parts of speech in total), but most of the labour is done by common nouns, proper nouns, and demonstratives.

part of speech	frequency (%)
null	50.7
common nouns	19.8
proper nouns	11.2
demonstrative pronouns	7.0
personal pronouns	3.8
adjectives	3.7
indefinite pronouns	1.4
relative pronouns	0.8
verbal nouns	0.5
cardinal numerals	0.5
possessive pronouns	0.1
quantifier	<0.1
interrogative pronouns	<0.1
personal reflexive pronoun	<0.1
interrogative adverb	<0.1
possessive reflexive pronoun	<0.1
ordinal numerals	<0.1
relative adverb	<0.1

Table 2.11: Subject parts of speech among conjunct participles that may head an overt subject in standard treebanks (i.e. leftmost conjunct participles preceding the matrix clause)

When we consider the position of the subject relative to overt-subject conjunct participles, the results are overall in line with the observations made in Chapter 1. First, inherently anaphoric parts of speech (personal pronouns and demonstratives) are much more common among SV (Table 2.12) than VS conjunct participles (Table 2.13).

In Chapter 1, we explained this difference with the observation that the VS configuration in conjunct participles is much more likely than the SV one to be used when reinstating older or inactive referents in the discourse, a function which (following Haug 2012) we associated with FRAMES. This would explain why referentially more explicit parts of speech, such as proper nouns or common nouns, tend to be more frequent among VS conjunct participles, whereas inherently anaphoric parts of speech, such as demonstratives and personal pronouns, are much more common among SV conjunct participles. The latter are more likely to be used when there is a topic continuation from the immediately preceding discourse when anaphoric relations are thus easier to resolve. However, we also observed that among dative absolutes, which typically function as FRAMES like VS conjunct participles, the vast majority of subjects belonged to inherently anaphoric parts of speech (predominantly personal pronouns), which was tentatively

part of speech	frequency
common nouns	32.9
proper nouns	21.9
demonstrative pronouns	19.6
personal pronouns	10.2
adjectives	6.8
indefinite pronouns	3.5
relative pronouns	2.6
cardinal numerals	1
verbal nouns	0.7
possessive pronouns	0.38
interrogative pronouns	0.2
interrogative adverbs	0.1

Table 2.12: Pre-matrix SV conjunct participles in standard treebanks: subject parts of speech

part of speech	frequency
common nouns	54.1
proper nouns	24.6
adjectives	8.6
demonstrative pronouns	3.6
personal pronouns	3.2
indefinite pronouns	1.7
verbal nouns	1.7
cardinal numerals	1.3
quantifiers	0.5
personal reflexive pronouns	0.1
possessive reflexive pronouns	0.1
ordinal numerals	0.1
possessive pronouns	0.1

Table 2.13: Pre-matrix VS conjunct participles in standard treebanks: subject parts of speech

explained with the fact that dative absolutes may tend to attract only very highly prominent referents as their subject. In the Gospels, the VS configuration was found to be by far the most frequent one among dative absolutes, to the extent we were not able to look at differences with the SV configuration, since occurrences of the latter were far too few. In this dataset, however, we see that the predominance of personal pronouns among VS dative absolutes was not an idiosyncrasy of the previous dataset, since this is what we also find in our standard treebanks, as Table 2.15 shows. But we also see that among SV dative absolutes demonstratives are clearly much more frequent in the VS configuration, as Table 2.14 shows.

The constant between sentence-initial conjunct participles and dative absolutes is that demonstratives are clearly more frequent in the SV configuration, which we can

part of speech	frequency
common nouns	41.7
proper nouns	19.1
demonstrative pronouns	17.5
personal pronouns	14.8
adjectives	3.3
indefinite pronouns	1.7
verbal nouns	0.5
relative pronouns	0.3
ordinal numerals	0.3
cardinal numerals	0.2
interrogative pronouns	0.2
quantifiers	0.2
possessive pronouns	0.2

Table 2.14: Pre-matrix SV dative absolutes in standard treebanks: subject parts of speech

part of speech	frequency
personal pronouns	45.4
common nouns	32
proper nouns	12.6
adjectives	4.2
indefinite pronouns	1.7
demonstrative pronouns	1.5
verbal nouns	1.2
cardinal pronouns	0.5
quantifiers	0.2

Table 2.15: Pre-matrix VS dative absolutes in standard treebanks: subject parts of speech

easily explain with regular principles of anaphora resolution. What is peculiar, then, is the predominance of personal pronouns among VS absolutes. A possibility, already tentatively offered in the previous chapter, is that absolutes (VS absolutes, specifically, which were the vast majority there) attract very few, but very highly prominent subjects in the discourse. The lack of information-structural annotation on standard treebanks prevents us from looking at variables such as the average distance of the subject antecedents or referential activation in the previous discourse. However, we can look at the lexical variation among the subjects of sentence-initial participle constructions.

Similarly to what we did in the previous Chapter for participle lemmas, Table 2.16 shows two different metrics capturing lexical diversity among overt subjects: the percentage of subjects belonging to the 10 most frequent lemmas (10MFL) and the (moving-average) type-token ratio (MATTR) among subject lemmas (with a moving window of 40 tokens). These two metrics are calculated for the subjects of dative absolutes and conjunct participles over different subcorpora. ‘(O)CS’ refers to (Old) Church Slavonic,

(O)ES to Old East Slavic and Middle Russian texts together. When providing the metrics for the whole dataset, ‘non-N’ shows the numbers obtained from the data without normalizing the (O)CS and OES/MRus spellings, whereas ‘N’ indicates that the (O)CS and OES/MRus lemma variants corresponding to the same lemma were normalized and merged into one. This is needed because some differences in orthographic conventions between the (O)CS and the OES/MRus subcorpora in the TOROT treebank would result in listing as two separate lemmas some lexical items which should be regarded as one, as in *tъ* ‘that one’ according to the (O)CS convention and *tyi* according to the OES/MRus one, for example. Naturally, further normalization of the two subcorpora can only result in less overall variation than no normalization, but it is still useful to be aware of the extent to which our observations may be affected.⁸ The remaining rows show the lexical variation in the individual subcorpora, including different subject positions.

Subsample	Absolutes		Conjuncts	
	10MFL	MATTR	10MFL	MATTR
(O)CS + (O)ES + Mar (non-N)	46.34%	0.54	27.11%	0.58
(O)CS + (O)ES + Mar(N)	49.85%	0.53	29.09%	0.58
(O)CS + (O)ES (non-N)	46.34%	0.54	27.11%	0.58
(O)CS + (O)ES (N)	49.85%	0.53	29.09%	0.58
(O)CS	56.58%	0.54	29.16%	0.61
(O)ES	47.73%	0.53	36.82%	0.55
(O)CS + (O)ES + Mar (SV)	48.55%	0.57	35.69%	0.58
(O)CS + (O)ES + Mar (VS)	58.97%	0.48	22.98%	0.64
(O)CS + (O)ES (SV)	48.55%	0.57	35.69%	0.58
(O)CS + (O)ES (VS)	58.97%	0.48	22.98%	0.64
(O)CS (SV)	63.31%	0.55	36.32%	0.65
(O)CS (VS)	61.21%	0.51	30.23%	0.60
(O)ES (SV)	43.90%	0.58	41.75%	0.53
(O)ES (VS)	68.03%	0.42	31.69%	0.69

Table 2.16: Lexical variation among subjects in standard treebanks, by subcorpus and subject position. *10MFL* = 10 most frequent lemmas; *MATTR* = moving-average type-token ratio; *Mar* = Codex Marianus; *(O)ES* = Old East Slavic and Middle Russian; *N* = normalized; *non-N* = non-normalized. Note that the closer the *MATTR* is to 1 (and the lower the percentage for the *10MFL*), the greater the lexical variation.

As we can see by comparing the scores in Table 2.16 before and after normalization of mixed (O)CS and OES/MRus datasets, the further normalization only yielded a slight decrease in variation according to the *10MFL* value, whereas the *MATTR* remained the same regardless of normalization. The first clear observation is that lexical variation is overall greater among conjunct participles than dative absolutes, as Figures 2.1a and 2.1b

⁸The main changes made are: 1) *je* and word-initial *a* were changed to *e* and *ja* respectively, 2) Cyrillic characters for nasal vowels were changed to the respective east Slavic outcome (*e/ję > ja*, *o/jǫ > ju*), 3) *ě* was changed to *e*, 4) multiple variants for *y* were merged (and all set to <ы>), 5) word-final *ii* and *yi* in OES texts were changed to *ь* and *ѣ* respectively, 6) ТѣRT ТѣRT > TRѣT TRѣT. The normalization script can be found in the data depository at <https://doi.org/10.6084/m9.figshare.24166254>.

show. The difference in both average MATTR and percentage of subjects belonging to the ten most frequent lemmas is statistically significant with a p -value < 0.01 obtained from both a Welch's t -test and a Mann–Whitney U -test.

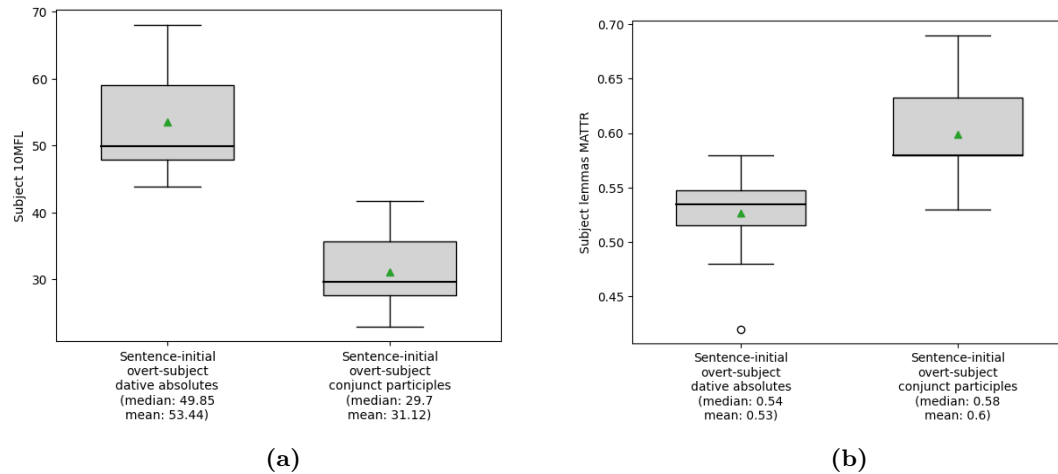


Figure 2.1: Lexical variation among subject lemmas in sentence-initial dative absolutes and conjunct participles. (a) Average percentage of subjects belonging to the 10 most frequent lemmas (10MFL). Welch's t -test: 8.94, p -value < 0.01 ; one-tailed Mann–Whitney U -test: 196, p -value < 0.01 . (b) Average moving-average type-token ratio (MATTR). Welch's t -test: -4.51, p -value < 0.01 ; one-tailed Mann–Whitney U -test: 15, p -value < 0.01 .

When we compare the lexical variation of subjects in sentence-initial SV and VS conjunct participles (Figure 2.3), we see that the VS configuration shows significantly more variation than the SV configuration, whereas the opposite is true of sentence-initial dative absolutes (Figure 2.2).

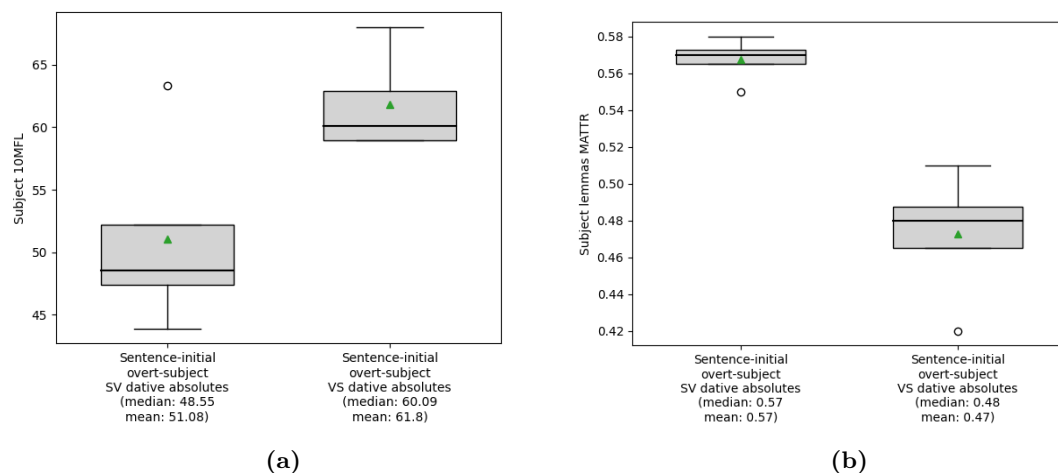


Figure 2.2: Lexical variation among subject lemmas in sentence-initial dative absolutes in the SV and VS configuration. (a) Average percentage of subjects belonging to the 10 most frequent lemmas (10MFL). Welch's t -test: -2.26, p -value = 0.04; one-tailed Mann–Whitney U -test: 3, p -value = 0.09. (b) Average moving-average type-token ratio (MATTR). Welch's t -test: 4.77, p -value < 0.01 ; one-tailed Mann–Whitney U -test: 16, p -value = 0.01.

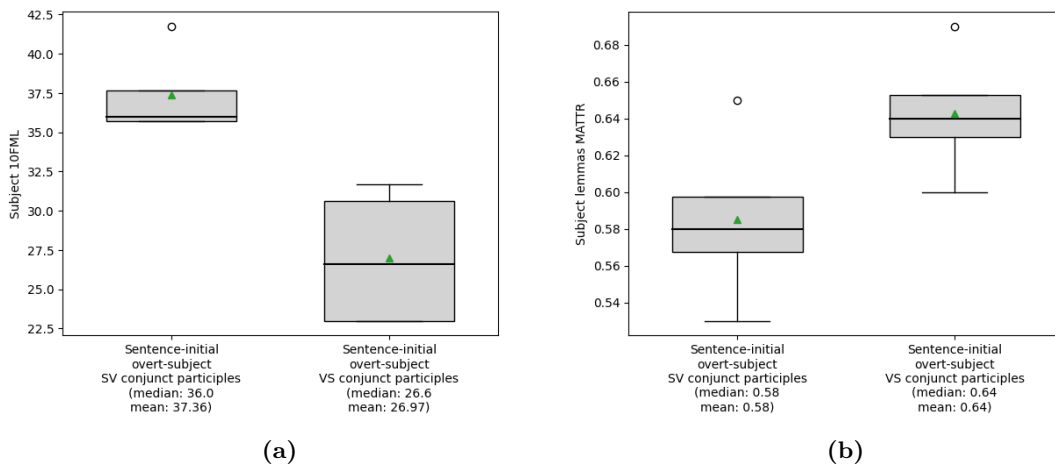


Figure 2.3: Lexical variation among subject lemmas in sentence-initial conjunct participles in the SV and VS configuration. (a) Average percentage of subjects belonging to the 10 most frequent lemmas (10FML). Welch’s t -test: 3.78, $p < 0.01$; one-tailed Mann–Whitney U -test: 16, $p = 0.01$. (b) Average moving-average type-token ratio (MATTR). Welch’s t -test: -1.86, $p = 0.05$; one-tailed Mann–Whitney U -test: 3, $p = 0.09$.

The much lower overall lexical variation among the subjects of sentence-initial dative absolutes and the much lower lexical variation among those in the VS configuration (contrary to conjunct participles) seems to support the explanation offered in Chapter 1 that dative absolutes, particularly those in the VS configuration, tend to attract few but very prominent referents as their subjects.

Given the heterogeneity of the corpus in terms of genres, and geographic and historical varieties, it may be useful to directly check whether there are clear differences between the most frequent subjects of sentence-initial dative absolutes and conjunct participles in the (O)CS and in the OES/MRus subcorpora among. Table 2.17 shows the most frequent subject lemmas by subcorpus and position in the sentence relative to the matrix clause.

Subcorpus	Absolutes	Conjuncts
(O)CS + (O)ES + Mar. (N)	* <i>i</i> ‘he’, <i>onǔ</i> ‘that one (there)’, <i>tǔ</i> ‘that one’, <i>sb</i> ‘this one’, <i>dbnǔ</i> ‘day’, <i>bogǔ</i> ‘God’, <i>leto</i> ‘year’, <i>svętǔ</i> ‘saint’, <i>azǔ</i> ‘I’, <i>Izǔjaslavǔ</i> ‘Izjaslav’	<i>onǔ</i> ‘that one (there)’, <i>azǔ</i> ‘I’, <i>cęsarǔ</i> ‘emperor’, <i>kǔnędzǔ</i> ‘king’, <i>blaženǔ</i> ‘(the) blessed (one)’, <i>Izǔjaslavǔ</i> ‘Izjaslav’, <i>my</i> ‘we’, <i>svętǔ</i> ‘saint’, <i>samǔ</i> ‘self’, Volodimerǔ ‘Volodymyr’

(O)CS + (O)ES (N)	*i ‘he’, on̄ ‘that one (there)’, t̄ ‘that one’, s̄ ‘this one’, d̄n̄ ‘day’, boḡ ‘God’, leto ‘year’, sv̄et̄ ‘saint’, az̄ ‘I’, Iz̄jaslav̄ ‘Izjaslav’	on̄ ‘that one (there)’, az̄ ‘I’, c̄sar̄ ‘emperor’, k̄n̄edz̄ ‘king’, blažen̄ ‘(the) blessed (one)’, Iz̄jaslav̄ ‘Izjaslav’, my ‘we’, sv̄et̄ ‘saint’, sam̄ ‘self’, Volodim̄r̄ ‘Volodymyr’
(O)ES	*i ‘he’, on̄ ‘that one (there)’, boḡ ‘God’, d̄n̄ ‘day’, Iz̄jaslav̄ ‘Izjaslav’, Feodosii ‘Feodosij’, az̄ ‘I’, k̄n̄edz̄ ‘king’	on̄ ‘that one (there)’, Iz̄jaslav̄ ‘Izjaslav’, az̄ ‘I’, k̄n̄edz̄ ‘king’, Volodim̄r̄ ‘Volodymyr’, blažen̄ ‘(the) blessed (one)’, c̄sar̄ ‘emperor’, my ‘we’, sam̄ ‘self’, polov̄č̄in̄ ‘Cuman’
(O)CS	*i ‘he’, t̄ ‘that one’, s̄ ‘this one’, sv̄et̄ ‘saint’, on̄ ‘that one (there)’, leto ‘year’, narod̄ ‘crowd’, d̄n̄ ‘day’, ty ‘you’, čas̄ ‘hour’	on̄ ‘that one (there)’, sv̄et̄ ‘saint’, c̄sar̄ ‘emperor’, vojevoda ‘commander’, k̄n̄edz̄ ‘king’, episkup̄ ‘bishop’, an̄tipat̄ ‘proconsul’, az̄ ‘I’, t̄ ‘that one’, blažen̄ ‘(the) blessed (one)’
(O)CS + (O)ES (SV)	*i ‘he’, on̄ ‘that one (there)’, t̄ ‘that one’, s̄ ‘this one’, boḡ ‘God’, d̄n̄ ‘day’, leto ‘year’, Iz̄jaslav̄ ‘Izjaslav’, Feodosii ‘Feodosij’, sl̄n̄ce ‘sun’	on̄ ‘that one (there)’, az̄ ‘I’, Iz̄jaslav̄ ‘Izjaslav’, c̄sar̄ ‘emperor’, my ‘we’, blažen̄ ‘(the) blessed (one)’, s̄ ‘this one’, sam̄ ‘self’, ty ‘you’, polov̄č̄in̄ ‘Cuman’
(O)CS + (O)ES (VS)	*i ‘he’, az̄ ‘I’, sv̄et̄ ‘saint’, d̄n̄ ‘day’, leto ‘year’, c̄sar̄ ‘emperor’, ot̄c̄ ‘father’, ty ‘you’, praz̄dn̄ik̄ ‘celebration, festivity’, narod̄ ‘crowd’	k̄n̄edz̄ ‘king’, c̄sar̄ ‘emperor’, blažen̄ ‘(the) blessed (one)’, sv̄et̄ ‘saint’, an̄tipat̄ ‘proconsul’, episkup̄ ‘bishop’, star̄ ‘old man’
(O)CS (SV)	*i ‘he’, on̄ ‘that one (there)’, t̄ ‘that one’, s̄ ‘this one’, leto ‘year’, d̄n̄ ‘day’, narod̄ ‘crowd’, čas̄ ‘hour’, boḡ ‘God’, sluga ‘servant’, vr̄m̄e ‘time’	on̄ ‘that one (there)’, az̄ ‘I’, c̄sar̄ ‘emperor’, s̄ ‘this one’, sv̄et̄ ‘saint’, ty ‘you’, jedin̄ ‘one, a certain one’, t̄ ‘that one’, vojevoda ‘commander’, vs̄ ‘all’
(O)CS (VS)	*i ‘he’, sv̄et̄ ‘saint’, ot̄c̄ ‘father’, ty ‘you’, narod̄ ‘crowd’, az̄ ‘I’, Pionii ‘Pionos’, Iona ‘Jonah’, vs̄ ‘all’, vr̄m̄e ‘time’, an̄tipat̄ ‘proconsul’	k̄n̄edz̄ ‘king’, sv̄et̄ ‘saint’, an̄tipat̄ ‘proconsul’, episkup̄ ‘bishop’, vojevoda ‘commander’, c̄sar̄ ‘emperor’, sluga ‘servant’, farisei ‘Pharisee’, mož̄ ‘man’, star̄ ‘old man’

and inanimate objects (e.g. *dvr̥v̥* ‘door’ or *sl̥n̥b̥ce* ‘Sun’) are quite common, which once again points to the typical framing nature of absolute constructions, since these nouns are generally used to describe the background to a main eventuality (as in ‘when the day ended’ or ‘after shutting the doors’).

Summary

In this section, the properties of subjects were analysed, focussing on dative absolutes and conjunct participles in sentence-initial position, since that is where we observe the clearest functional overlap (since both can function as FRAMES) and since that is where conjunct participles were argued to be able to syntactically head a subject. We saw clear differences between the two constructions. Among sentence-initial dative absolutes, overt subjects are overwhelmingly more frequent, with only 5.6% (39) of absolutes having a null subject. Among sentence-initial conjunct participles, overt and null subjects are almost equally frequent (50.8% overt versus 49.2% null). Among both overt-subject conjunct participles and dative absolutes, there are some differences between the SV and VS configurations. The subjects of conjunct participles display very similar properties to those observed in the Codex Marianus. Namely, besides common nouns, which are frequent in both configurations (albeit more so among the VS one), among the subjects of SV conjunct participles, anaphoric parts of speech (demonstratives and personal pronouns) are much more common than among VS conjunct participles. This was explained by the fact that VS conjunct participles are much more likely to be used when reintroducing older or inactive referents in the discourse, which may call for referentially more explicit parts of speech to be used. Among sentence-initial dative absolutes, demonstratives, as with conjunct participles, are much more frequent in the SV configuration, whereas personal pronouns are overwhelmingly more present in the VS one. Tentatively, this was explained by the fact that VS absolutes attract very few but highly discourse-prominent subjects. The analysis of lexical variation among subjects showed that the subjects of sentence-initial dative absolutes display overall significantly smaller lexical variation than conjunct participles but also that subjects in VS dative absolutes are much less lexically varied than those in the SV configuration.

2.2.3 Lexical variation among participles

The pattern observed regarding lexical variation among the subjects of conjunct participles and dative absolutes can, to some extent, also be observed in the lexical variation among the participles themselves. Table 2.18 shows the number of participles belonging to the ten most frequent lemmas (10MFL) and the MATTR for all participle lemmas at a moving window of 40 tokens, divided into subcorpora and position in the sentence relative to the matrix clause.

Subsample	Absolutes		Conjuncts	
	10MFL	MATTR	10MFL	MATTR
(O)CS + (O)ES + Mar (non-N)	36.15%	0.68	23.20%	0.78
(O)CS + (O)ES + Mar (N)	38.53%	0.68	26.78%	0.78
(O)CS + (O)ES (non-N)	36.05%	0.68	23.07%	0.79
(O)CS + (O)ES (N)	38.46%	0.68	26.38%	0.79
(O)CS + Mar	42.74%	0.66	27.41%	0.78
(O)CS	43.31%	0.66	26.92%	0.80
(O)ES	37.93%	0.69	26.38%	0.79
Mar	50.54%	0.64	36.79%	0.72
(O)CS + (O)ES + Mar (SV)	40.97%	0.70	33.11%	0.76
(O)CS + (O)ES + Mar (VS)	37.99%	0.67	53.55%	0.55
(O)CS + (O)ES (SV)	39.24%	0.70	32.41%	0.77
(O)CS + (O)ES (VS)	39.74%	0.68	54.51%	0.57
(O)CS (SV)	54.68%	0.62	36.04%	0.75
(O)CS (VS)	38.79%	0.74	50.17%	0.61
(O)ES (SV)	35.12%	0.76	31.86%	0.77
(O)ES (VS)	52.38%	0.61	65.27%	0.51
Mar (SV)	NA	NA	47.34%	0.70
Mar (VS)	45.83%	0.64	64.71%	0.50
(O)CS + (O)ES + Mar (left, N)	39.83%	0.68	28.42%	0.77
(O)CS + (O)ES + Mar (right, N)	40.71%	0.71	32.00%	0.74
(O)CS + (O)ES (left, N)	39.36%	0.68	26.88%	0.79
(O)CS + (O)ES (right, N)	41.31%	0.71	32.95%	0.74
(O)CS (left)	44.22%	0.67	27.36%	0.79
(O)CS (right)	45.78%	0.72	33.76%	0.74
(O)ES (left)	39.47%	0.70	27.06%	0.78
(O)ES (right)	46.92%	0.68	32.84%	0.73
Mar (left)	51.46%	0.64	44.12%	0.67
Mar (right)	NA	NA	30.11%	0.81

Table 2.18: Lexical variation among participles by subcorpus. *10MFL* = ten most frequent lemmas; *MATTR* = moving-average type-token ratio; *Mar* = Codex Marianus; *left* and *right* = pre-matrix and post matrix, respectively; *N* = normalized; *non-N* = non-normalized. Note that the closer the MATTR is to 1 (and the lower the percentage for the 10MFL), the greater the lexical variation. Bold = configurations where conjunct participles show a lower degree of lexical variation than dative absolutes.

Lexical variation among conjunct participles is almost always consistently greater than among dative absolutes across subcorpora and positions in the sentence. This is not surprising, considering what we have observed regarding the typical framing character of dative absolutes, as opposed to the varied functions of conjunct participles. As argued in Chapter 1 (following Haug (2012: 320)), FRAMES are assumed to be more ‘predictable’, since they are presupposed (anaphoric or accommodated), whereas ELABORATIONS and INDEPENDENT RHEMES typically introduce new information, so we do expect to find less lexical variation among FRAMES than participle constructions with other functions. As

Figures 2.4a and 2.4b show, both mean and median 10MFL/MATTR for conjunct participles are well beyond the midspread of the values for dative absolutes (they are in fact well beyond the whiskers of the plot for the latter, meaning that their means/medians are beyond the maximum or minimum observed values for dative absolutes). The difference is significant according to all tests (p -value < 0.01).

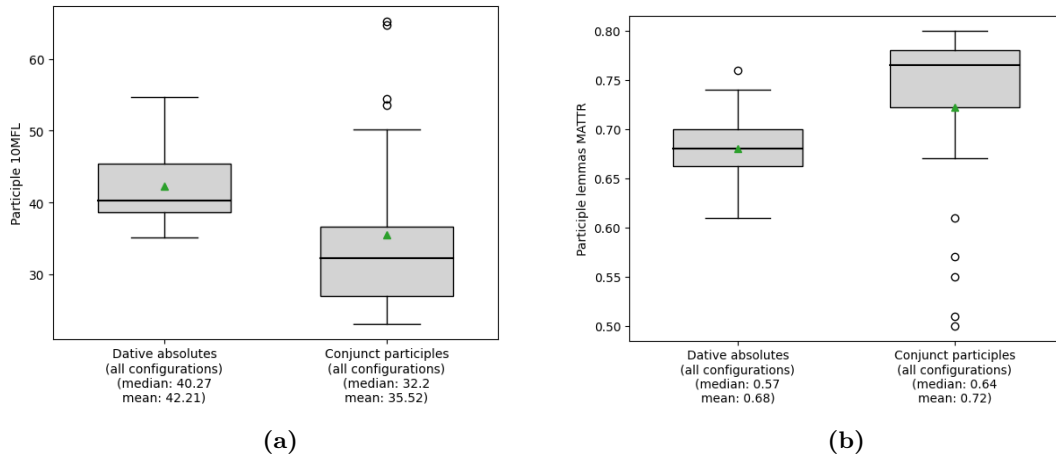


Figure 2.4: Lexical variation among participles, any configuration. (a) Average percentage of participles belonging to the ten most frequent lemmas (10MFL). Welch’s t -test: 2.56, p -value = 0.01; one-tailed Mann–Whitney U -test: 530, p -value < 0.01 . (b) Average moving-average type-token ratio (MATTR) among lemmas. Welch’s t -test: -2.16, p -value = 0.03; one-tailed Mann–Whitney U -test: 158, p -value < 0.01 .

As in the Gospels, the level of lexical variation among conjunct participles is very different depending on the position of the subject relative to the participle, with the conjunct participles in the VS configuration showing much less lexical variation according to both tests, which could, again, be associated with the clearer FRAME function of sentence-initial conjunct participles. The difference is highly significant according to both tests (Figure 2.5).

A different situation is observed instead among sentence-initial dative absolutes: there is no significant difference in lexical variation between dative absolutes in the SV and in the VS configuration, as Figure 2.6 shows.

Overall, this result is also not surprising in itself. As we have already observed when looking at the properties of subjects above and in the previous Chapter (Section 1.3), the principles driving the choice of VS and SV configuration in sentence-initial dative absolutes and conjunct participles are clearly not identical. In light of all the results obtained so far, it is clear that sentence-initial dative absolutes (and, in fact, *most* participles in any configuration) have relatively consistent functions throughout. In Table 2.18, rows in bold indicate when variation is smaller among conjunct participles than dative absolutes: in all cases, these are VS configurations. Given the consistently small lexical variation among dative absolutes and the clear correlation between subject position and lexical variation among conjunct participles, it seems evident that, even

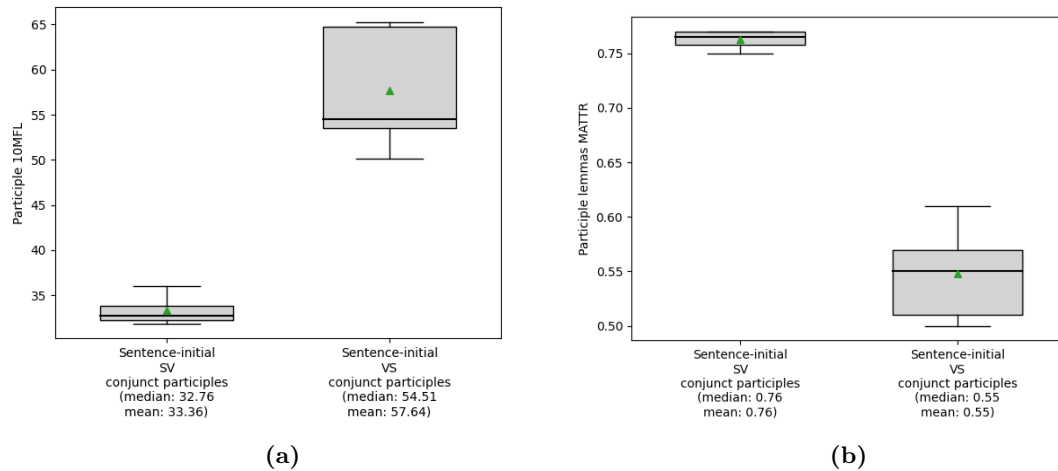


Figure 2.5: Lexical variation among participles in sentence-initial participles, SV versus VS configuration. (a) Average percentage of participles belonging to the 10 most frequent lemmas (10MFL). Welch's t -test: -7.53 , p -value < 0.01 ; one-tailed Mann-Whitney U -test: 0 , p -value < 0.01 . (b) Average moving-average type-token ratio (MATTR) among lemmas. Welch's t -test: 10.38 , p -value < 0.01 ; one-tailed Mann-Whitney U -test: 20 , p -value < 0.01 .

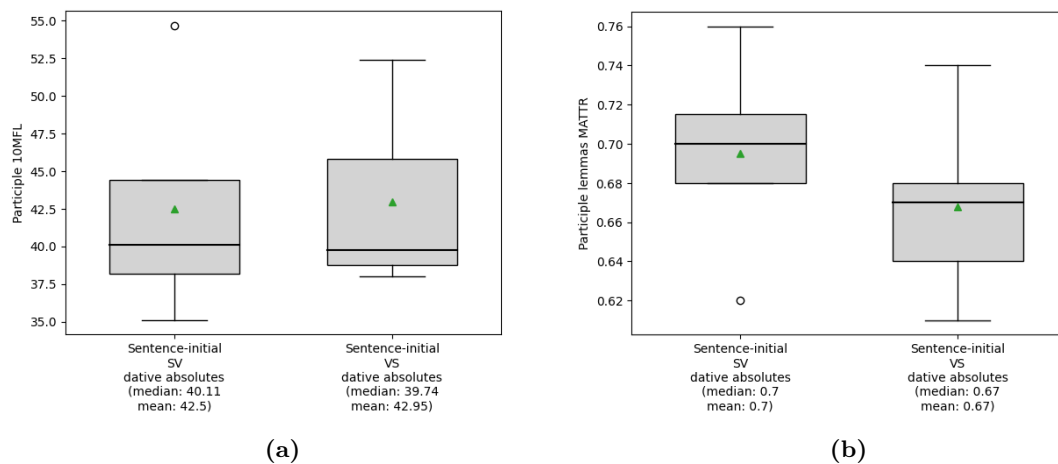


Figure 2.6: Lexical variation among participles in sentence-initial dative absolutes (SV versus VS). (a) Average percentage of participles belonging to the 10 most frequent lemmas (10MFL). Welch's t -test: -0.08 , p -value = 0.93 ; one-tailed Mann-Whitney U -test: 10 , p -value 1 . (b) Average moving-average type-token ratio (MATTR) among lemmas. Welch's t -test: 0.74 , p -value = 0.48 ; one-tailed Mann-Whitney U -test: 14 , p -value = 0.19 .

beyond the Gospels, VS conjunct participles are very likely to function as FRAMES and SV ones more as either INDEPENDENT RHEMES or ELABORATIONS, depending on the position relative to the matrix clause, whereas dative absolutes tend to have a consistently small lexical variation throughout, reflecting their typical framing function.

Table 2.19 show the most frequent lemmas among dative absolutes and conjunct participles, divided by subcorpus.

Subcorpus	Absolutes	Conjuncts
(O)CS + (O)ES + Mar. (N)	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>iti</i> ‘go’, <i>chotěti</i> ‘want’, <i>rešti</i> ‘say’, <i>sěděti</i> ‘sit’, <i>moliti se</i> ‘pray’, <i>minęti</i> ‘pass’, <i>vęniti</i> ‘go in’	<i>glagolati</i> ‘speak, say’, <i>viděti</i> ‘see’, <i>byti</i> ‘be’, <i>slyšati</i> ‘hear’, <i>priti</i> ‘come, arrive’, <i>rešti</i> ‘say’, <i>iti</i> ‘go’, <i>chotěti</i> ‘want’, <i>prijęti</i> ‘take, receive’, <i>vęzęti</i> ‘take’
(O)CS + (O)ES (N)	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>chotěti</i> ‘want’, <i>rešti</i> ‘say’, <i>iti</i> ‘go’, <i>sěděti</i> ‘sit’, <i>moliti se</i> ‘pray’, <i>minęti</i> ‘pass’, <i>stojati</i> ‘stand’	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>rešti</i> ‘say’, <i>priti</i> ‘come, arrive’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>vęzęti</i> ‘take’, <i>prijęti</i> ‘take, receive’
(O)CS + Mar.	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>priti</i> ‘come, arrive’, <i>rešti</i> ‘say’, <i>vęniti</i> ‘go in’, <i>iti</i> ‘go’, <i>minęti</i> ‘pass’, <i>chotěti</i> ‘want’, <i>moliti se</i> ‘pray’, <i>doiti</i> ‘arrive’	<i>glagolati</i> ‘speak, say’, <i>viděti</i> ‘see’, <i>byti</i> ‘be’, <i>slyšati</i> ‘hear’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>prijęti</i> ‘take, receive’, <i>chotěti</i> ‘want’, <i>iměti</i> ‘have’, <i>sętvoriti</i> ‘make’
(O)CS	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>chotěti</i> ‘want’, <i>doiti</i> ‘arrive’, <i>minęti</i> ‘pass’, <i>moliti se</i> ‘pray’, <i>iti</i> ‘go’, <i>konęčati</i> ‘finish’	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>priti</i> ‘come, arrive’, <i>prijęti</i> ‘take, receive’, <i>vęzęti</i> ‘take’, <i>sętvoriti</i> ‘make’
(O)ES	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>sěděti</i> ‘sit’, <i>prispěti</i> ‘rush, happen’, <i>stojati</i> ‘stand’, <i>žiti</i> ‘live’, <i>rešti</i> ‘say’, <i>nastati</i> ‘begin’	<i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>slyšati</i> ‘hear’, <i>viděti</i> ‘see’, <i>priti</i> ‘come, arrive’, <i>byti</i> ‘be’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>vęzęti</i> ‘take’, <i>iměti</i> ‘have’
Mar.	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>vęniti</i> ‘go in’, <i>iti</i> ‘go’, <i>isęchoditi</i> ‘exit, go out’, <i>vęzležati</i> ‘lay down’, <i>jasti</i> ‘eat’, <i>priti</i> ‘come, arrive’, <i>iziti</i> ‘go down’, <i>iměti</i> ‘have’	<i>viděti</i> ‘see’, <i>priti</i> ‘come, arrive’, <i>slyšati</i> ‘hear’, <i>iti</i> ‘go’, <i>pristępiti</i> ‘ascend, approach’, <i>iziti</i> ‘go down’, <i>prijęti</i> ‘take, receive’, <i>vęstati</i> ‘stand up’, <i>byti</i> ‘be’, <i>vęzvręti</i> ‘look up at’
(O)CS + (O)ES + Mar. (SV)	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>priti</i> ‘come, arrive’, <i>rešti</i> ‘say’, <i>chotěti</i> ‘want’, <i>moliti se</i> ‘pray’, <i>minęti</i> ‘pass’, <i>iti</i> ‘go’, <i>pomagati</i> ‘help’, <i>sěděti</i> ‘sit’	<i>glagolati</i> ‘speak, say’, <i>viděti</i> ‘see’, <i>byti</i> ‘be’, <i>slyšati</i> ‘hear’, <i>rešti</i> ‘say’, <i>chotěti</i> ‘want’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>iměti</i> ‘have’, <i>prijęti</i> ‘take, receive’
(O)CS + (O)ES + Mar. (VS)	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>sěděti</i> ‘sit’, <i>glagolati</i> ‘speak, say’, <i>chotěti</i> ‘want’, <i>vęniti</i> ‘go in’, <i>cęsaręstvovati</i> ‘rule, serve as emperor’, <i>doiti</i> ‘arrive’, <i>umręti</i> ‘die’	<i>slyšati</i> ‘hear’, <i>viděti</i> ‘see’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>pristępiti</i> ‘ascend, approach’, <i>vęstati</i> ‘stand up’, <i>byti</i> ‘be’, <i>prijęti</i> ‘take, receive’, <i>iziti</i> ‘go down’, <i>otęvręstati</i> ‘reply’

(O)CS + (O)ES (SV)	<i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>moliti se</i> ‘pray’, <i>minŕti</i> ‘pass’, <i>pomagati</i> ‘help’, <i>sěděti</i> ‘sit’	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>rešti</i> ‘say’, <i>slyšati</i> ‘hear’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>přiti</i> ‘come, arrive’, <i>iměti</i> ‘have’, <i>vzěti</i> ‘take’
(O)CS + (O)ES (VS)	<i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>sěděti</i> ‘sit’, <i>chotěti</i> ‘want’, <i>glagolati</i> ‘speak, say’, <i>česarŕstvovati</i> ‘rule, serve as emperor’, <i>doiti</i> ‘arrive’, <i>umrěti</i> ‘die’, <i>moliti se</i> ‘pray’	<i>slyšati</i> ‘hear’, <i>viděti</i> ‘see’, <i>přiti</i> ‘come, arrive’, <i>otověštati</i> ‘reply’, <i>byti</i> ‘be’, <i>razgněvati</i> ‘anger’, <i>iti</i> ‘go’, <i>uvěděti</i> ‘discover’, <i>vzěti</i> ‘take’, <i>přijeti</i> ‘take, receive’
(O)CS (SV)	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>přiti</i> ‘come, arrive’, <i>chotěti</i> ‘want’, <i>minŕti</i> ‘pass’, <i>umrěti</i> ‘die’, <i>moliti se</i> ‘pray’, <i>stati</i> ‘become’, <i>koněčati</i> ‘finish’	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>chotěti</i> ‘want’, <i>iměti</i> ‘have’, <i>přijeti</i> ‘take, receive’, <i>vzěti</i> ‘take’, <i>přiti</i> ‘come, arrive’, <i>iti</i> ‘go’
(O)CS (VS)	<i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>doiti</i> ‘arrive’, <i>česarŕstvovati</i> ‘rule, serve as emperor’, <i>rešti</i> ‘say’, <i>iti</i> ‘go’, <i>moliti se</i> ‘pray’, <i>vŕniti</i> ‘go in’, <i>ležati</i> ‘lie, recline’	<i>slyšati</i> ‘hear’, <i>viděti</i> ‘see’, <i>razgněvati</i> ‘anger’, <i>otověštati</i> ‘reply’, <i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>vzěti</i> ‘take’, <i>izlěsti</i> ‘crawl, climb out’, <i>uvěděti</i> ‘discover’
(O)ES (SV)	<i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>chotěti</i> ‘want’, <i>rešti</i> ‘say’, <i>pomagati</i> ‘help’, <i>sěděti</i> ‘sit’, <i>iti</i> ‘go’, <i>prispěti</i> ‘rush, happen’, <i>stojati</i> ‘stand’, <i>žiti</i> ‘live’	<i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>chotěti</i> ‘want’, <i>byti</i> ‘be’, <i>iti</i> ‘go’, <i>přiti</i> ‘come, arrive’, <i>prositi</i> ‘beg, ask’, <i>vzěti</i> ‘take’
(O)ES (VS)	<i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>sěděti</i> ‘sit’, <i>chotěti</i> ‘want’, <i>nastati</i> ‘begin’, <i>prispěti</i> ‘rush, happen’, <i>běžati</i> ‘run’, <i>umrěti</i> ‘die’, <i>minŕti</i> ‘pass’	<i>slyšati</i> ‘hear’, <i>viděti</i> ‘see’, <i>přiti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>uvěděti</i> ‘discover’, <i>uslyšati</i> ‘hear’, <i>otověštati</i> ‘reply’, <i>vzěti</i> ‘take’, <i>sŕdumati</i> ‘think’, <i>přičhati</i> ‘come’
Mar. (SV)	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>zatvoriti</i> ‘close’, <i>isŕchoditi</i> ‘exit, go out’, <i>četrŕtovlastŕstvovati</i> ‘serve as a tetrarch’, <i>vŕsijati</i> ‘begin to shine’, <i>čuditi</i> ‘marvel’, <i>plěsati</i> ‘dance’, <i>vzležati</i> ‘lay down’, <i>ugoditi</i> ‘please, satisfy’	<i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>byti</i> ‘be’, <i>přiti</i> ‘come, arrive’, <i>učiti</i> ‘learn, teach’, <i>iměti</i> ‘have’, <i>iti</i> ‘go’, <i>věděti</i> ‘know’, <i>přijeti</i> ‘take, receive’, <i>iskušati</i> ‘tempt, try’

Mar. (VS)	<i>byti</i> ‘be’, <i>v̄niti</i> ‘go in’, <i>glagolati</i> ‘speak, say’, <i>iti</i> ‘go’, <i>iziti</i> ‘go down’, <i>priti</i> ‘come, arrive’, <i>v̄zlezati</i> ‘lay down’, <i>is̄choditi</i> ‘exit, go out’, <i>jasti</i> ‘eat’, <i>s̄choditi</i> ‘go down’	<i>viděti</i> ‘see’, <i>iti</i> ‘go’, <i>priti</i> ‘come, arrive’, <i>pristopiti</i> ‘ascend, approach’, <i>slyšati</i> ‘hear’, <i>iziti</i> ‘go down’, <i>v̄stati</i> ‘stand up’, <i>prijęti</i> ‘take, receive’, <i>v̄niti</i> ‘go in’, <i>v̄zvręti</i> ‘look up at’
(O)CS + (O)ES + Mar. (left-N)	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>iti</i> ‘go’, <i>chotęti</i> ‘want’, <i>rešti</i> ‘say’, <i>sędęti</i> ‘sit’, <i>minęti</i> ‘pass’, <i>v̄niti</i> ‘go in’, <i>moliti sę</i> ‘pray’	<i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>v̄zęti</i> ‘take’, <i>prijęti</i> ‘take, receive’, <i>v̄stati</i> ‘stand up’, <i>s̄tvoriti</i> ‘make’, <i>chotęti</i> ‘want’
(O)CS + (O)ES + Mar. (right-N)	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>pomagati</i> ‘help’, <i>sędęti</i> ‘sit’, <i>cęsaręstvovati</i> ‘rule, serve as emperor’, <i>moliti sę</i> ‘pray’, <i>chotęti</i> ‘want’, <i>stojati</i> ‘stand’, <i>s̄pati</i> ‘sleep’, <i>bežati</i> ‘run’	<i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>chotęti</i> ‘want’, <i>byti</i> ‘be’, <i>imęti</i> ‘have’, <i>prositi</i> ‘beg, ask’, <i>moliti sę</i> ‘pray’, <i>tvoriti</i> ‘make’, <i>viděti</i> ‘see’, <i>slaviti</i> ‘praise’
(O)CS + (O)ES (left-N)	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>chotęti</i> ‘want’, <i>glagolati</i> ‘speak, say’, <i>iti</i> ‘go’, <i>rešti</i> ‘say’, <i>sędęti</i> ‘sit’, <i>minęti</i> ‘pass’, <i>moliti sę</i> ‘pray’, <i>v̄niti</i> ‘go in’	<i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>v̄zęti</i> ‘take’, <i>prijęti</i> ‘take, receive’, <i>s̄tvoriti</i> ‘make’, <i>chotęti</i> ‘want’, <i>v̄stati</i> ‘stand up’
(O)CS + (O)ES (right-N)	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>pomagati</i> ‘help’, <i>sędęti</i> ‘sit’, <i>cęsaręstvovati</i> ‘rule, serve as emperor’, <i>moliti sę</i> ‘pray’, <i>bežati</i> ‘run’, <i>rešti</i> ‘say’, <i>stojati</i> ‘stand’, <i>v̄schoditi</i> ‘go up’	<i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>chotęti</i> ‘want’, <i>byti</i> ‘be’, <i>imęti</i> ‘have’, <i>tvoriti</i> ‘make’, <i>prositi</i> ‘beg, ask’, <i>moliti sę</i> ‘pray’, <i>viděti</i> ‘see’, <i>slaviti</i> ‘praise’
(O)CS (left)	<i>byti</i> ‘be’, <i>priti</i> ‘come, arrive’, <i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>chotęti</i> ‘want’, <i>iti</i> ‘go’, <i>doiti</i> ‘arrive’, <i>minęti</i> ‘pass’, <i>v̄niti</i> ‘go in’, <i>kon̄čati</i> ‘finish’	<i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>priti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>v̄zęti</i> ‘take’, <i>prijęti</i> ‘take, receive’, <i>s̄tvoriti</i> ‘make’, <i>chotęti</i> ‘want’, <i>otvęštati</i> ‘reply’
(O)CS (right)	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>cęsaręstvovati</i> ‘rule, serve as emperor’, <i>podajati</i> ‘give’, <i>umręti</i> ‘die’, <i>moliti sę</i> ‘pray’, <i>rešti</i> ‘say’, <i>zatvoriti</i> ‘close’, <i>s̄kon̄čavati</i> ‘finish’, <i>s̄v̄prašati</i> ‘ask’	<i>glagolati</i> ‘speak, say’, <i>byti</i> ‘be’, <i>chotęti</i> ‘want’, <i>imęti</i> ‘have’, <i>viděti</i> ‘see’, <i>rešti</i> ‘say’, <i>tvoriti</i> ‘make’, <i>moliti sę</i> ‘pray’, <i>prijęti</i> ‘take, receive’, <i>slaviti</i> ‘praise’

(O)ES (left)	<i>byti</i> ‘be’, <i>priiti</i> ‘come, arrive’, <i>chotěti</i> ‘want’, <i>iti</i> ‘go’, <i>sěděti</i> ‘sit’, <i>nastati</i> ‘begin’, <i>prispěti</i> ‘rush, happen’, <i>žiti</i> ‘live’, <i>stojati</i> ‘stand’, <i>minōti</i> ‘pass’	<i>slyšati</i> ‘hear’, <i>viděti</i> ‘see’, <i>priiti</i> ‘come, arrive’, <i>iti</i> ‘go’, <i>byti</i> ‘be’, <i>vzěti</i> ‘take’, <i>vzstati</i> ‘stand up’, <i>pojěti</i> ‘take’, <i>sъtvoriti</i> ‘make’, <i>chotěti</i> ‘want’
(O)ES (right)	<i>byti</i> ‘be’, <i>pomagati</i> ‘help’, <i>sěděti</i> ‘sit’, <i>běžati</i> ‘run’, <i>vъschoditi</i> ‘go up’, <i>chotěti</i> ‘want’, <i>stojati</i> ‘stand’, <i>izvoliti</i> ‘want, prefer’, <i>moliti se</i> ‘pray’, <i>svitati</i> ‘dawn’	<i>glagolati</i> ‘speak, say’, <i>rešti</i> ‘say’, <i>chotěti</i> ‘want’, <i>prositi</i> ‘beg, ask’, <i>iměti</i> ‘have’, <i>tvoriti</i> ‘make’, <i>moliti se</i> ‘pray’, <i>věděti</i> ‘know’, <i>velěti</i> ‘command’, <i>byti</i> ‘be’
Mar. (left)	<i>byti</i> ‘be’, <i>glagolati</i> ‘speak, say’, <i>vъniti</i> ‘go in’, <i>iti</i> ‘go’, <i>isъchoditi</i> ‘exit, go out’, <i>vъzležati</i> ‘lay down’, <i>jasti</i> ‘eat’, <i>iziti</i> ‘go down’, <i>priiti</i> ‘come, arrive’, <i>iměti</i> ‘have’	<i>viděti</i> ‘see’, <i>priiti</i> ‘come, arrive’, <i>slyšati</i> ‘hear’, <i>iti</i> ‘go’, <i>pristopiti</i> ‘ascend, approach’, <i>iziti</i> ‘go down’, <i>prijěti</i> ‘take, receive’, <i>vzstati</i> ‘stand up’, <i>vъzvrěti</i> ‘look up at’, <i>byti</i> ‘be’
Mar. (right)	<i>byti</i> ‘be’, <i>pristopiti</i> ‘ascend, approach’, <i>vъsijati</i> ‘begin to shine’, <i>sъpati</i> ‘sleep’, <i>vlasti</i> ‘rule’, <i>vъniti</i> ‘go in’, <i>zatvoriti</i> ‘close’, <i>vъzdati</i> ‘give (in homage)’, <i>pospěšъstvovati</i> ‘assist, promote’, <i>utvrъždati</i> ‘establish, build’	<i>učiti</i> ‘learn, teach’, <i>byti</i> ‘be’, <i>iměti</i> ‘have’, <i>iskušati</i> ‘tempt, try’, <i>viděti</i> ‘see’, <i>iskati</i> ‘seek’, <i>rešti</i> ‘say’, <i>věděti</i> ‘know’, <i>moliti se</i> ‘pray’, <i>propovědati</i> ‘preach’

Table 2.19: Dative absolutes and conjunct participles: ten most-frequent lemmas by subcorpus and position of the participle relative to the matrix clause. Lemmas for the mixed (O)CS/(O)ES subcorpus are given on the basis of the normalized data. Only lemmas occurring at least twice are included. If fewer than ten lemmas occur at least twice, then only those are included. *Mar.* = Codex Marianus; *(O)ES* = Old East Slavic and Middle Russian; *left* and *right* = pre-matrix and post matrix, respectively; *N* = normalized.

The first observation to make is that a number of verbs, such as *byti* and motion verbs like *priiti* ‘come, arrive’ are frequent among both conjunct participles and dative absolutes across all configurations. At the same time, however, there are lemmas that are very frequent among the occurrences of one of the two constructions but not present at all, or much less frequent, among the occurrences of the other. Among dative absolutes, we find several verbs denoting activities and states, such as *sěděti* ‘to be sitting’, *sъpati* ‘sleep’, *žiti* ‘live’, *moliti se* ‘pray’, *stojati* ‘be standing’, which are most often imperfective and, as is most typical, function as FRAMES, as in (20)-(21).

- (20) *i se vъ jedinъ dnъ sъpęštju*
 and behold on one.SG.M.ACC day.SG.M.ACC sleep.PTCP.IPFV.M.DAT.SG
tomu vъ poludnie i se javi se
 that.M.DAT.SG at midday.SG.N.ACC and behold appear.AOR.3.SG REFL

jemu *ocb* *našb* *feodosii*
 3.SG.M.DAT father.M.NOM.SG our.1.M.NOM.SG Feodosij.M.NOM.SG
glę
 say.PTCP.IPFV.M.NOM.SG

‘And one day, as he was sleeping at midday, our Father Feodosij appeared to him, saying’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 65b)

- (21) *i* *sice* ***moleštīm*** *sę* *poidoša* *polovci*
 and thus pray.PTCP.IPFV.DAT.PL REFL advance.AOR.3.PL Polovtsy.NOM.PL
 ‘And as they [the Rusian princes] were praying, the Polovtsy advanced’ (*Primary Chronicle*, Codex Laurentianus f. 93d)

Also among dative absolutes, we find verbs, such as *minōti* ‘pass’, clearly involved in time expressions, as in (22).

- (22) *časoma* *že* *dvěma* ***minovšěma.***
 hour.M.DAT.DU PTC two.M.DAT.DU pass.PTCP.PFV.M.DAT.DU
raschoždaaše *sę* *po malu* *tōma*
 disperse.IMPF.3.SG REFL slowly darkness.F.NOM.SG
 ‘After two hours had passed, darkness slowly started to disperse’ (*Vita of Kodratos*, Codex Suprasliensis f. 56v)

Very frequent among absolutes are also verbs indicating under which ruler or religious figure a particular event occurs within the narrative. Lemmas such as *česarbstvovati* ‘rule, serve as emperor’ and *četvrtovlastbstvovati* ‘serve as a tetrarch’ lend themselves well as framing devices in a discourse, setting the main event within a timeframe. (23) is one of several such examples.

- (23) *vō lēt* *.dz.* *ps.* *vi.*
 in year.SG.N.ACC six thousand seven hundred twelve
crstvujuštju. *olbkse.* *vō csrigradě.* *vō*
 reign.PTCP.IPFV.M.DAT.SG Oleksa.M.DAT.SG in Constantinople.SG.M.LOC in
crstvě *isakově.* *brat* *svojego.*
 reign.SG.N.LOC Isak.SG.N.LOC.STRONG brother.SG.M.GEN his.GEN.SG
jegože *slěpivō.* *a samō*
 who.SG.M.GEN blind.PTCP.PFV.M.NOM.SG and self.M.NOM.SG
csrmō *sta.* *a sýna* *jego*
 emperor.M.INST.SG become.AOR.3.SG and son.SG.M.GEN his.GEN.SG
oleksu. *zatvori* *vō stěnachō* *vyšokychō.*
 Oleksa.SG.M.ACC shut.AOR.3.SG in wall.PL.F.LOC high.PL.F.LOC
stražěju. *jako ne vynidetō.*
 guard.F.INST.SG that NEG go out.PRS.3.SG

‘In the year 6712 [1204], when Oleksa was reigning in Constantinople: during the reign of his brother Isak, whom Oleksa blinded before becoming emperor himself, Isak shut his son Oleksa in high walls under guard so that he could not get out’ (*First Novgorod Chronicle* f. 64r)

On the other hand, among conjunct participles we find several telic verbs, such as *prijeti* ‘take, receive’, *vzēti* ‘take’, *slyšati* ‘hear’ *sotvoriti* ‘make’, *viděti* ‘see’, *vzstati* ‘stand up’. We can instead more easily envision these as part of a series of foreground clauses following one another, as is often the case with INDEPENDENT RHEMES. (24)-(26) are typical examples of conjunct participles with these lemmas.

- (24) *Kagan že vьzъmъ čašu reče*
 khan PTC take.PTCP.PFV.M.NOM.SG cup.F.ACC.SG say.AOR.3.SG
 ‘The Khan then took the cup and said’ (*Vita Constantini* 9)

- (25) *plaměň velikъ zělo ot vьrbcha*
 flame.M.NOM.SG big.M.NOM.SG.STRONG very from top.SG.M.GEN
crkŭvnaago išьdъ i akŭ komara
 church’s.SG.M.GEN go out.PTCP.PFV.M.NOM.SG and like vault.F.NOM.SG
sъtvorivъ se přide na drugŭi
 make.PTCP.PFV.M.NOM.SG REFL reach.AOR.3.SG on other.SG.M.ACC
chъlъmъ
 hill.SG.M.ACC
 ‘A very tall flame rose from the dome of the church, bent like a vault and reached another hill’ (*Life of Feodosij Pečerskij*, *Uspenskij Sbornik* f. 56a)

- (26) *onъ že priimъ jęgo*
 that.M.NOM.SG PTC receive.PTCP.PFV.M.NOM.SG 3.SG.M.GEN
oglasŭ jęgo slovesŭ dchovъnyimi
 catechize.AOR.3.SG 3.SG.M.GEN word.PL.N.INST spiritual.PL.N.INST
 ‘But he received him and instructed him in the holy scriptures’ (*Vita of St. Aninas the Wonderworker*, *Codex Suprasliensis* f. 140r)

Mental-state verbs such as *chotěti* ‘want’ and *věděti* ‘know’ generally involve a different reading than the typical temporal ones seen in the OCS dataset and are often likely to receive a causal interpretation, as in the conjunct participles in (27)-(28).

- (27) *dvjabolъ radvaše se semu. ne*
 devil.NOM rejoice.IMPV.3.SG REFL this.DAT.SG NEG
vědyi jako blizъ pogibelъ chotěše byti
 know.PTCP.IPFV.M.NOM.SG that near death.NOM want.IMPV.3.SG be.INF
emu
 3.SG.M.DAT

‘The devil rejoiced at that, not knowing [i.e. because he did not know] that his death was near’ (*Primary Chronicle*, *Codex Laurentianus* f. 26v)

- (28) *on že ne chotę krovŭ proliti.*
 that.M.NOM.SG PTC NEG want.PTCP.IPFV.M.NOM.SG blood.GEN.SG shed.INF
ne bi se s nimi
 NEG fight.AOR.3.SG REFL with 3.PL.M.INST

‘Not wanting to [i.e. since he did not want to] shed blood, he did not fight with him’. (*Suzdal Chronicle*, *Codex Laurentianus* f. 102r)

This, however, does not necessarily entail overall different functions than the one observed so far: as already mentioned, unless augmented with overt connectives (e.g. ‘because’, ‘when’, ‘while’), the semantics of participial adjuncts is underspecified, which is why a semantic framework integrating discourse relations was needed in the first place to fully account for dative absolutes and conjunct participles. *Chotěti* ‘want’ is found among the most frequent lemmas of both dative absolutes and conjunct participles. Among the latter it is most often used imperfectively and can generally be interpreted as an ELABORATION, as in (29) and (30).

- (29) *Michal* *že viděvš* *to*
 Michal.M.NOM.SG PTC see.PTCP.PFV.M.NOM.SG that.SG.N.ACC
sskoči s koně chotě pomoči
 jump off.AOR.3.SG from horse.SG.M.GEN want.PTCP.IPFV.M.NOM.SG help.INF
volodimeru
 Volodimer.M.DAT.SG
 ‘Michal, when he saw that, jumped off the horse wanting to help Volodimer’
 (*Suzdal Chronicle*, Codex Laurentianus f. 106r)

- (30) *i v monastyri poidoša ne chotjašte v*
 and in monastery.PL.F.ACC go.AOR.3.PL NEG want.PTCP.IPFV.F.NOM.PL in
polon poiti ot svoego grada vo
 captivity.SG.M.ACC go.INF from their.GEN.SG town.SG.M.GEN in
inye grady
 other.PL.M.ACC town.PL.M.ACC
 ‘And they went in monasteries, not wishing to go into captivity from their own town to other towns’ (*The Taking of Pskov*)

In dative absolutes, *chotěti* is not always used in its lexical meaning of ‘wanting to’, but rather as a modal auxiliary translatable as ‘being about to’, as in (31)-(32), where it can be interpreted as a generic FRAME (‘when *x* was about to do *y*, *z* happened’):

- (31) *na sud že jemu choteštu iti.*
 to judgement.ACC PTC 3.SG.DAT want.PTCP.IPFV.M.DAT.SG go.INF
plaka se mati otročete sego.
 cry.AOR.3.SG REFL mother.NOM child.GEN this.M.GEN.SG
glagoljušti
 say.PTCP.IPFV.F.NOM.SG
 ‘But when he was ready to go for the Judgement, the mother of the child cried, saying’ (*Vita Constantini* 2)

- (32) *i chotěvšū jemu po dvoju dniju*
 and want.PTCP.PFV.M.DAT.SG 3.SG.DAT on second.LOC day.LOC
otplouti javi se jemu v sně aggel
 sail.away.INF appear.AOR.3.SG REFL 3.SG.DAT in dream.LOC angel.NOM
gospodn *glagolē*
 lord.ADJ.NOM.SG say.PTCP.IPFV.M.NOM.SG

‘And as he was going to sail away on the second day, an angel of the Lord appeared to him in dream saying’ (*Vita of John the Hesychast*, Codex Suprasliensis f. 13v)

It is interesting to notice that *věděti* ‘know’ (all imperfective) also occurs as the participle lemma among dative absolutes but does not make it to the 10 most frequent lemmas, seemingly because they all occur in a specific literary genre and language variety (the OES chronicles⁹). If we look at these occurrences more closely, we see that their subjects are often names of kings (i.e. *jaroslavъ* ‘Jaroslav’, *mstislavъ* ‘Mstislav’, *volodimerъko* ‘Volodimirko’, *borisъ* ‘Boris’, *izjaslavъ* ‘Izjaslav’; see examples (33)-(34)). It is tempting to view this as a tendency for dative absolutes to occur in fixed or semi-fixed verb-subject combinations, especially considering that the common noun *кѡнѡзѡ* ‘king’, despite being among the most-frequent lemmas in the subcorpus, is never found as the subject of *věděti* ‘know’.

- (33) *jaroslavu že ne věduštju otbñě*
 Jaroslav.DAT PTC NEG know.PTCP.IPFV.M.DAT.SG father.ADJ.F.DAT.SG
smrti. varezi bęchu mnozi ou
 death.F.DAT Varangian.NOM.PL be.IMPF.3.PL much.NOM.PL at
jaroslava
 Jaroslav.GEN

‘While Jaroslav had not heard of his father’s death yet, many Varangians were under his command’ (*Primary Chronicle*, Codex Laurentianus f. 48a)

- (34) *volodimerku že togo ne věduštju so*
 Volodimerko.DAT PTC that.N.GEN NEG know.PTCP.IPFV.M.DAT.SG with
andrějem. i stasta u mičbska
 Andrej.INS and set.out.AOR.3.DU at Mičesk.GEN

‘Volodimerko and Andrej did not know that and went to Mičesk’ (*Suzdal Chronicle*, Codex Laurentianus f. 110r)

The rest of the occurrences of *věděti* ‘know’ among dative absolutes appear instead in clearly fixed expressions meaning ‘without anyone knowing’, ‘not known to anyone’, all in the OES subcorpus (as in (35)) except one example in *Suprasliensis* (36):

- (35) *ne věduštju nikomuže pridoch v pečeru.*
 NEG know.PTCP.IPFV.M.DAT.SG nobody.DAT come.AOR.1.SG in crypt.ACC
 ‘Without anyone knowing I went to the crypt’ (*Primary Chronicle*, Codex Laurentianus f. 70b)

⁹Of the 12 total occurrences, 5 are found in the *Suzdal Chronicle* (Codex Laurentianus), 2 in the *Primary Chronicle* (Codex Laurentianus); outside the chronicles, 4 are in the *Life of Feodosij Pečerskij* (Uspenskij Sbornik) and only 1 in the Codex Suprasliensis.

- (36) *i zatvori i v̄ chyzině p̄ti deset̄ ti*
 and lock-up.AOR.3.SG 3.SG.M.ACC in shack.LOC five.GEN ten.GEN PTC
p̄ti l̄t̄ s̄šta nikomuže inomu
 five.GEN year.GEN.PL be.PTCP.IPFV.M.GEN.SG nobody.DAT.SG other.DAT
ot̄ družiny věd̄štu jako episkup̄
 from company.GEN know.PTCP.IPFV.M.DAT.SG that bishop.NOM.SG
jest̄
 be.PRS.3.SG

‘And he locked him up in a shack when he was fifty-five, without anyone else from the company knowing that he was a bishop’ (*Vita of John the Hesychast*, Codex Suprasliensis f. 13r)

Overall, lexical variation among the constructions supports what we observed through other variables. Dative absolutes have a relatively consistent framing function throughout different configurations, as emerged from the overall smaller lexical variation among participle lemmas compared to conjunct participles, as well as from the lack of significant differences between sentence-initial dative absolutes in the SV and VS configurations. Conjunct participles, on the other hand, are functionally more varied, and the different functions are reflected in the different levels of lexical variation in different configurations, particularly depending on the position of the subject in sentence-initial occurrences, whereby sentence-initial VS conjunct participles are overall less lexically varied than those in the SV configuration, supporting the interpretation of the former as more typical FRAMES.

Summary

The analysis of lexical variation among *participles* showed that conjunct participles are overall more lexically varied than dative absolutes as a whole, except in the VS configuration, where conjunct participles show less lexical variation. When we cross these results together with the properties of subjects, we see that dative absolutes tend to occur in very predictable participle lemma-subject lemma combinations and, not infrequently, in what seems like fixed or semi-fixed expressions. These can occur across different texts and genres as ‘formulae’¹⁰ such as those seen with elaborating dative absolutes (of the type ‘God <HELP> <PERSON>’, where HELP is generally one of a limited set of verbs all with a similar semantics to *help* (e.g., again, *s̄obljudati* ‘guard, watch over’, *pomagati* ‘help’, *ukr̄epljati* ‘give strength’), but also variations of the basic phrase *ne věduštju nikomuže* (‘without anyone knowing’). But they may also occur as semi-fixed combinations contextually only to a specific text or genre, as in ‘reigning <RULER> in <PLACE>’ (e.g. *crstvujuštju obkse v̄ csrigradě* ‘when Oleksa was reigning in Constantinople’), which are, for example, found overwhelmingly more frequently in the chronicles.

¹⁰I am referring to *formulae* informally here since I have not systematically focussed on testing the formulaicity of dative absolutes, although there are strong reasons to think that such a study is warranted.

2.3 Conjunct participles and dative absolutes in strategically-annotated texts

In this final section, I summarize results from a case study which was carried out in Pedrazzini (2022) on strategically annotated treebanks. Pedrazzini (2022: 31-36) looked at the extent to which corpora with considerably fewer and shallower levels of annotation than deeply annotated and standard treebanks can be exploited to investigate a discourse-driven syntactic phenomenon (i.e. the discourse functions of participle clauses), specifically as a means of corroborating results emerging from treebanks with deeper annotation. This section first provides general frequencies on the position and aspect of conjunct participles and dative absolutes and lexical variation among participles (Section 2.3.1) as they emerged from the small corpus of strategically annotated treebanks as presented in Section 0.5.1 of the Introduction. However, the main part of the case study (Section 2.3.2) presents a ‘distant-to-close-reading’ analysis of a specific text, in the sense that it shows the extent to which closer-reading of automatically extracted occurrences of the constructions may complement potential patterns found from automatic methods. The text selected by Pedrazzini (2022) for this case study is the Slavonic *Story of Abraham of Qidun and his niece Mary* (AQM), contained in the 14th-century Middle Bulgarian Bdinski Sbornik (Dujčev 1972; Scharpe & Vyncke 1973). The reason for this choice is mainly incidental to the content of TOROT at the time of this study, where the vast majority of non-liturgical texts are of East Slavic provenance. AQM, despite being a translation from Greek, is instead of South Slavic (Middle Bulgarian) provenance and narrates the life of Abraham Qidun, a 4th-century Syrian hermit, and his niece Mary, in a rather different style and with a different vocabulary range than what we are used to from hagiographies included in previous datasets.¹¹

2.3.1 Position, aspect and lexical variation of participles

Tables 2.20 and 2.21 show the distribution of aspect among dative absolutes and conjunct participles preceding and following the matrix clause.

	imperfectives	perfectives
pre-matrix	47.3% (124)	52.7% (138)
post-matrix	59.9% (94)	40.1% (63)
tot.	52% (218)	48% (201)

Table 2.20: Dative absolutes in strategically annotated treebanks: aspect distribution by sentence position (row percentage)

¹¹For a detailed discussion of the text’s relation to its Greek sources, including translation techniques in this and related hagiographies, see Stern (2013, 2015, 2016, 2018). I am grateful to Dieter Stern for pointing me to the relevant literature on the topic and for kindly sending me copies of the cited material.

	imperfectives	perfectives
pre-matrix	27.7% (319)	72.3% (833)
post-matrix	64.5% (453)	35.5% (249)
tot.	41.6% (772)	58.4% (1082)

Table 2.21: Conjunct participles in strategically annotated treebanks: aspect distribution by sentence position (row percentage)

It is immediately clear that the distribution of aspect among conjunct participles follows the same pattern as the one observed in all the other datasets. Namely, the majority of pre-matrix conjunct participles are perfective, while the majority of post-matrix ones are imperfective. This suggests that, overall, it is very likely that their typical functions as INDEPENDENT RHEMES and ELABORATIONS, respectively, are also reflected in this fully automatically annotated dataset. The association between aspect and position is statistically significant, with a moderate to strong effect size ($\phi = 0.36$, Cramér's V): $\chi^2(1) = 242.08$, $p < 0.01$, with the odds of perfectives in pre-matrix position being 4.75 times higher than in post-matrix position. It is, in fact, not hard to find examples with the expected functions only by looking for different combinations of aspect and position in the sentence. (37)-(38) are examples of pre-matrix, perfective conjunct participles, clearly functioning as INDEPENDENT RHEMES. (39) is instead more likely an example of FRAME conjunct participles. The framing function of the latter can also be inferred from the fact that we have two coordinated conjunct participles, suggesting fronting and, therefore, a higher adjunction site than INDEPENDENT RHEMES. (40)-(41) are post-matrix, imperfective conjunct participles, clearly functioning as ELABORATIONS.

- (37) *abie že v̄stāv̄ iz̄m̄*
 immediately PTC stand up.PTCP.PFV.M.NOM.SG take.PTCP.PFV.M.NOM.SG
nož̄b̄ svoī i s̄tvorī ot dr̄eva kr̄st̄b̄
 knife and make.AOR.3.SG from tree.GEN.SG cross.ACC.SG
 ‘At once he stood up, took his knife and made a cross from the tree’ (Euthymius, *Praise for St. Constantine and Helena*, Sbornici of Vladislav Grammarian f. 428v)

- (38) *i on̄b̄ bo is̄pad̄ ot s̄đania*
 and that.NOM.SG because fall.PTCP.PFV.M.NOM.SG from seat.GEN.SG
kon̄naago padaet̄ na zem̄e
 horse's.GEN.SG fall.PRS.3.SG on ground.ACC
nizvr̄žen̄b̄ byv̄
 overthrow.PTCP.PASS.PFV.M.NOM.SG be.PTCP.PFV.M.NOM.SG
 ‘Because he will fall from the saddle and land on the ground, overthrown’
 (*Manasses Chronicles* p. 5447)

- (39) [*Sia* *že* *vəsi* ***slyšavše*** *i*
 this.N.ACC.PL PTC this.N.ACC.PL hear.PTCP.PFV.M.NOM.SG and
viděvše] *množaišq* *věrq* *i* *usrōdie* *kz*
 see.PTCP.PFV.M.NOM.SG more.F.ACC faith.F.ACC and zeal.N.ACC to
stmu *pokazovaachq*
 saint.M.DAT.SG show.IMP.3.PL
 ‘After hearing and seeing all these things, they started showing greater faith and
 zeal towards the Saint’ *Life of Ivan of Rila*, Zografski Sbornik f. 102r)
- (40) *staše* *ereie* ***čaušte*** *čļovanija*
 stand up.AOR.3.PL priest.NOM.PL wait.PTCP.IPFV.NOM.PL kiss.ACC.PL
ego
 3.SG.GEN
 ‘The priests stood up, waiting for his kiss’ (Jagičev Zlatoust, f. 29b)
- (41) *i* *slōzami* *sebe* *obliavz* *padaet* *na*
 and tear.INST.PL REFL dress.PTCP.PFV.M.NOM.SG fall.PRS.3.SG on
lici ***mole*** *se* *i* ***glagole***
 face.LOC pray.PTCP.IPFV.M.NOM.SG REFL and say.PTCP.IPFV.M.NOM.SG
 ‘And, in tears, he puts on clothes and falls on his face, praying and saying (*Life of
 Ivan of Rila*, Zografski Sbornik f. 96v)

Also among dative absolutes, there is a statistically significant correlation between aspect and position, although with a very low effect size ($\phi = 0.12$, Cramér’s V): $\chi^2(1) = 5.69$, $p = 0.02$, with the odds of perfectives in pre-matrix position being only 1.66 times higher than in post-matrix position. We can see that, as in standard treebanks, the difference in proportion between the two aspects (47.3% imperfective versus 52.7% perfectives) is nearly equal. A further test indicates that the difference is not statistically significant among pre-matrix absolutes,¹² whereas it is so among post-matrix ones.¹³ This result suggests that the observations made previously regarding the differences between pre- and post-matrix absolutes are likely supported by this dataset, namely, post-matrix absolutes may also work as ELABORATIONS (temporally identical to their matrix), or as clause-final restrictive adjuncts (temporally closer to FRAMES).

As Table 2.22 shows, the difference in lexical variation among the constructions is overall very similar to the one observed in the previous dataset.¹⁴ If we consider the total occurrences, lexical variation is smaller among dative absolutes than conjunct participles. For both constructions, lexical variation is greater among post-matrix than pre-matrix occurrences, but more evidently so for dative absolutes. The difference in variation is evident among pre-matrix occurrences, with dative absolutes showing a much smaller variation than conjunct participles. Variation among post-matrix occurrences is quite similar between the two constructions, although slightly lower for conjunct participles.

¹²Binomial test, $p = 0.21$.

¹³Binomial test, $p < 0.01$.

¹⁴Note that among post-matrix conjunct participles *glagolati* is once again used imperfectively in its quotative particle-like function, so it has been removed from the frequencies in the table.

Subsample	Absolutes		Conjuncts	
	10MFL	MATTR	10MFL	MATTR
total	33.42%	0.73	26.18%	0.79
left	39.84%	0.67	27.55%	0.78
right	28.77%	0.82	28.83%	0.80

Table 2.22: Lexical variation among participles by subcorpus in strategically annotated treebanks. *10MFL* = ten most frequent lemmas; *MATTR* = moving-average type-token ratio; *left* and *right* = pre-matrix and post matrix, respectively. Note that the closer the MATTR is to 1 (and the lower the percentage for the 10MFL), the greater the lexical variation.

Table 2.23 shows the ten most frequent lemmas among conjunct participles and dative absolutes, both overall and by position relative to the matrix clause.

Subcorpus	Absolutes	Conjuncts
total	<i>byti</i> ‘be’, <i>chotěti</i> ‘want’, <i>rešti</i> ‘say’, <i>prěiti</i> ‘pass away’, <i>žiti</i> ‘live’, <i>držati</i> ‘hold’, <i>priiti</i> ‘come, arrive’, <i>ostaviti</i> ‘leave’, <i>iměti</i> ‘have’, <i>ljubiti</i> ‘love’	<i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>chotěti</i> ‘want’, <i>slyšati</i> ‘hear’, <i>rešti</i> ‘say’, <i>vzěti</i> ‘take, seize’, <i>uvědět</i> ‘know’, <i>iti</i> ‘go’, <i>iměti</i> ‘have’, <i>ostaviti</i> ‘leave’
left	<i>byti</i> ‘be’, <i>chotěti</i> ‘want’, <i>prěiti</i> ‘pass away’, <i>rešti</i> ‘say’, <i>ostaviti</i> ‘leave’, <i>držati</i> ‘hold’, <i>sčkončati</i> ‘end, finish’, <i>všijati</i> ‘become light’, <i>žiti</i> ‘live’, <i>iti</i> ‘go’	<i>byti</i> ‘be’, <i>viděti</i> ‘see’, <i>slyšati</i> ‘hear’, <i>uvědět</i> ‘know’, <i>iti</i> ‘go’, <i>priiti</i> ‘come, arrive’, <i>vstati</i> ‘stand up’, <i>rešti</i> ‘say’, <i>vzěti</i> ‘take, seize’, <i>chotěti</i> ‘want’
right	<i>byti</i> ‘be’, <i>žiti</i> ‘live’, <i>ljubiti</i> ‘love’, <i>chotěti</i> ‘want’, <i>sčbrati</i> ‘gather’, <i>priiti</i> ‘come, arrive’, <i>iměti</i> ‘have’, <i>rešti</i> ‘say’, <i>vědět</i> ‘know’, <i>obladati</i> ‘rule, reign’	<i>byti</i> ‘be’, <i>chotěti</i> ‘want’, <i>viděti</i> ‘see’, <i>rešti</i> ‘say’, <i>tvoriti</i> ‘make’, <i>iměti</i> ‘have’, <i>vzěti</i> ‘take, seize’, <i>nositi</i> ‘carry, take’, <i>mnět</i> ‘think’, <i>ostaviti</i> ‘leave’

Table 2.23: Dative absolutes and conjunct participles in strategically annotated treebanks: ten most-frequent lemmas by position of participle relative to the matrix clause

Besides the high frequency of *byti* ‘be’ and certain verbs of motion such as *priiti* ‘come, arrive’ among both constructions, there are lemmas occurring only among the most frequent of one of the other construction, most evidently state and activity verbs such as *držati* ‘hold’, *žiti* ‘live’, and *obladati* ‘rule, reign’ among dative absolutes and telic verbs such as *vzěti* ‘take, seize’, *viděti* ‘see’, and *slyšati* ‘hear’ among conjunct participles. We can also find verbs among dative absolutes which are likely to occur in time expressions,

such as *vъsijati* ‘become light’, to be most likely found in framing expressions such as ‘when it dawned, when day came’ or similar, as in (42).

- (42) *Světú že vъsiovšú i povъsodu*
 light.DAT PTC dawn.PTCP.PFV.M.DAT.SG and everywhere
proliavšú se, i světlostio
 spread.PTCP.PFV.M.DAT.SG REFL and light.inst.sg
isplъnenomu i světozarnu dnju
 fill.PTCP.PASS.PFV.M.DAT.SG and bright.dat.sg day.dat.sg
byvšú, javišo se přěžde nevidimaa,
 be.PTCP.PFV.M.DAT.SG appear.AOR.3.PL REFL before invisible.N.NOM.PL
i mračnaa tōma prognana
 and dark.F.NOM.SG darkness.F.NOM.SG disperse.PTCP.PASS.PFV.F.NOM.SG
bъstō sianiemō světlyichō lučō
 bright.GEN.PL ray.GEN.PL
 ‘And when the light dawned and spread everywhere, and the day became bright and full of light, all the invisible things appeared and the dark gloom was dispersed by the shining of the bright rays’ (*Manasses Chronicles* p. 27)

Overall, aspect, position in the sentence and lexical variation all point to the properties of absolutes and conjunct participles that we observed in deeply annotated and standard treebanks. The results of the closer-reading analysis in the next section, which looks at the *Story of Abraham of Qidun and his niece Mary* from the *Bdinski Sbornik*, while agreeing with these observations, will also highlight a more specialized potential usage of absolute constructions as clause-bridging (and topic-shifting) devices.

2.3.2 Zooming in: dative absolutes and conjunct participles in a Middle Bulgarian text

Potential conjunct participles and dative absolutes were first extracted using the same criteria used for extracting the constructions from the rest of the strategically annotated treebank: dative absolutes by looking for participles in the dative with an ADVCL dependency relation, conjunct participles by looking for participles in the nominative also with an ADVCL dependency relation.

Before post-correction, 15 potential dative absolutes and 22 conjunct participles were identified on the basis of automatic annotation alone. All 22 potential conjunct participles were confirmed as actual conjunct participles. Of the potential dative absolutes, two turned out to be part of the same passive absolute construction (*prizvaně byvši* ‘having been called’), where one of the dative participles (*byvši*) is the copula (i.e. two of the potential absolutes were actually one). One occurrence was found to be a finite verb (*nareku* ‘I call, will call’), which was tagged as a dative participle, possibly because of the stem *narek-* (morphologically potentially suggesting a past active participle) and the *-u* ending (suggesting a masculine or neuter dative). Finally, two of the potential

occurrences were revealed to be regular dative participles agreeing with a preceding dative noun. All remaining potential absolutes (11 occurrences) were confirmed as such.

First, predictions were made based on the patterns observed in deeply annotated and standard treebanks. The majority of dative absolutes were expected to be found to the left of the matrix, while conjunct participles were expected to be found in both pre- and post-matrix position with some frequency, with pre-matrix occurrences more likely to be perfective and post-matrix occurrences more likely to be imperfective. This is indeed what we find, as Tables 2.24 and 2.25 show.

	imperfectives	perfectives
pre-matrix	4	7
post-matrix	0	0

Table 2.24: Dative absolutes in AQM: aspect distribution by sentence position

	imperfectives	perfectives
pre-matrix	0	19
post-matrix	3	0

Table 2.25: Conjunct participles in AQM: aspect distribution by sentence position

All dative absolutes occur in pre-matrix position, both imperfectively and perfectly. Conjunct participles occur overwhelmingly more in pre-matrix position, with only three occurrences following the matrix. All pre-matrix occurrences are perfective, whereas the only three post-matrix conjunct participles are imperfective. This is in line with the typical functions of pre-matrix conjunct participles as INDEPENDENT RHEMES and of post-matrix ones as ELABORATIONS.

Tables 2.26 and 2.27 show all the subject lemmas and their frequencies for dative absolutes and conjunct participles respectively.

subject	<i>n</i>
<i>*i</i> (3 rd pers. pron.)	6
<i>sv</i> ‘this’	1
<i>ot'cb</i> ‘father’	1
<i>l'eto</i> ‘year’	1
<i>utro</i> ‘morning’	1
<i>dnb</i> ‘day’	1

Table 2.26: Dative absolutes in AQM: subject lemmas

Very similarly to dative absolutes in both deeply annotated and standard treebanks, the majority of the subjects of dative absolutes are third-person personal pronouns (6 occurrences), as in (43).

subject	<i>n</i>
<i>on̄</i> ‘that one’	3
<i>sam̄</i> ‘self’	2
<i>znaem̄</i> ‘acquaintance’	1
<i>ogn̄</i> ‘fire’	1
<i>poslan̄</i> ‘(the) sent (one)’	1
null	14

Table 2.27: Conjunct participles in AQM: subject lemmas

- (43) *potom̄ s̄edeštu jem̄u na odr̄ě, reč kb*
 after sit.PTCP.IPFV.M.DAT.SG 3.SG.DAT on bed.LOC say.AOR.3.SG to
n̄jemu
 3.SG.DAT
 ‘After that, as he was was sitting on the bed, she said to him (AQM, Bdinski Sbornik f. 10v)

Time-denoting nouns, namely *dn̄n̄* ‘day’ and *l̄eto* ‘year’ (one occurrence each), also occur as subjects in dative absolutes and appear in constructions very similar to the widespread generic temporal FRAMES found in the Gospels. A particularly clear echo of the latter can be detected in (44), where *utro* ‘morning’ would be analyzed in TOROT as a predicative complement, consistently with similar examples headed by *byti* ‘be’ (e.g. *pozdě byv̄ošu* ‘once it got late’, Matthew 14:23).

- (44) *outru že byv̄ošu reč kb nei*
 morning.DAT PTC be.PTCP.PFV.N.DAT.SG say.AOR.3.SG to s3.SG.DAT
 ‘When morning came, he said to her’ (AQM, Bdinski Sbornik f. 13r)

All 10 overt subjects of the dative absolutes follow the participle, which, as we already commented on, can be associated with their framing function, whereby older referents get reinstated in a new discourse segment.

Most conjunct participles have a null subject. The demonstrative *on̄* ‘that one’ is the most frequent among overt subjects, which was also the case in the previous dataset. Of the 8 overt subjects, three precede the participle and five follow it. VS conjunct participles are once again clearly more ambiguous between a FRAME and an INDEPENDENT RHEME interpretation. Compare (45) and (46).

- (45) *slyšav̄ši že ona ěko kamen̄*
 hear.PTCP.PFV.F.NOM.SG PTC that.F.NOM as stone.NOM.SG
bezvd̄šen̄ přebys v̄ roukou ego
 without.soul.NOM.SG remain.AOR.3.SG in arm.LOC.DU his
 ‘After listening, she remained like a soulless stone in his arms’ (AQM, Bdinski Sbornik f. 12r)

- (46) *ona že vstavši i obvetb i*
 she PTC stand-up.PTCP.PFV.F.NOM.SG and embrace.AOR.3.SG 3.SG.M.ACC
i načetb lobizati po vii ego
 and start.AOR.3.SG kiss.INF on neck.LOC 3.SG.M.GEN
 ‘She stood up, and embraced him, and started kissing his neck’ (AQM, Bdinski Sbornik f. 8r)

While (46) is unmistakably an INDEPENDENT RHEME, a FRAME reading for (45) seems more likely. In the latter, one may also notice a causal relation between the participle clause and its matrix, which, as we have seen in the previous section, is more likely to emerge from FRAMES and generally becomes more evident when *framing expressions* follow the matrix.

Most null-subject pre-matrix conjunct participles, as in (47), more clearly receive an INDEPENDENT RHEME reading.

- (47) *vstavši idi sb nami*
 stand up.ptcp.pfv.f.nom.sg go.imp.2.sg with us.inst.pl
 ‘Stand up and come with us’ (AQM, Bdinski Sbornik f. 12v)

In some instances, a FRAME reading is instead triggered by the fact that the participle in question is not interpreted as modally dependent on the matrix verb (cf. Section 0.2.2 of the Introduction). In (48), for example, *izbšedb* is not part of the indirect question (signalled by *li*) nor of the negation (i.e. the question is arguably whether the fire will *burn*, not whether it will *come out*).

- (48) *ašte že i drvznu ognb izbšedb*
 if PTC even dare.PRS.1.SG fire.M.NOM.SG come out.PTCP.PFV.M.NOM.SG
izb dvercb těch ne sžžet li mene
 from door.gen.pl that.GEN.PL NEG burn Q 1.SG.GEN
 ‘Even if I dared, won’t the fire burn me when it comes out of those doors?’ (AQM, Bdinski Sbornik f. 4v)

Some examples, such as (49), are, as is common with sentence-initial conjunct participles, ambiguous between a FRAME and an INDEPENDENT RHEME reading.

- (49) *i rastrzavši ryzi svoe bišše se po*
 and tear.PART.PRF.M.NOM.SG robe.ACC.PL her beat.IMP.3.SG REFL on
licju i chotěše se sama udaviti ot
 face.DAT and want.IMP.3.SG REFL self suffocate.INF from
pečaly
 sorrow.GEN.SG

‘And after tearing her own robes, she started striking her own face and wanted to suffocate herself from sorrow’ (AQM, Bdinski Sbornik f. 3v)

participle	<i>n</i>
<i>sěděti</i> ‘sit’	2
<i>byti</i> ‘be’	1
<i>pěti</i> ‘sing’	1
<i>lobyzati</i> ‘kiss’	1
<i>umřěti</i> ‘die’	1
<i>prizvatı</i> ‘summon’	1
<i>vřniti</i> ‘enter’	1
<i>prılučiti (se)</i> ‘happen’	1
<i>končati</i> ‘finish’	1
<i>pomesti</i> ‘change, replace’	1

Table 2.29: Dative absolutes in AQM: frequency of participle lemmas

- (52) *i vřstavřaa idosta abie*
 and stand.up.PTCP.PFV.M.NOM.DU go.AOR.3.DU immediately
 ‘And they stood up and left without further ado’ (AQM, Bdinski Sbornik f. 13r)

- (53) *ona ře otvěřtavři reč kb*
 that.NOM.F.SG PTC answer.PTCP.PFV.F.NOM.SG say.AOR.3.SG to
nemu
 3.SG.M.GEN
 ‘She then replied to him saying’ (AQM, Bdinski Sbornik f. 13r)

As in the previous datasets, telic verbs, such as *vřzeti* ‘take, seize’, *prijęti* ‘receive’, *slyřati* ‘hear’, *ızeti* ‘remove’, are particularly common among conjunct participles. As we observed, these can be easily envisioned as part of a series of events following one another, as in (54).

- (54) *ıřm ře cętu samř dastř*
 take out.PTCP.PFV.M.NOM.SG PTC daric.ACC self.M.NOM.SG give.AOR.3.SG
gostınyku
 innkeeper.DAT.SG
 ‘The blessed man took out a daric and handed it to the innkeeper’ (AQM, Bdinski Sbornik f. 9r)

Verbs of motion, constituting the majority in the previous two case studies, are also represented in the new datasets for both constructions, as in the conjunct participle in (55) and the dative absolute in (56).

- (55) *i vřzlęřř sęde na odrę*
 and climb.up.PTCP.PFV.M.NOM.SG sit.AOR.3.SG on bed.LOC
 ‘So he climbed up and sat on the bed’ (AQM, Bdinski Sbornik f. 10r)

- (56) *i vŝedŝima že ima. vidě odrb*
 and enter.PTCP.PFV.DAT.DU PTC 3.DU.M.DAT see.AOR.3.SG bed.ACC
vysokb nastlanb
 large.ACC spread.PTCP.PFV.PASS.ACC.SG
 ‘As they entered, he saw a large bed made up’ (*AQM*, Bdinski Sbornik f. 10r)

If taken out of context, some of the lemmas among dative absolutes hardly seem ‘predictable predications’ from the (Old) Church Slavonic standpoint, based on the pool of lemmas observed in the previous dataset. This is the case with *lobyzajušti* ‘kissing’ in (57).

- (57) *ona že vstavši i objetb i*
 she PTC stand-up.PTCP.PFV.F.NOM.SG and embrace.AOR.3.SG 3.SG.M.ACC
i načetb lobizati po vii jego.
 and start.AOR.3.SG kiss.INF on neck.LOC 3.SG.M.GEN
lobyzajušti že jee. vzbvovnja ubo vonjeju
 kiss.PTCP.IPFV.F.DAT.SG PTC 3.SG.F.GEN exhale.AOR.3.SG then smell.INS
čr̃norizbčvskoju tělo jego
 monk.ADJ.INS body.NOM 3.SG.M.GEN
 ‘She stood up, and embraced him, and started kissing his neck. As she was kissing him, the smell of monasticism issued from his body’ (*AQM*, Bdinski Sbornik f. 8r)

By close-reading the surrounding discourse, however, it becomes clear that even uncommon lemmas do not present new information, but refer back to events that had already been introduced in the previous discourse. The dative absolute’s lemma *lobyzati* ‘kiss’ in (57), for example, had already occurred as part of the main predication in the previous sentence (*načetb lobizati* ‘(he) started kissing’), and gets reinstated by the dative absolute not as a main event but as a FRAME for a following new event.

It is tempting to see this as a general feature of the phenomenon, whereby dative absolutes may often be used as clause-bridging devices. After closely checking the discourse surrounding each dative absolutes, for most of the occurrences, either one of two patterns emerged. In some of the occurrences, the dative absolute’s lemma appeared as the main verb in one of the previous two sentences, as in (57) above and (58). In other occurrences, the subject of the dative absolute appeared in the previous sentence as a focal¹⁵ element or as one of the event participants in a foregrounded eventuality (59)-(60).

- (58) *prizovi mi ju da se poveselju dñs*
 summon.IMP 1.SG.DAT 3.SG.F.ACC so-that REFL enjoy.PRS.1.SG today
s njeju. zělo bo jako slyšalb jesm
 with 3.SG.F.INST much for that hear.PTCP.RESULT.M.SG be.PRS.1.SG

¹⁵The information-structural notion of focus is intended here, along with Halliday (1967: 204) as ‘a message block which he [the speaker] wishes to be interpreted as informative. What is focal is “new” information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse’. In this sense, it overlaps with the notion of ‘rheme’ in other terminologies.

o njei godě mi jestь.
 about 3.SG.F.LOC pleasant.ADV 1.SG.DAT be.PRS.3.SG
prizvaně že **byvši** *ei.*
 summon.PTCP.PFV.PASS.F.DAT.SG PTC be.PTCP.PFV.F.DAT.SG s3.SG.DAT
pride k njemu
 come.AOR.3.SG to 3.SG.DAT

‘Summon her, so that I can enjoy myself with her today. From what I’ve heard of her, I am much attracted by her’. After being summoned, she came to him.’ (AQM, Bdinski Sbornik f. 7v)

- (59) *k lět, čr̃ñnorizbč̃stvova sь nimь jako*
 twenty year.GEN.PL live-monastically.AOR.3.SG with 3.M.INST.SG like
agnica čstaa, i golubyca neskr̃ṽnnaa. i
 lamb.NOM pure.NOM and dove.NOM innocent.NOM and
končavšu *se dvadeset̃nu lētu vř̃meni. i*
 end.PTCP.PFV.N.DAT.SG REFL twentieth.N.DAT year.DAT time.GEN and
neistoṽstvo s̃tvori dijavol̃ na nju i sět̃
 fury.ACC make.AOR.3.SG devil.NOM on 3.SG.F.ACC and snare.ACC
poleče da ulovit̃ ju.
 set-up.AOR.3.SG to catch.PRS.3.SG 3.SG.F.ACC

‘For twenty years she lived monastically after him, like a chaste lamb, and an innocent dove. When twenty years had elapsed, the Devil turned his fury on her and set up nets to ensnare her’ (AQM, Bdinski Sbornik 2v)

- (60) *iměše blaženy sь brata po pl̃ti.*
 have.IMPF.3.SG blessed.NOM this.NOM brother.ACC by flesh.DAT
imušta d̃b̃štere edinočedju. umeršu
 have.PTCP.IPFV.M.ACC.SG daughter.ACC only.ACC die.PTCP.PFV.M.DAT.SG
že ocju jee. př̃bys že junotka s̃r
 PTC father.DAT 3.SG.F.GEN remain.AOR.3.SG PTC girl.NOM orphan.NOM
 ‘The blessed one [Abraham] had a brother who had an only daughter [Mary].
 When the father [of Mary] died, the girl remained orphan.’ (AQM, Bdinski
 Sbornik f. 1v)

In (58), the dative absolute *prizvane že byvši ei* ‘after she was called’ picks up the lemma of the imperative *prizovi ju* ‘call her’. In (59), *20 let* ‘for 20 years’ is also focussed (implying ‘for as long as 20 years’), whereas it then reappears as the subject of the dative absolute, where it is topical (*končavšu se dvadeset̃nu letu vř̃meni* ‘after 20 years had passed’). In (60), *ocju* ‘father’, the dative absolute’s subject, and *brata* ‘brother’, the object of the previous sentence, are actually the same referent, but the different lemma is due to a topic-shift, from the referent’s brother (Abraham) to the referent’s daughter (Mary).

It is clear that, in this dataset, dative absolutes are quite systematically used as clause-bridging devices, whereby previously foregrounded eventualities or event participants are ‘repurposed’ as part of the background for a new portion of foregrounded

discourse to come. As we will discuss in more detail in the second part of the thesis (see Section 6.3.2 of Chapter 6), the clause-bridging function of absolutes observed in this dataset is consistent with the very clear typical framing function of the construction and find clear parallels in the world's languages (particularly switch-reference systems; cf. Section 6.3.2 of Chapter 6).

2.4 Summary

This first part of this chapter analysed the properties of dative absolutes and conjunct participles as they emerge from standard treebanks, a much larger (in terms of number of tokens) and more diverse (in terms of genre and historical periods) dataset than the Codex Marianus. As in Chapter 1, the analysis focussed on the extent to which the functions of participle constructions can be derived indirectly from (a combination of) variables such as position in the sentence, aspect distribution, properties of their subjects, and lexical variation among the participles.

As in the Codex Marianus, conjunct participles showed a strong association between position relative to their matrix clause and their aspect. This supports the observation, already made in Chapter 1, that the most typical functions of pre-matrix conjunct participles, which are much more likely to be perfective, is as INDEPENDENT RHEMES, whereas post-matrix conjunct participles, which are much more likely to be imperfective, are much more typically ELABORATIONS.

Among dative absolutes, post-matrix occurrences, while still less frequent (13.9%), are not as uncommon as the first dataset suggested. The majority of post-matrix absolutes are very similar to FRAMES from the temporal semantic perspective, and several of these may be considered, following Fabricius-Hansen & Haug (2012b), as clause-final restrictive adjuncts, which are very similar to FRAMES in that they restrict the domain of the matrix clause by introducing a *state* within which the main eventuality holds. Like clause-final restrictive adjuncts, several post-matrix absolutes may receive a (strongly or mildly) contrastive reading. However, based on their temporal semantic contribution, we also found evidence that a smaller number of post-matrix occurrences can be interpreted as ELABORATIONS, whose runtime is identical to the one of the matrix eventuality (unlike FRAMES, which have their own runtime and may include the matrix runtime).

We then analysed the properties of subjects, focussing on sentence-initial participle constructions. The subjects of sentence-initial conjunct participles showed similar properties to those observed in Chapter 1. Anaphoric parts of speech, namely demonstratives and personal pronouns, are much less frequent in the VS configuration, which can be explained by the specific function of VS conjunct participles of reintroducing older or inactive referents in the discourse. Among sentence-initial dative absolutes, demonstratives, as with conjunct participles, are much more common in the SV configuration, whereas personal pronouns are much more frequent so in the VS one. The analysis of

lexical variation both among subjects and among the participles themselves supports the intuition that dative absolutes as a whole, but particularly the VS configuration, are more likely to attract highly discourse-prominent referents as their subjects and to be found in fixed expressions (e.g. *ne věduštju nikomuže* ‘without anyone knowing’ and variations thereof) or predictable subject-verb lemma combinations depending on text or genre (e.g. ‘reigning <RULER> in <PLACE>’, as in *crstvujuštju obkse vč csrigradě* ‘when Oleksa was reigning in Constantinople’). The new data generally suggests, besides the pool of typically framing verbs common to most texts and historical stages, different genres and language varieties might involve different high-frequency verbs and subjects. The phenomena thus seem to lend itself well to genre-based distributional analyses.¹⁶

The patterns observed in strategically annotated treebanks are consistent with what we observed in the other two datasets. The closer-reading analysis of a Middle Bulgarian Slavonic text, the *Story of Abraham and Mary His Niece* from the Bđinski Sbornik, besides confirming the general patterns found previously, showed that absolutes in this text systematically work as a clause-bridging (or topic-shifting) device. This either emerged from the predicates themselves, whereby a previous foreground clause is repurposed as a background clause using the dative absolutes, or by only repeating event participants that may have been focussed (or constituents of a foreground clause) in the previous discourse and are shifted to topics as subjects of a dative absolute.

¹⁶See Kurešević (2006) for a study on dative absolutes using a genre-based approach, although not from a quantitative perspective.

Chapter 3

Finite competitors: *jegda*-clauses

3.1 *Jegda*-clauses as finite competitors

Distributional overlap between participle constructions and *jegda*-clauses is evident from parallel examples such as (1)-(2) and (3)-(4).

- (1) a. *i egda pride v̄o crkv̄o. pristopiše*
and when come.AOR.3.SG in temple.ACC.F.SG move forward.AOR.3.PL
k̄o nemu učestju. archierei i
to 3.SG.M.DAT teach.PTCP.IPFV.M.SG.DAT chief priest.PL.M.NOM and
star̄ci ljud̄stii
elder.PL.M.NOM people.PL.M.NOM
- b. *Kai elthontos autou eis to*
and come.PTCP.PFV.M.GEN.SG 3.SG.M.GEN in the.N.ACC.SG
hieron prosēlthon autōi
temple.N.ACC.SG move forward.AOR.3.PL 3.SG.M.DAT
didaskonti hoi arkhierais kai
teach.PTCP.IPFV.M.DAT.SG the.PL.M.NOM chief priest.M.NOM.PL and
hoi presbuteroi tou laou
the.M.NOM.PL elder.M.NOM.PL the.M.GEN.SG people.M.GEN.SG
‘And when he entered the temple, the chief priests and the elders of the
people came to him as he was teaching’ (Matthew 21:23)
- (2) a. *prišed̄šou že emu v̄o dom̄o.*
come.PTCP.PFV.M.SG.DAT PTC 3.SG.M.DAT in house.SG.M.ACC
pristo[p]ste k̄o nemu slēp̄ca.
move forward.AOR.3.DU to 3.SG.M.DAT blind man.DU.M.NOM
- b. *elthonti de eis tēn oikian*
come.PTCP.PFV.M.DAT.SG PTC in the.F.ACC.SG house.F.ACC.SG
prosēlthon autōi hoi
move forward.AOR.3.PL him.3.SG.M.DAT the.M.NOM.PL
tuphloi
blind man.M.NOM.PL
‘When he entered the house, the blind men came to him’ (Matthew 9:28)

- (3) a. *egda molite se glte Otče našo*
 when pray.PRS.2.PL REFL say.IMP.2.PL Father.VOC 1.PL.M.VOC
iže esi na nbsch
 that.M.NOM.SG be.PRS.2.SG in heaven.LOC.PL
- b. *hotan proseukhēsthe legete pater*
 when pray.SBJV.PRS.2.PL say.IMP.PRS.2.PL Father.VOC
hagiasthētō to onoma sou
 dedicate.IMP.AOR.PAS.3.SG the.N.NOM.SG name.N.NOM.SG 2.SG.GEN
 ‘When you pray, say: “Our Father in heaven, Hallowed be Your name. Your kingdom come”’ (Luke 11:2)
- (4) a. *Molešte že se ne lichō glte*
 pray.PTCP.IPFV.NOM.PL PTC REFL NEG evil.N.ACC.SG say.IMP.2.PL
ěkože i języčōnici
 like also heathen.M.NOM.PL
- b. *Proseukhomenoi de mē battalogēsēte hōsper*
 pray.PTCP.IPFV.NOM.PL PTC NEG speak idly.SBJV.AOR.2.PL like
hoi ethnikoī
 the.M.NOM.PL heathen.M.NOM.PL
 ‘And when you pray, do not use vain repetitions as the heathen do’
 (Matthew 6:7)

(1) and (2) are an example of clear functional overlap between *jegda*-clauses and dative absolutes. They involve the same lemma and verbal aspect, and in both examples the subject of the adverbial is not co-referential with that of the matrix clause. Likewise, (3) and (4), involving a *jegda*-clause and a conjunct participle, respectively, are structurally very similar, with a framing adverbial (‘when praying’) followed by a command.

Such parallel examples can also be easily found outside of the Gospels. See, for instance, (5)-(6) and (7)-(8), showing the use of *jegda*-clauses and conjunct participles, and of *jegda*-clauses and dative absolutes, respectively, in very similar contexts.

- (5) *i egda viditъ mestъ rucě svoi*
 and when see.PRS.3.SG vengeance.F.ACC.SG hand.F.DU.ACC his.3.DU.F.ACC
umyetъ v krovī grěšnika
 wash.PRS.3.SG in blood.LOC sinner.M.GEN.SG
 ‘And when he sees vengeance, he will wash his hands in the blood of the sinner’
 (*Primary Chronicle*, Codex Laurentianus f. 78d)
- (6) *on že vidę se i*
 that.M.SG.NOM PTC see.PTCP.IPFV.M.SG.NOM REFL and
stuživ si ostavlęet
 grieve.PTCP.PFV.M.SG.NOM this.3.SG.M.DAT leave.3.SG.PRS.ACT
pustynju kupno ž i brata svoego
 hermitage.F.SG.ACC together with also brother.M.GEN.SG his.3.SG.M.GEN

‘When he sees these things and starts grieving, he will leave the hermitage, as well as his own brother’ (*Life of Sergij of Radonezh*, f. 61r)

- (7) *egda ispolni se m dnii povelě*
 when finish.AOR.3.SG REFL forty day.M.GEN.PL command.AOR.3.SG
imъ iti v goru elevonьskuju
 3.PL.M.DAT go.INF in mount.SG.F.ACC olive’s.SG.F.ACC
 ‘When forty days had passed, he commanded them to go to the Mount of Olives’
 (*Primary Chronicle*, Codex Laurentianus f. 35v)

- (8) *i koncejuštju se lět tomu.*
 and finish.PTCP.IPFV.N.DAT.SG REFL year.SG.N.DAT that.SG.N.DAT
vygnaša žirolava is posadnicьstva. i
 drive out.AOR.3.PL Žirolav.SG.M.GEN from posadnichestvo.SG.N.GEN and
daša zavidu nerevinicju
 give.AOR.3.PL Zavid.SG.M.DAT Nerevinič.SG.M.DAT
 ‘And when that year was ending, they drove Žirolav out of the *posadnichestvo*
 and gave it to Zavid Nerevinič’ (*Novgorod First Chronicle* f. 40r)

The competition between *jegda*-clauses and participle constructions was also clear from some of the mismatches between Greek and Slavic seen in Section 1.5 of Chapter 1. In (9) and (10), for instance, Greek uses a genitive absolute and a nominalized accusative with infinitive construction, respectively, whereas in both cases Slavic has a *jegda*-clause.

- (9) a. *da ne egda položiti osnovaniě. i ne*
 so that NEG when lay.PRS.3.SG foundation.ACC and NEG
možetъ sьvršiti. vsi videštei
 be able.PRS.3.SG finish.INF all.NOM.PL see.PTCP.IPFV.NOM.PL
načnъti rogati se emu ěko
 begin.PRS.3.PL mock.INF REFL 3.SG.DAT that
- b. *hina mēpote thentos autou themelion kai*
 thus not-ever lay.PTCP.PFV.M.GEN.SG he.GEN.SG foundation.ACC and
mē ischyontos ektelesai pantes hoi
 NEG be-able.PTCP.IPFV.M.GEN.SG finish.AOR.INF all.NOM.PL the
heōrountes arxōntai autōi empaizein
 see.PTCP.IPFV.M.NOM.PL begin.SBJV.AOR.MID he.DAT mock.INF.PRS
legontes hoti
 say.PTCP.IPFV.NOM.PL that
 ‘lest, after he has laid the foundation, and is not able to finish, all who see it
 begin to mock him’ (Luke 14:29-30)
- (10) a. *bystъ že egda približi se isъ vъ*
 happen.AOR.3.SG PTC when approach.AOR.3.SG REFL Jesus.NOM in
ericho. slēpecъ eterъ sēdēaše pri poti
 Jericho.ACC blind-man.NOM certain.NOM sit.IMP.F.3.SG beside way.LOC
prose
 beg.PTCP.IPFV.NOM.SG

- b. *egeneto* *de en tōi* *engizein* *auton eis*
 happen.AOR.3.SG PTC in the.N.DAT approach.PRS.INF he.ACC to
Hiereichō typhlos tis ekathēto para tēn
 Jericho.ACC blind-man.NOM certain.NOM sit.IMPV.3.SG beside the
hodon epaitōn.
 road.ACC beg.PTCP.IPFV.M.NOM.SG
 ‘It came to pass that, as Jesus approached Jericho, a blind man was sitting
 by the road begging’ (Luke 18:35)

In Chapter 1, we observed that, even though a large number of dative absolutes correspond to genitive absolutes, there is one coherent group of mismatches where Greek has a nominalized accusative with infinitive and Old Church Slavonic has a dative absolute. We also saw that *bystǎ-/egeneto*-clauses are somewhat formulaic, in that they generally involve slots filled by constructions with predictable discourse functions—with the slot between *bystǎ* and the main verb occupied by a framing, forward-looking construction, in many cases corresponding to an absolute construction. In (10) the *jegda*-clause fills that same slot and, in fact, if we compare the Greek constructions corresponding to *jegda*-clauses in Slavic (Table 3.1), we can see that accusative with infinitives, just like with dative absolutes, are translated relatively often with a *jegda*-clause.

construction	<i>n</i>
finite subordinate	124
accusative with infinitive	24
genitive absolute	2
backward-controlled participle	1
conjunct participle	1

Table 3.1: Ancient Greek constructions corresponding to a *jegda*-clause in Old Church Slavonic

With the goal of clarifying what drives the choice between a *jegda*-clause and participle constructions, this chapter and the next look into the issue of competition from the perspectives of temporal relations and discourse interpretation. Our approach to the analysis of *jegda*-clauses will necessarily differ in some important respects from the one taken so far to the analysis of participle clauses.

First, because of their finiteness, the addition of the tense variable to the aspectual one introduces some complexity at the level of temporal reference that did not concern the temporal interpretation of participle clauses. As discussed in Section 0.3 of the Introduction, there are several distinctions at the broad level of event structure (e.g. event/states, durativity/punctuality, telicity/atelicity) that interact with tense-aspect in the interpretation of the temporal relations between finite temporal adverbials and their matrix clause as much as between sequences of independent clauses. As noted in the Introduction, this type of complexity, specific to finite clauses, is reflected, for instance, in the temporal symmetry (or absence thereof) that has long been observed between *when*-clauses and their matrix clauses when either of the two is imperfective (Kamp & Rohrer 1983; Partee 1984; Hinrichs 1986; Sandström 1993; Bonomi 1997; Sæbø 2012), as in (11).

- (11) a. Irina and I were preparing the documents when Boris called.
 b. When Irina and I were preparing the documents Boris called.

The temporal relation between the imperfective clause (*were preparing*) and the perfective clause (*called*) remains the same if the roles of the two clauses are swapped: in this case, the eventuality described by the perfective clause is included in the eventuality described by the imperfective clauses in both (11a) and (11b). What may have changed is their information structure and/or the discourse relations involved. We can assume a similar situation to hold in Early Slavic. In (12a), for example, the *jegda*-clause is imperfective, whereas its matrix clause is perfective, and their temporal relation is one of inclusion of the bounded matrix in the unbounded *jegda*-clause. This same temporal relation would be preserved, without having to change the tense-aspect of either of the predicates, if the roles were swapped, which would give (12b).¹

- (12) a. [*i egda sěaše*] [ovo pade pri poti]
 and when sow.IMPF.3.SG this.NOM.SG fall.AOR.3.SG along path.LOC
 ‘And when he was sowing some [seeds] fell along the path’ (Luke 8:5)
- *b. [*i sěaše*] [*egda ovo pade pri poti*]
 and sow.IMPF.3.SG when this.NOM.SG fall.AOR.3.SG along path.LOC
 ‘And he was sowing when some [seeds] fell along the path’

While (12b) is not, of course, a real example, similar readings are easily found, as in (13).

- (13) *Toma že edin̄ ot̄ oboju na desęte naricaemy*
 Thomas.SG.M.NOM PTC one.SG.M.NOM from twelve called
bliznec̄. ne bě tu s̄ nimi egda
 Twin.SG.M.NOM NEG be.IMPF.3.SG here with 3.PL.M.INST when
pride is
 come.AOR.3.SG Jesus.NOM
 ‘Now Thomas, one of the twelve, called the Twin, was not with them when Jesus came’ (John 20:24)

Now, the same observations do not seem to apply to non-finite adjuncts. (14a), for instance, is an almost identical example to (12) above, except that a dative absolute is used instead of a *jegda*-clause. The dative absolute is imperfective, whereas the matrix clause is perfective, and the temporal relation between the two, once again, is one of inclusion of the perfective event in the imperfective event. Albeit a very dubious procedure in itself when it comes to nonfinite adjuncts, to test the possibility of temporal symmetry between a participle construction and its matrix clause, in order to swap their role, we would have to add tense distinctions to the first clause and, conversely, remove

¹The asterisk here does not indicate ungrammaticality, but simply that it is not an attested example.

them from the second clause. Potentially, if we made the first clause a past-tense one, we would obtain an imperfect, as in the *jegda* example in (12a), whereas the second clause would become a perfective dative absolute, presumably resulting in (14b).

- (14) a. [*i sějōštoumou*] [*ova oubo padō pri*
and sow.PTCP.IPFV.M.DAT.SG this.NOM.PL fall.AOR.3.PL along path
poťi]

‘And when he was sowing some [seeds] fell along the path’ (Matthew 13:4)
- *b. [*i sěaše*] [*ověmъ oubo padъšimъ*
and sow.IMPF.3.SG this.DAT.PL fall.PTCP.IPFV.M.DAT.PL along
pri poťi]
path
?’And he was sowing, as some [seeds] had fallen along the path’

If such configuration was at all possible, then the event described by the dative absolute in (14b) would most likely receive a pluperfect interpretation, whereby it is interpreted as having occurred sometime before the imperfect clause, not one of inclusion of the event described by the absolute construction in the matrix event. We find very few of such occurrences (i.e. an imperfect matrix followed by a perfective dative absolute, but (15) is one such example, where the absolute is indeed interpreted as a pluperfect.

- (15) *ězdechu po onoi storoně dněpra*
ride.IMPF.3.PL along one.SG.F.DAT side.SG.F.DAT Dnepr.SG.M.GEN
ljudi emljušte a drugyja
people.PL.M.ACC capture.PTCP.IPFV.NOM.PL and other.PL.M.ACC
sěkušte ně outegšimъ perevesti se
kill.PTCP.IPFV.NOM.PL NEG manage.PTCP.PFV.DAT.PL cross over.INF REFL
imъ
3.PL.M.DAT
‘They were riding along the other side of the river Dnepr, capturing some people and killing others, as they had not been able to cross over’ (*Kiev Chronicle*, year 6643, Codex Hypatianus)

Second, *when*-clauses are *overtly subordinated* temporal adverbials, even though the temporal relation conveyed by the subordinator *when* is semantically quite underspecified compared to more explicit temporal connectives, such as *after* or *before*. The temporal subordinator *when* has, in fact, long been identified as a clue to infer rhetorical relations closely connected to frame adverbials: most notably, SDRT considers it an ‘easily recognized trigger’ of BACKGROUND structures (Asher et al. 2007: 13), which immediately places *jegda*-clauses in competition specifically to participle FRAMES, rather than participle constructions in any configurations. This narrows down our investigation from the very start and allows us to focus particularly on differences and similarities

between *jegda*-clauses and participle constructions functioning as FRAMES.

If the main point of reference for the study of participle clauses was Bary & Haug's (2011) framework and Haug's (2012) corpus evidence, the analysis of *jegda*-clauses draws from the plethora of studies on the topic of temporal relations between temporal adverbials, including *when*-clauses, and their matrix clause, particularly Hinrichs (1986), Sandström (1993), and Declerck (1997). It also engages with previous work specifically on the competition between finite and non-finite adjunct clauses, particularly Behrens et al. (2012), and attempts to reconcile *jegda*-clauses and participle constructions in Early Slavic within the same discourse representation analysis drawing from SDRT accounts of the rhetorical relations triggered by various temporal connectives.

The analysis is split across this chapter and the next. This chapter provides a descriptive and quantitative analysis of *jegda*-clauses as they appear in the Early Slavic corpus with a focus on variables that may capture differences in their properties compared to dative absolutes and conjunct participles so as to motivate their distribution. The next chapter analyses the range of temporal readings triggered by *jegda*-clauses based on different attested combinations of tense-aspect between the *jegda*-clause and its matrix clause and provides a discourse-representation-based analysis of *jegda*- and participle clauses.

Similarly to the previous two chapters, the corpus analysis in this chapter looks at position of *jegda*-clauses in the sentence and the distribution of tense-aspect in different configurations (Section 3.2.1); the properties of subjects, including their information-structural properties (e.g. anaphoric distance to their antecedent) where possible (i.e. in deeply annotated treebanks) (Section 3.2.2); the level of lexical variation of verbs in *jegda*-clauses compared to participle constructions in different configurations; (Section 3.2.3). Finally, it also touches on the average clause length (in number of words) of *jegda*-clauses as opposed to participle clauses, as this has been brought forward in previous literature, as we will see, as a potential functionally motivated difference between finite and nonfinite competitors (Section 3.2.4).

3.2 *Jegda*-clauses in the corpus

3.2.1 Position in the sentence and tense-aspect distribution

Tables 3.2 and 3.3 show the position of *jegda*-clauses relative to the matrix clause, in the Codex Marianus and in standard treebanks, respectively. In both subcorpora, *jegda*-clauses to the left of the matrix clause are significantly more frequent than those to the right ($p < 0.01$, binomial test, one-tailed).²

²The *jegda*-matrix configuration in all the following frequencies also includes 17 occurrences of *bysto*-clauses where, syntactically, the *jegda*-clause follows the main verb (*bysto*). For the same reasons pointed out in Chapter 1, these have been included among *jegda*-clauses to the left of the matrix.

<i>jegda</i> -matrix	matrix- <i>jegda</i>
84.9% (129)	15.1% (23)

Table 3.2: Relative order of *jegda*-clause and matrix clause in the Codex Marianus

<i>jegda</i> -matrix	matrix- <i>jegda</i>	NA
62.7% (158)	29% (73)	8.3% (21)

Table 3.3: Relative order of *jegda*-clause and matrix clause in standard treebanks

Some occurrences do not have an identifiable matrix clause (*NA* in Table 3.3) and, upon closer inspection, these appear largely in contexts such as (16), namely as part of preambles preceding the narrative proper. In such contexts, *jegda*-clauses are used to either specify the time and location in which the narrative takes place or to summarize it.

- (16) *EGDA POS_vLA SAOUL_v I S_vCHRANI*
 when send.AOR.3.SG Saul.SG.M.NOM and watch.AOR.3.SG
DOM_v EGO OUBITI I
 house.SG.M.ACC 3.SG.M.GEN kill.INF 3.sg.m.acc

‘When Saul had sent men to watch David’s house in order to kill him’ (*Psalterim Sinaiticum*, Psalm 58)

As Tables 3.4 and 3.5 show, present-tense *jegda*-clauses are overall the most frequent across the corpus, with aorist *jegda*-clauses not far behind. Imperfect and future *jegda*-clauses (the latter all forms of *byti*, i.e. *b_{gd}-*) are also found with some frequency.

	present	aorist	imperfect	future
<i>jegda</i> -matrix	40.3% (52)	39.5% (51)	11.6% (15)	8.5% (11)
matrix- <i>jegda</i>	56.5% (13)	34.8% (8)	8.7% (2)	0% (0)
Tot.	42.8% (65)	38.8% (59)	11.2% (17)	7.2% (11)

Table 3.4: *Jegda*-clauses in the Codex Marianus: tense distribution by position relative to the matrix clause (row percentage)

	present	aorist	imperfect	future
<i>jegda</i> -matrix	39.9% (63)	36.7% (58)	20.9% (33)	2.5% (4)
matrix- <i>jegda</i>	54.8% (40)	27.4% (20)	16.4% (12)	1.4% (1)
NA	4.8% (1)	57.1% (12)	38.1% (8)	0% (0)
Tot.	41.3% (104)	35.7% (90)	21% (53)	2% (5)

Table 3.5: *Jegda*-clauses in standard treebanks: tense distribution by position relative to the matrix clause (row percentage)

It is hard to make a direct comparison with the distribution of aspects among participle clauses. On the one hand, we can confidently establish the aspect of past-tense forms

on the basis of inflectional morphology alone, with the aorist indicating perfective aspect and the imperfect indicating imperfective aspect. Establishing the aspect of nonpast forms is a more complex matter, since the modern Slavic derivational or ‘lexical’ aspect (i.e. aspectual-pair) system was still developing in Old Church Slavonic (cf. Eckhoff & Janda 2014). Table 3.6 reports the frequency of perfectives and imperfectives among *jegda*-clauses in the Codex Marianus after establishing the aspect of present-tense forms manually on a case-by-case basis. As a general rule, if a present-tense form allows a future interpretation (i.e. it is seen as a completed whole), then it was considered perfective (which is Dostál’s (1954) main diagnostic for present forms); otherwise, it was considered imperfective. All aorist forms were instead as perfective, all imperfect forms as imperfective, and all future forms of *byti* (*bǫd-*) as imperfective (cf. Kamphuis 2020: 36).³

	imperfectives	perfectives
<i>jegda</i> -matrix	48.5% (63)	51.5% (67)
matrix- <i>jegda</i>	18.2% (4)	81.8% (18)
Tot.	44.1% (67)	55.9% (85)

Table 3.6: *Jegda*-clauses in the Codex Marianus: aspect distribution by position relative to the matrix clause (row percentage)

Based on the manually-tagged aspect, perfectives seem slightly more frequent than imperfectives in both pre-matrix and post-matrix position. A binomial test suggests, however, that the difference in proportion is not statistically significant overall ($p = 0.08$, one-tailed), nor in pre-matrix position ($p = 0.39$, one-tailed). In the post-matrix position, we get a significant value ($p < 0.01$, one-tailed), although we should mind that occurrences are relatively few in that position.

3.2.2 Properties of subjects

As shown in Table 3.7, in the Codex Marianus *jegda*-clauses with an overt subject are slightly more frequent than those with a null subject, among both pre- and post-matrix occurrences. However, the difference in proportion is not clear-cut and is not, in fact, statistically significant ($p = 0.08$, two-tailed binomial test). Outside of the Codex Marianus, as shown in Table 3.8, *jegda*-clauses with a null subject are instead slightly more frequent, but once again, the difference between overt- and null-subject *jegda*-clauses is not statistically significant ($p = 0.11$, two-tailed binomial test).

Thanks to the anaphoric-link annotation on the Greek New Testament in the PROIEL treebanks, from nominal referents to their antecedents (including null arguments), we can infer whether the subject of a *jegda*-clause in the Codex Marianus has the same referent as the subject of its matrix clause. Table 3.9 shows the number of coreferential versus non-coreferential *jegda*-clauses by position relative to the matrix.

³The dataset with the added ‘aspect’ variable can be found in the project repository (<https://doi.org/10.6084/m9.figshare.24166254>).

	Overt subject	Null subject
<i>jegda</i> -matrix	54.3% (70)	45.7% (59)
matrix- <i>jegda</i>	73.9% (17)	26.1% (6)
tot.	57.2% (87)	42.8% (65)

Table 3.7: Overt and null subjects in *jegda*-clauses in the Codex Marianus (row percentage)

	overt subject	null subject
<i>jegda</i> -matrix	44.3% (70)	55.7% (88)
matrix- <i>jegda</i>	41.1% (30)	58.9% (43)
NA	61.9% (13)	38.1% (8)
tot.	44.8% (113)	55.2% (139)

Table 3.8: Overt and null subjects in *jegda*-clauses in standard treebanks (row percentage)

	coreferential	non-coreferential
<i>jegda</i> -matrix	37.2% (48)	62.8% (81)
matrix- <i>jegda</i>	30.4% (7)	69.6% (16)
Tot.	36.8% (56)	63.2% (96)

Table 3.9: Subject coreference and non-coreference between *jegda*-clauses and their matrix clause in the Codex Marianus

Jegda-clauses with a different subject from the one of the matrix are significantly more frequent than co-referential *jegda*-clauses ($p < 0.01$, two-tailed binomial test). Intuitively, it is non-coreferential *jegda*-clauses that should show the clearest functional overlap with dative absolutes, since, in the Codex Marianus, the vast majority of absolutes have a different subject from that of the matrix clause. Conversely, coreferential *jegda*-clauses should be especially useful for understanding their competition with conjunct participles, since the latter predominantly have the same subject as the matrix one.

As we did in Chapter 1 for dative absolutes and conjunct participles, we can look at the average distance of the immediate antecedent of the subject of *jegda*-clauses and compare it to the results obtained on the subjects of dative absolutes and conjunct participles. Significant differences between average distances may point to functional differences between the constructions. In Section 1.3.1 of Chapter 1, for example, we argued that the greater average distance of the antecedent of subjects in VS conjunct participles compared to SV conjunct participles may be indicative of the framing function of the former (whereby older referents are reinstated as part of a FRAME). In Chapter 1, we only compared sentence-initial dative absolutes and conjunct participles, since that is where the clearest functional overlap occurs between the two and because that is the position in which conjunct participles may head an overt subject. This time, however, even though over 70% of pre-matrix *jegda*-clauses (93) are sentence-initial, if we were to further subset these into two subgroups of *jegda*-clauses (coreferential

and non-coreferential, to compare them with conjunct participles and dative absolutes, respectively), the number of sentence-initial occurrences in each subset would be too few to allow for a comparison with sentence-initial participle constructions. We shall therefore consider *pre-matrix*, rather than strictly sentence-initial, *jegda*-clauses, dative absolutes, and conjunct participles. From conjunct participles, only the *leftmost* occurrence is considered (which, most of the time, but not always, coincides with the sentence-initial position), since, as we have seen, that is where they may syntactically head a subject. Figure 3.1a compares pre-matrix dative absolutes and pre-matrix *jegda*-clauses whose subject is *not* coreferential with the one of the matrix. Figure 3.1b compares (the leftmost) pre-matrix conjunct participles and pre-matrix *jegda*-clauses whose subject *is* coreferential with the one of the matrix.

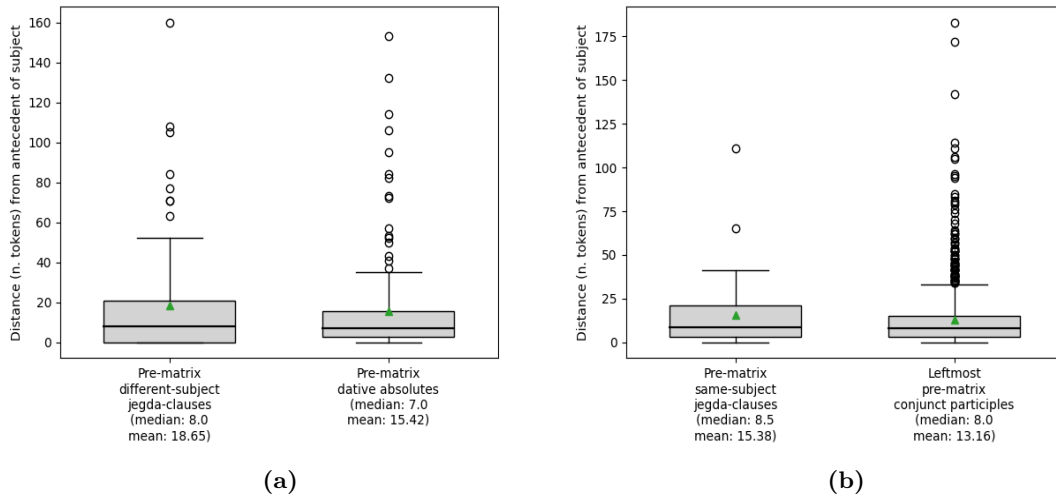


Figure 3.1: (a) Average anaphoric distance between the subject and immediate antecedent in pre-matrix non-coreferential *jegda*-clauses and pre-matrix dative absolutes. Test results: Welch’s t -test: 0.82, p -value = 0.41; one-tailed Mann–Whitney U -test: 6006.5, p -value = 0.78. (b) Average anaphoric distance between the subject and immediate antecedent in pre-matrix coreferential *jegda*-clauses and (the leftmost) pre-matrix conjunct participles. Test results: Welch’s t -test: 0.74, p -value = 0.45; one-tailed Mann–Whitney U -test: 21547.5, p -value = 0.40.

As the test results show, no significant differences emerge from such comparison, namely, the average distance of the antecedent of the subjects of *jegda*-clauses in pre-matrix position is similar to that of the subjects of pre-matrix dative absolutes and conjunct participles regardless of subject coreference. Note that, even when we compare the three constructions in sentence-initial position (regardless of subject co-reference), we still do not observe significant differences.

Some potentially interesting observations can instead be made on the pick-up *rates* of the subject referents in the preceding discourse. In Section 1.3.1 of Chapter 1, we compared the average number of mentions (‘relative saliency’) of the subject referents of sentence-initial conjunct participles and dative absolutes in the preceding 1, 5, 30, and 60 sentences, to gauge whether there are differences in how many times the subject

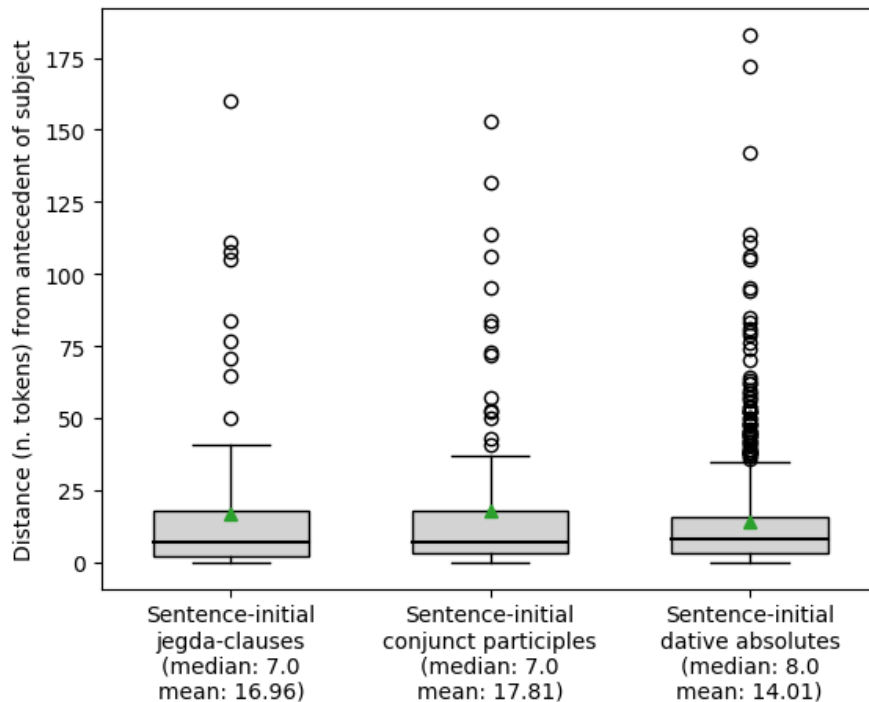


Figure 3.2: Average anaphoric distance between subject and immediate antecedent in sentence-initial *jegda*-clauses, conjunct participles and dative absolutes (all subjects). Test results: 1. *jegda*-clauses versus dative absolutes: Welch’s *t*-test: -0.71 , p -value = 0.47 ; one-tailed Mann–Whitney *U*-test: 2366.5 , p -value = 0.22 ; 2. *jegda*-clauses versus conjunct participles: Welch’s *t*-test: 0.42 , p -value = 0.67 ; one-tailed Mann–Whitney *U*-test: 9768.5 , p -value = 0.89 .

referents of participle constructions in sentence-initial position are mentioned in the near or distant preceding discourse. We observed that the subjects of sentence-initial participle constructions show overall very similar relative saliency. Similarly, Table 3.10 compares the average relative saliency of subject referents in pre-matrix dative absolutes and pre-matrix non-coreferential *jegda*-clauses. Table 3.11 compares the average relative saliency of (the leftmost) pre-matrix conjunct participles and pre-matrix *jegda*-clauses sharing the subject with the matrix clause. For each window of preceding sentences (w), the mean (\bar{x}), median (med) and standard deviation (σ) are provided, together with the results of the statistical tests (Welch’s *t*-test and one-tailed Mann–Whitney *U*-test) comparing the constructions at the relevant window.

As the test results indicate, there are no significant differences between the pick-up rates of the subject referents of pre-matrix coreferential *jegda*-clauses and (the leftmost) pre-matrix conjunct participles, at any window of preceding sentences. Instead, when we compare pre-matrix non-coreferential *jegda*-clauses and pre-matrix dative absolutes, we observe that the referents of the subjects of dative absolutes are overall significantly more activated in the previous discourse than the referents of the subjects of *jegda*-clauses,

	<i>w</i>	non-coref. <i>jegda</i>			absolutes			tests			
		\bar{x}	<i>med</i>	σ	\bar{x}	<i>med</i>	σ	<i>t</i>	<i>p</i>	<i>U</i>	<i>p</i>
<i>all subj.</i>	1	1.08	1.0	0.98	1.39	1.0	1.09	-2.21	0.03	4912.5	0.03
	5	2.52	1.0	2.96	3.24	3.0	2.99	-1.75	0.08	4983.5	<i>0.05</i>
	30	8.6	1.0	11.8	9.27	5.0	10.38	-0.42	0.67	5284.5	0.2
	60	12.08	1.0	18.53	13.36	5.0	17.79	-0.51	0.61	5175.5	0.15
<i>overt</i>	1	0.9	1.0	0.88	1.42	1.0	1.09	-3.32	$<\alpha$	2501.5	$<\alpha$
	5	2.35	1.0	3.14	3.34	3.0	2.99	-1.9	<i>0.06</i>	2612.0	0.01
	30	6.56	1.0	10.48	9.56	6.0	10.44	-1.72	0.09	2571.5	0.01
	60	7.85	1.0	13.13	13.81	6.0	17.94	-2.46	0.02	2490.5	0.01

Table 3.10: Pick-up rates of subject referents in the previous discourse: pre-matrix non-coreferential *jegda*-clauses and pre-matrix dative absolutes, by subject realization and window of preceding sentences. *w* = window of preceding sentences; *non-coref.* = non-coreferential; *all subj.* = null and overt subjects; *overt* = overt subjects only; \bar{x} = mean; *med* = median; σ = standard deviation; *t* = Welch’s *t*-test statistics; *U* = two-tailed Mann-Whitney *U*-test; *p* = *p*-value relative to the statistics to the left; $<\alpha$ = < 0.01 , i.e., highly significant; bold = significant; italics = marginally significant. The *t*-test statistics reflect the order of the column, so if it is negative, *jegda*-clauses have a lower average value than absolutes; if positive, they have a higher average value.

	<i>w</i>	coref. <i>jegda</i>			conjuncts			tests			
		\bar{x}	<i>med</i>	σ	\bar{x}	<i>med</i>	σ	<i>t</i>	<i>p</i>	<i>U</i>	<i>p</i>
<i>all subj.</i>	1	1.77	1.0	1.61	1.51	1.0	1.21	1.09	0.28	21730.5	0.53
	5	4.17	3.5	3.39	3.5	3.0	3.08	1.33	0.19	23042.5	0.18
	30	13.67	7.0	15.78	10.77	6.0	11.54	1.25	0.22	22451.0	0.31
	60	20.08	7.0	26.87	15.81	6.0	19.97	1.08	0.28	22256.5	0.37
<i>overt</i>	1	1.05	1.0	0.65	1.7	1.0	1.14	-2.1	0.04	4737.0	0.39
	5	3.0	2.5	2.65	3.73	3.0	2.99	-0.56	0.58	5171.0	0.86
	30	10.95	4.0	13.99	10.62	6.0	11.01	0.02	0.98	5204.0	0.9
	60	15.09	4.0	21.11	15.6	6.0	19.55	-0.19	0.85	5198.5	0.89

Table 3.11: Pick-up rates of subject referents in the previous discourse: pre-matrix coreferential *jegda*-clauses and (leftmost) pre-matrix conjunct participles, by subject realization and window of preceding sentences; *non-coref.* = non-coreferential; *all subj.* = null and overt subjects; *overt* = overt subjects only; \bar{x} = mean; *med* = median; σ = standard deviation; *t* = Welch’s *t*-test statistics; *U* = two-tailed Mann-Whitney *U*-test; *p* = *p*-value relative to the statistics to the left; $<\alpha$ = < 0.01 , i.e., highly significant; bold = significant; italics = marginally significant. The *t*-test statistics reflect the order of the column, so if it is negative, *jegda*-clauses have a lower average value than conjunct participles; if positive, they have a higher average value.

with some differences depending on whether we consider all subjects (null *and* overt) or overt subjects only. When we consider all subjects, we observe a significant difference in the number of mentions only if the immediately preceding sentence is considered and a marginally significant difference if the 5 preceding sentences are considered. Namely, when we consider the nearby preceding discourse, the subjects of pre-matrix dative absolutes are referentially more activated than the subjects of pre-matrix *jegda*-clauses,

regardless of realization (i.e. this is true of both null- and overt-subject absolutes and *jegda*-clauses). When longer stretches of discourse are considered, there are instead no significant differences in their level of referential activation. When we only consider overt subjects, then the difference in relative saliency between the subjects of the two constructions is highly significant in both the preceding nearby discourse (1, 5 preceding sentences) and the preceding more distant discourse (30, 60 preceding sentences), with the subject referents of dative absolutes showing significantly more activation.

Recall that *jegda*-clauses equally have null and overt subjects, whereas dative absolutes in the Gospels virtually always have overt subjects. Null subjects are generally much more likely to signal topic continuation from the preceding discourse and are, therefore, generally the most likely candidate for topics in the sentence in which they appear (cf. Frascarelli 2007; Rizzi 2018; Di Domenico, Baroncini & Capotorti 2020). If we cross the results on relative saliency with those on the average distance of subject antecedents, it is tempting to think that, when choosing a framing construction involving a change of subject from the FRAME itself to the matrix clause (as is most often the case with dative absolutes in the Codex Marianus and with non-coreferential *jegda*-clauses), then, *with very salient subjects jegda*-clauses may be preferred if there is a topic continuation from the previous discourse, dative absolutes if there is a topic shift. This is potentially suggested by the equally high relative saliencies obtained from larger sentence windows and both null and overt subjects. However, *jegda*-clauses may be preferred as topic-shifter when the FRAME involves *less salient* subjects. This may instead be suggested by the fact that overt-subject non-coreferential *jegda*-clauses are significantly less activated (at any window of preceding sentences) than overt-subject dative absolutes.

The PROIEL treebank does not include annotation of complex information-structural categories such as topic or focus, since these are harder to assign with confidence. It does, however, include several other levels of annotation that can be used to collectively estimate the *topicworthiness* of nominal referents relative to other nominals in the surrounding discourse. Topicworthiness is here meant as ‘an inherent property of an NP, such that it is more likely to serve as the topic (although it need not do so), and is closely related to such concepts as the hierarchies of grammatical person, animacy, definiteness and salience’ (Comrie 2003: 329). We can use a well-tested algorithm⁴ to assign a *topic score* measuring the relative topicworthiness of nominal referents based on properties known to correlate with aboutness topics cross-linguistically (Eckhoff 2018: 34). Summarizing from Eckhoff (2018), these properties and the respective scores for their values are:

- givenness status, namely OLD (+15), NEW (+0), and different types of referent accessibility: ACC-INF (accessible by inference, +10), ACC-GEN (accessible from world knowledge, +5), and ACC-SIT (accessible by deixis, +13);

⁴The algorithm was written by Dag Haug and Hanne Eckhoff and is described in great detail in Eckhoff (2018).

- *relative saliency*, measured as the number of mentions in the 30 preceding sentences (+10 for the candidate in the sentence with the highest relative saliency);
- word order (+15 to the linearly first topic candidate in the sentence, since topics tend to be fronted);
- realization: +30 for null subjects (except for those with a NON-SPEC, +5 for personal pronouns, +5 for human proper nouns, KIND, or NEW givenness status), since pro-drops, personal pronouns, and (human) proper nouns are more likely to be topics;
- syntactic relation: +10 for SUB (since subjects are more likely to be topics), +5 for OBJ, +2 for OBL, +1 for COMP and ADV clauses.
- animacy hierarchy: based on the animacy of referents established at lemma level (see Haug et al. 2009), +10 HUMAN (since nominals with human referents are more likely to be topics), +5 ORG (human, collective), +3 ANIMAL, +3 CONCRETE, +0 TIME, PLACE, NONCONC, VEH.
- the properties of the immediate antecedent and of the intervening competing candidates (+2 for the candidate whose immediate antecedent outranks intervening candidates on syntactic relation, and animacy and givenness hierarchies).

The subjects of coreferential pre-matrix *jegda*-clauses and those of pre-matrix conjunct participles have overall very similar average topic scores, whether we consider all subjects (Figure 3.3a) or overt subject only (Figure 3.3b).

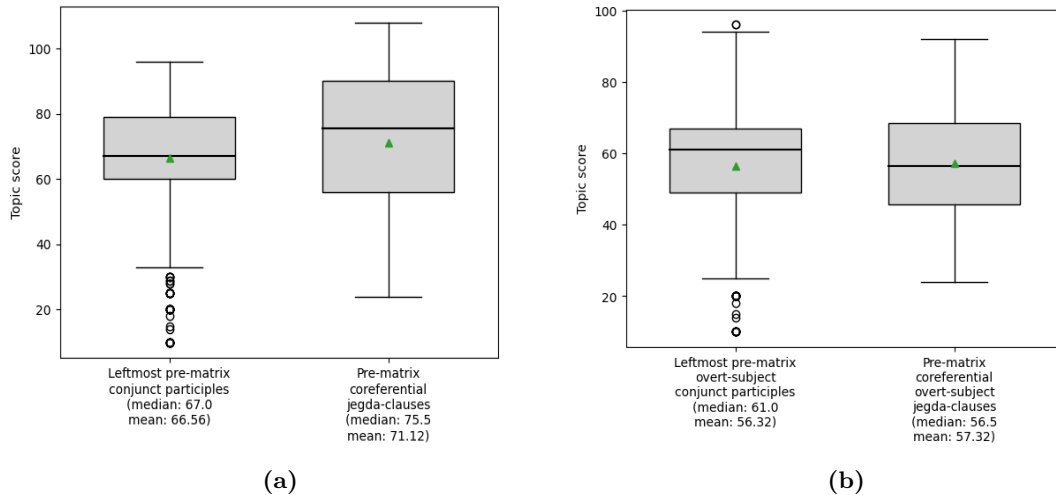


Figure 3.3: Average topic score of subjects in (the leftmost) pre-matrix conjunct participles and pre-matrix coreferential *jegda*-clauses, by subject realization. **(a.)** Null and overt subjects: Welch’s *t*-test: 0.90, $p = 0.37$; one-tailed Mann–Whitney *U*-test: 75994, $p = 0.10$. **(b.)** Overt subjects only: Welch’s *t*-test: -0.10, $p = 0.91$; one-tailed Mann–Whitney *U*-test: 25643, $p = 0.42$.

The subjects of non-coreferential pre-matrix *jegda*-clauses and pre-matrix dative absolutes show instead some differences depending on subject realization. Figures 3.4a and 3.4b

compare the average topic scores of subjects in non-coreferential pre-matrix *jegda*-clauses and pre-matrix dative absolutes. Figure 3.4a includes the scores for both overt and null subjects, whereas Figure 3.4b includes those for overt subjects only.

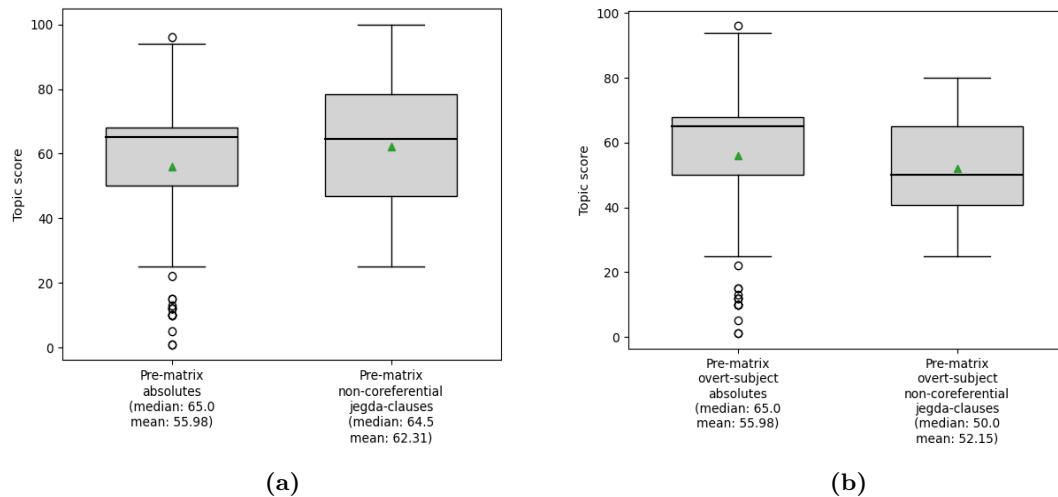


Figure 3.4: Average topic score of subjects in pre-matrix dative absolutes and pre-matrix non-coreferential *jegda*-clauses, by subject realization. **(a.)** Null and overt subjects: Welch’s t -test: -2.33, $p = 0.01$; one-tailed Mann–Whitney U -test: 4563, $p = 0.06$. **(b.)** Overt subjects only: Welch’s t -test: 1.41, $p = 0.08$; one-tailed Mann–Whitney U -test: 3874, $p < 0.01$.

On average, when considering both null and overt subjects, the subjects of *jegda*-clauses have a higher topic score than the subjects of dative absolutes. According to the Mann–Whitney U -test, the difference is not significant or only marginally significant, with $p = 0.06$, whereas according to Welch’s t -test the difference is highly significant at $p = 0.01$. This is as expected, since, as we have seen, *jegda*-clauses equally have null and overt subjects, whereas dative absolutes most often have an overt subject. Null subjects are much more likely candidates as aboutness topics in the sentence in which they appear and we generally mainly expect relatively topicworthy nominal referents to be able to be prodropped in the first place.

Instead, when we only consider overt subjects, the subjects of dative absolutes have an overall greater topic score than the subjects of *jegda*-clauses. In this case, according to Welch’s t -test the difference is not significant at $p = 0.08$, whereas according to the Mann–Whitney U -test, the difference is highly significant at $p < 0.01$. These figures seem to support what we observed about the relative saliency of the subject referents in the two constructions. Namely, given a change in subject from a framing construction to the matrix clause (i.e. the need to use a non-coreferential adjunct), *jegda*-clauses may preferred when there is a topic continuation from the previous discourse, which may license subject prodropping, or when there is a topic shift from the previous discourse but involving less salient or less topicworthy subjects. Dative absolutes may instead overall be preferred when there is a topic shift from the previous discourse, especially if more salient or more topicworthy subjects are involved.

3.2.3 Aktionsart and lexical variation

As seen in Chapter 1, a large portion of the Greek New Testament in PROIEL includes annotation on the Aktionsart of verbs, including the categories ACTIVITY, SEMELFACTIVE, STATE, and TELIC. In Section 1.4 of Chapter 1, we saw that telic verbs are the most common among both conjunct participles and dative absolutes, although more prominently so among the former, whereas activity verbs are markedly more common among dative absolutes. Activity verbs are intuitively easily found as part of framing constructions (e.g. *while they were singing, when you pray*), whereas telic verbs can be expected to be particularly frequent in series of events with narrative progression (as is the case with INDEPENDENT RHEMES).

As Table 3.12 shows, telic verbs are also the most frequent among *jegda*-clauses. Stative and activity verbs are relatively more frequent among *jegda*-clauses than conjunct participles (where they represented 14.82% and 3.58% of the verbs, respectively). This makes the Aktionsarten of verbs among *jegda*-clauses overall more similar to dative absolutes, even though activity verbs remain markedly more frequent among the latter (20.47% of all dative absolutes). The overall similarity between absolutes and *jegda*-clauses in this respect is not surprising, since both activity and stative verbs can be expected to occur in framing constructions relatively often.

		matrix				<i>tot.</i>
		ACT	SEM	ST	TEL	
<i>jegda</i>	ACT	1.38 (2)	0	2.1 (3)	6.9% (10)	10.34% (15)
	SEM	0	0	0	0	0
	ST	2.1 (3)	0	8.27% (12)	9% (13)	19.31% (28)
	TEL	3.4% (5)	0	11.72% (17)	55.17% (80)	70.34% (102)

Table 3.12: Aktionsarten of *jegda*-clauses and their matrix clause in the Codex Marianus. ACT = activity, SEM = semelfactive, ST = state, TEL = telic.

The analysis of the lexical variation among participles in Section 2.2.3 of Chapter 2 also provided additional evidence regarding the different functions of conjunct participles and dative absolutes. As in Chapter 2, Table 3.13 shows the percentage of verbs belonging to the ten most frequent lemmas (10MFL), as well as the MATTR with a moving window of 40 lemmas, for each construction in different subcorpora and configurations. The values for absolutes and conjunct participles are repeated from Chapter 2 for ease of comparison. Figures 3.5a and 3.5b compare the average 10MFL and MATTR from Table 3.13 for each construction.

When adding *jegda*-clauses to the picture, it appears that *jegda*-clauses, overall, occupy an intermediate position between conjunct participles and dative absolutes. Namely, lexical variation among *jegda*-clauses is generally greater than or equal to dative absolutes but smaller than variation among conjunct participles. As the statistics in Figure 3.5

Subsample	Absolutes		Conjuncts		Jegda	
	10MFL	MATTR	10MFL	MATTR	10MFL	MATTR
(O)CS + (O)ES + Mar. (non-N)	36.15%	0.65	20.51%	0.78	35.15%	0.69
(O)CS + (O)ES + Mar. (N)	38.53%	0.65	24.29%	0.78	38.12%	0.69
(O)CS + (O)ES (non-N)	36.05%	0.65	19.67%	0.79	33.33%	0.73
(O)CS + (O)ES (N)	38.46%	0.65	23.20%	0.79	35.71%	0.72
(O)CS + Mar.	42.74%	0.64	25.12%	0.77	41.38%	0.65
(O)CS	43.31%	0.64	23.32%	0.80	40.72%	0.67
(O)ES	37.93%	0.66	23.49%	0.78	37.65%	0.78
Mar.	50.54%	0.61	36.79%	0.69	50%	0.60
(O)CS + (O)ES + Mar. (left, N)	39.83%	0.65	28.72%	0.74	40.62%	0.68
(O)CS + (O)ES + Mar. (right, N)	40.71%	0.69	22.90%	0.80	48.42%	0.69
(O)CS + (O)ES (left, N)	39.36%	0.66	27.22%	0.76	36.08%	0.72
(O)CS + (O)ES (right, N)	41.31%	0.70	22.85%	0.80	49.32%	0.68
(O)CS (left)	44.22%	0.64	27.61%	0.77	43.96%	0.68
(O)CS (right)	45.78%	0.71	19.98%	0.84	50.85%	0.66
(O)ES (left)	39.47%	0.67	27.20%	0.75	40.30%	0.77
(O)ES (right)	46.92%	0.67	25.17%	0.77	NA	NA
Mar. (left)	51.46%	0.60	44.12%	0.63	53.08%	0.60
Mar. (right)	NA	NA	30.11%	0.78	NA	NA

Table 3.13: Lexical variation among the verbs in *jegda*-clauses and participle constructions, by subcorpus and configuration. *10MFL* = ten most frequent lemmas; *MATTR* = moving-average type-token ratio; *Mar* = Codex Marianus; (O)ES = Old East Slavic and Middle Russian; *left* and *right* = pre-matrix and post matrix, respectively; *N* = normalized. Note that the closer the *MATTR* is to 1 (and the lower the percentage for the 10MFL), the greater the lexical variation.

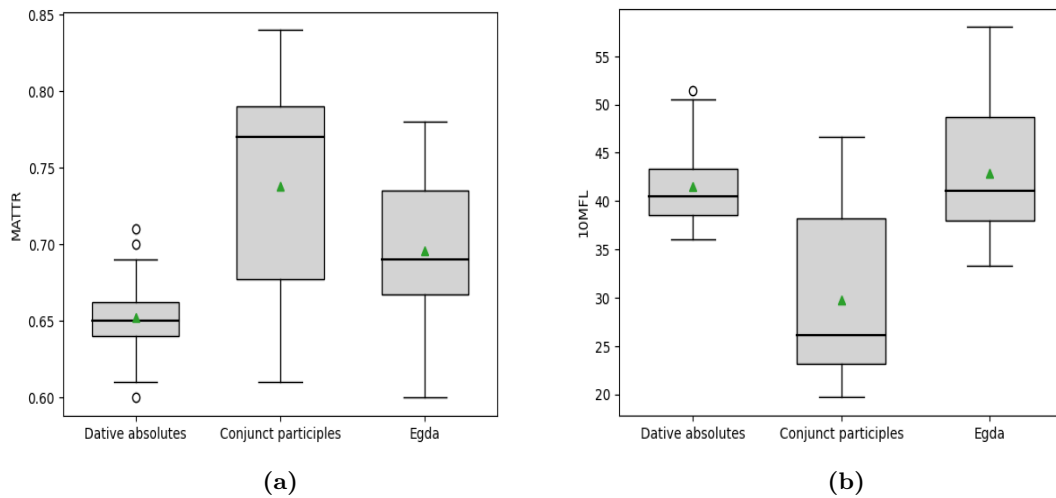


Figure 3.5: Lexical variation among verbs in *jegda*-clauses, conjunct participles and dative absolutes. **(a.)** average moving-average type-token ratio (MATTR). Test results: 1. *jegda*-clauses versus dative absolutes: Welch’s t -test: -3.16 , $p < 0.01$; one-tailed Mann–Whitney U -test: 93 , $p < 0.01$; 2. *jegda*-clauses versus conjunct participles: Welch’s t -test: 2.1 , $p = 0.04$; one-tailed Mann–Whitney U -test: 283 , $p = 0.02$. **(b.)**: average percentage of subjects belonging to the ten most frequent lemmas (10MFL). Test results: 1. *jegda*-clauses versus dative absolutes: Welch’s t -test: -0.8 , $p = 0.4$; one-tailed Mann–Whitney U -test: 186 , $p = 0.7$; 2. *jegda*-clauses versus conjunct participles: Welch’s t -test: -5.1 , $p < 0.01$; one-tailed Mann–Whitney U -test: 72 , $p < 0.01$.

indicate, we get a significant difference between conjunct participles and both *jegda*-clauses and dative absolutes, according to both the MATTR and the 10MFL scores. This is as expected, since the most typical functions of conjunct participles are as INDEPENDENT RHEMES and as ELABORATIONS, both of which normally encode new information, as opposed to absolutes and *jegda*-clauses, which are more typically framing constructions and as such refer to previously mentioned or inferrable events, resulting in overall more ‘predictable’ predications.

Variation among dative absolutes is only significantly smaller than *jegda*-clauses when considering the MATTR but not the 10MFL value. Judging from the types of variation the two metrics tend to capture, this suggests that, similarly to absolutes, many *jegda*-clauses are accounted for by the 10 most frequent lemmas, that is, just like absolutes, *jegda*-clauses are used very often with the same few verbs. However, outside of those 10 lemmas, *jegda*-clauses show significantly more variation than absolutes, as the difference in MATTR indicates. This suggests that *jegda*-clauses, while being overall functionally very similar to dative absolutes, may also function as less ‘predictable’ frames than absolutes, and also points to the nearly-formulaic character of absolutes, which often occur in what appears to be fixed or semi-fixed verb-subject combinations.

3.2.4 Clause length

Behrens et al. (2012: 182) suggest that finite adverbial clauses, having explicit tense and a subordinating conjunction, tend to make a ‘less context-dependent, more precise contribution’ to the interpretation of the sentence in which they occur than their non-finite competitors, ‘highlighting one single aspect of the relation between the events or propositions involved, or narrowing down the set of possible interpretations’ (cf. also Sæbø 2011a). According to Behrens et al., non-finite clauses may therefore be chosen merely because of their semantic underspecifity and economy from the production perspective, since they typically distinguish fewer morphosyntactic categories than finite competitors. In the case of Old Church Slavonic and Greek, participles make aspect, but not tense, distinctions. They are, however, inflected for gender and case, unlike finite verb forms. Behrens et al. (2012) also claim that non-finite adjuncts are less explicit in the semantic dimensions relevant to finite adverbial clauses because the former are shorter than the latter, since they lack a subjunction and ‘have no explicit subject’. This is obviously referring specifically to *open* adjuncts (in our case, conjunct participles, but not absolute constructions), which share their subject with the matrix clause, but they also claim that the observations they make on the basis of open adjuncts are overall valid for *closed* adjuncts, too (i.e. absolutes) (Behrens et al. 2012: 197–198). We can test the claims on the length of participle adjuncts and *jegda*-clauses by leveraging dependency annotation to count the number of tokens occurring inside participle clauses and *jegda*-clauses. Haug (2012: 314) already compared the length of Ancient Greek participle clauses in different configurations (e.g. SV, VS, pre-matrix, post-matrix) to test the intuition that sentence-initial participles, particularly VS ones, should be shorter than others since they are more likely to function as FRAMES, which introduce known or presupposed events used to anchor other events and are therefore expected to be more ‘predictable’. His results confirmed that this was, in fact, the case.

We can reproduce the same figures as those in Haug (2012) on Early Slavic, while also adding *jegda*-clauses, to test whether Behrens et al.’s (2012) claim on the length of non-finite adjuncts and finite adverbials applies in our case. Mean, median, and standard deviation of the length of clauses in different configurations are provided in Table 3.14. The clear-cut differences between participle configurations observed by Haug (2012) on Ancient Greek are also found in the Early Slavic data to a large extent. Haug (2012: 315) reports that sentence-initial VS participle constructions (‘ptcp_sub’ in the table) and post-matrix participles as a whole (‘ptcp_right’) stand out as significantly shorter and longer than the average, respectively, which fits the observation that FRAMES are presupposed or known eventualities, hence presumably ‘more predictable’ and thus shorter on average, whereas ELABORATIONS convey new information and are thus generally expected to involve longer clauses. As Table 3.14 shows, both absolutes and conjunct participles

	mean	median	standard deviation
sub_ptcp	3.65	3.0	3.17
ptcp_sub	3.39	3.0	2.66
sub_da	4.62	4.0	3.24
da_sub	4.67	4.0	3.31
sub_xadv	4.13	3.0	3.93
xadv_sub	3.17	2.5	2.53
da_left	4.2	4.0	2.53
da_right	4.87	4.0	4.14
xadv_left	3.23	3.0	2.5
xadv_right	4.63	3.0	4.4
jegda_left	5.22	4.0	3.0
jegda_right	5.6	5.0	2.94

Table 3.14: Mean, median and standard deviation of the length (in number of tokens) of participle constructions and *jegda*-clauses in different configurations. *sub_ptcp* = sentence-initial SV participle constructions (absolutes and conjunct participles), *ptcp_sub* = sentence-initial VS participle constructions, *sub_da* = sentence-initial SV absolutes, *da_sub* = sentence-initial VS absolutes, *sub_xadv* = sentence-initial SV conjunct participles, *xadv_sub* = sentence-initial VS conjunct participles, *da_left* = pre-matrix absolutes, *da_right* = post-matrix absolutes, *xadv_left* = pre-matrix conjunct participles, *xadv_right* = post-matrix conjunct participles, *jegda_left* = pre-matrix *jegda*-clauses, *jegda_right* = post-matrix *jegda*-clauses.

are shorter, on average, in pre-matrix than in post-matrix position.⁵ Sentence-initial SV conjunct participles are also significantly longer than VS conjunct participles.⁶ Among sentence-initial dative absolutes, the difference in length between those in the SV and VS configuration is instead not statistically significant.⁷ The average length of *jegda*-clauses, in both pre- and post-matrix position, although particularly in the post-matrix one (*'jegda_right'* in the table), is greater than participle constructions in any configurations. This suggests that, although the lack of an explicit subject or a subjunction does not exclude relatively long clauses, it does affect the average length. The length of absolute constructions, in fact, also stands out, overall, compared to that of conjunct participles. Once again, the presence of an overt, internal subject in absolute constructions in the vast majority of cases is likely to have an impact on the average length.

3.2.5 Summary

The analysis of the corpus data in this section has pointed to some potential functional differences between *jegda*-clauses and participle constructions.

Jegda-clauses appear to be distributionally more similar to dative absolutes, which

⁵Pre-matrix absolutes versus post-matrix absolutes: Welch's *t*-test: -2.34, $p=0.01$; one-tailed Mann-Whitney *U*-test: 127016, $p = 0.10$. Pre-matrix conjunct participles versus post-matrix conjunct participles: Welch's *t*-test: -11.64, $p<0.01$; one-tailed Mann-Whitney *U*-test: 10382926.5, $p<0.01$.

⁶Welch's *t*-test: 9.71, $p<0.01$; one-tailed Mann-Whitney *U*-test: 10449661, $p<0.01$.

⁷Welch's *t*-test: -0.33, $p=0.73$; one-tailed Mann-Whitney *U*-test: 318913.5, $p=0.65$.

reflects the intuition that their typical function is as framing, backgrounding devices, which is also supported by the fact that, similarly to absolutes, *jegda*-clauses are significantly more frequent in pre-matrix than post-matrix position. This is to be expected from frames since they set the scene for the discourse to come and are thus more likely to occur sentence-initially.

We saw that establishing the aspect of *jegda*-clauses is not as straightforward as with participle constructions due to frequent present-tense forms, which cannot always be confidently disambiguated as perfective or imperfective on the basis of morphology alone since the modern Slavic derivational-aspect system had not fully developed in Old Church Slavonic. Although this makes a direct comparison of aspect distribution among participles and *jegda*-clauses using quantitative methods quite challenging, manual annotation of a subset of our dataset with information on aspect showed no significant difference in the number of perfectives and imperfectives among *jegda*-clauses.

The analysis of some of the properties of subjects among *jegda*-clauses indicated that there is overall more evident competition with absolutes than with conjunct participles. In the Gospels, *jegda*-clauses with a different subject than their matrix were found to be significantly more frequent than those with co-referential subjects. However, null subjects are not infrequent even among *jegda*-clauses with a different subject from their matrix, suggesting that *jegda*-clauses may be preferred when there is a topic continuation from the immediately preceding discourse. This also seems to be confirmed, quite unsurprisingly, when we compare the average topic scores of subject referents in non-coreferential *jegda*-clauses and dative absolutes, the former being significantly higher than the latter. When we only consider overt subjects, however, it becomes clear that this result is chiefly obtained *because* of null-subject *jegda*-clauses. With overt subjects, which we can take to indicate potential topic-shifting (though, of course, not necessarily), their difference in topic score is not significant, and the subjects of dative absolutes are found to have referents that are significantly more activated in the previous sentence than *jegda*-clauses. This may point to the frequent topic-shifting function of dative absolutes observed in Chapter 2, whereby a highly salient referent in the immediately preceding discourse, possibly part of a rhemic constituent, is repurposed as the topic of a new portion of discourse.

Finally, the analysis of lexical variation among the verbs indicates that *jegda*-clauses may be somewhat less predictable frames than absolute constructions, as suggested by the significantly greater MATTR among the former than the latter. However, when using the ten most frequent lemmas as the criterion, we observe no significant difference in lexical variation among the two, regardless of subcorpus or configuration, which suggests that, similarly to absolutes, *jegda*-clauses involve few lemmas occurring very often, but unlike the latter, they also involve several lemmas occurring very infrequently. Compared to conjunct participles, however, *jegda*-clauses show significantly smaller lexical variation

among verbs on both fronts, as was also expected given the more typical INDEPENDENT RHEME or ELABORATION function of conjunct participles, which are thus more likely to introduce new, less predictable information.

Chapter 4

Temporal relations and discourse representation

The previous chapter provided a descriptive and quantitative analysis of *jegda*-clauses as they appear in the Early Slavic corpus and focussed on variables that could capture differences in their properties compared to dative absolutes and conjunct participles, as they had emerged from the analyses in Chapters 1 and 2. The analysis of their distribution in the sentence and of the properties of their subjects suggested that *jegda*-clauses are, overall, functionally more similar to dative absolutes, with significant differences in the information status of their typical subjects. In particular, the results on subject realization and the analysis of the relative saliency and topicworthiness of their subject referents suggested that *jegda*-clauses may be preferred over dative absolutes when there is a topic continuation from the immediately preceding discourse or when there is a topic shift from the previous discourse but involving less salient or less topicworthy subjects. The analysis of the lexical variation among verbs indicated that *jegda*-clauses may also occur as much less ‘predictable’ FRAMES than dative absolutes, which instead seem to be formed from a much more limited pool of lemmas and not infrequently appear in fixed or formulaic verb-subject combinations.

This chapter discusses the temporal interpretation of *jegda*-clauses based on the different attested combinations of tense-aspect between the *jegda*-clause and its matrix clause (Section 4.1). It also proposes a possible formalization of the rhetorical relations typically introduced by *jegda*-clauses and participle constructions within an SDRT framework (Section 4.2).

4.1 *Jegda*-clauses and temporal relations interpretation

As shown in Table 4.1, most aorist *jegda*-clauses have an aorist matrix clause and most present-tense *jegda*-clauses have a present-tense matrix clause. Imperfect *jegda*-clauses can occur with both an aorist matrix or an imperfect matrix clause. This is

unlike Ancient Greek, where an imperfect *hote/hotan*-clause with an aorist matrix is instead exceedingly rare.¹

matrix egda	aorist	future	imperfect	present	NA
aorist	26.5% (102)	-	7% (29)	1.3% (6)	2.4% (12)
future	-	0.7% (3)	-	3.9% (13)	-
imperfect	5.4% (22)	-	8.5% (37)	0.6% (3)	1.6% (8)
present	1.6% (8)	2.4% (9)	-	37.5% (150)	0.5% (2)

Table 4.1: Tense of *jegda*-clauses and their matrix clause in the Codex Marianus and standard treebanks (combined)

In the past tense, an aorist *jegda*-clause followed by an aorist main clause normally produces narrative progression, whereby the matrix event is interpreted as occurring after the *jegda*-event. (1) and (2) are two such examples.

- (1) *i egda porogaše se emu. sɔvlěše sɔ*
 and when mock.AOR.3.SG REFL 3.SG.M.DAT take off.AOR.3.PL from
nego chlamidɔ i oblěše i vɔ
 3.SG.M.GEN robe.SG.F.ACC and dress.AOR.3.PL 3.SG.M.ACC in
rizy svoje.
 clothes.PL.F.ACC own.3.PL.ACC
 ‘After they had mocked him, they took off the robe and put his own clothes on him’ (Matthew 27:31)

- (2) *Egda že slyša pilatɔ se slovo.*
 when PTC hear.AOR.3.SG Pilate.NOM this.SG.N.ACC word.SG.N.ACC
pače ouboě se
 more.COMP be frightened.AOR.3.SG REFL
 ‘When Pilate heard this, he was even more afraid’ (John 19:8–9)

According to Partee (1984), a pre-posed eventive *when*-clause introduces a new temporal referent locating the main eventuality *just after* the event described by the *when*-clause. This is the temporal effect we see in (1) and (2), and it is similar to that created by perfective FRAMES and INDEPENDENT RHEMES. Hinrichs (1986: 75–76) instead claimed that two eventive clauses (accomplishments and achievements, that is, *telic* expressions) linked by *when* in English can have any temporal order with respect to each other and brings (3a)-(3c) as evidence of this.

- (3) a. John broke his arm when he wrecked the Pinto.
 b. When the Smiths moved in, they threw a party.
 c. When the Smiths threw a party, they invited all their old friends.

¹As observed in the Introduction, only 1 occurrence of pre-matrix *hôte*-clause with an imperfect verb followed by an aorist main verb was found (in Herodotus), and none of *hotan*.

where (3a), according to Hinrichs, conveys simultaneity, whereas in (3b) and (3c) the event described by the *when*-clause precedes and follows the one described by the main clause, respectively. The argument is that when both the *when*-clause and the main clause are non-durative, then they are *both* contained within the reference time, or *frame* (Hinrichs 1986: 78), introduced by *when*. Inside this reference time, the two events can have any order, as represented in Figure 4.1 from Hinrichs (1986: 75).

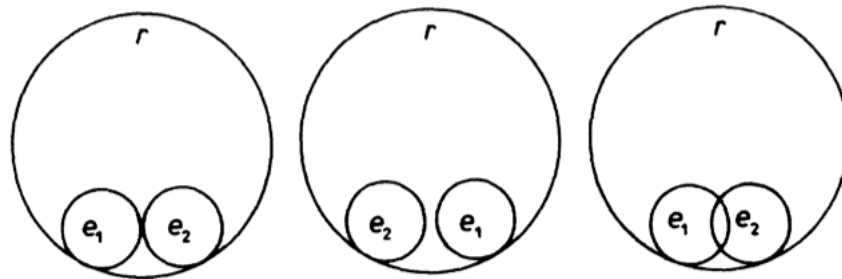


Figure 4.1: Possible orders of a non-durative *when*-clause and its non-durative matrix clause according to Hinrichs (1986: 78)

There are some clear shortcomings in Hinrichs' account—apart from relying only on native speakers' intuitions (cf. Hinrichs 1986: 74). First, as also noted by Sandström (1993), he seemingly considers the *when*-clause and the main clause equal, rather than the first being subordinated to the second, as is clear from Hinrichs disregarding the order of the clauses in (3a), which he compares to (3b)-(3c), even if the linear order has obvious importance in the organization of discourse, which, in turn, has clear effects on the range of possible readings. Another weakness of Hinrichs's account is that, despite working within a coarse stative versus eventive binary ontology as in DRT, it seems to make an equivalence between lack of progressive form on telic predicates in English and eventivity. Arguably, the *when*-clause in (3a), while being a telic, bounded expression, is pragmatically not the same as the one in (3b), for example, which is also telic and bounded. Even DRT and SDRT, which strictly work within that same coarse eventive-stative binary, recognize that while in languages such as French and English the distinction is *generally* marked by tense in narrative texts, it can result from a combination of several factors, including, crucially, the larger discourse context, so that finer-grained distinctions are often needed to distinguish the two (cf. Kamp & Reyle 1993; de Swart 1995, 1998; Caudal 2005). A more satisfactory account, in some respects, of the possible temporal interpretation of *when*-clauses is the one by Declerck (1997), who gives default, 'unmarked' interpretations based on the combination of durativity, telicity, and boundedness, but always leaves room for pragmatic incompatibility, where *generally* another default interpretation is given. A recurring 'exception' to the default interpretation of different combinations of *when*-clause and main clause is what Declerck calls 'occasion'-referring *when*-clauses, which he describes as those not introducing a

time referent, but defining an occasion in which the main eventuality actualizes, as in examples (4)-(6), from Declerck (1997: 218-226).

- (4) When John built his sailboat, he bought some timber from me
- (5) The chairman resigned when the secretary wrote a libellous article about him in the local newspaper
- (6) When he found the lost ring, he also found some secret letters

In such cases, according to Declerck, the order in which the two eventualities occur can remain unspecified. In a way, this is very similar to what is claimed by Hinrichs (1986), except that *inclusion* of one unbounded eventuality into another (the *when*-situation into the main situation or vice versa) does not appear to be an option in his account, as Figure 4.1 indicates. I argue that what Declerck refers to as *occasions* can often be loosely represented as durative, but bounded, even when the *when*-clause in isolation is punctual, in which case a non-durative (punctual) matrix can be included in the *when*-situation. This is arguably the case in (6), as well as (3a), as already discussed, but also (3c). Example (3c) is also discussed by Sandström (1993), among others, to show how temporal inclusion of an eventive main clause into an eventive *when*-clause is not only possible, but it involves a part-of relation, whereby the main clause event is a subevent, or a constitutive part (Moens 1987), of the *when*-clause event (but not vice-versa). Note that the part-of relation between an eventive when clause and an eventive main clause is meant by Sandström (1993) in terms of Boolean inclusion \subseteq , whereby identity relations are also considered as part-of relations while keeping the order of *when*-clause and main clause relevant. For example, (7a), from Sandström (1993: 197), is obviously not the same as (7b).

- (7) a. When Sue killed her husband, she put poison in his whisky.
 b. When Sue put poison in her husband's whiskey, she killed him.²

In (7a), or [when e_1 , e_2], the subevent relation $e_1 \subseteq e_2$ is in fact identity, whereby the poisoning constitutes the killing. Note, however, that the *when*-clause and the main clause are not equal, so it *does* matter what goes into what. If we switch the order in which the eventualities are presented, we do not get a \subseteq relation, but a consequentiality one, temporally $e_1 \prec e_2$.

The subevent relation between an eventive *jegda*-clause and an eventive main clause can be exemplified by (8).

²Sandström (1993) gives the example *When Sue put poison in her husband's whisky, he died*, but (7b) is properly parallel to (7a) and gives a clearer impression of the unidirectionality of the inclusion relation.

- (8) a. *Bystǫ že egda kr̃stišę se vsi*
 happen.AOR.3.SG PTC when baptize.AOR.3.PL REFL all.PL.M.NOM
ljudbe. isou kr̃štǫšju se i
 people.PL.M.NOM Jesus.DAT baptize.PTCP.PFV.M.SG.DAT REFL and
moleštju se. otvrze se nebo
 pray.PTCP.IPFV.M.SG.DAT REFL open.AOR.3.SG REFL heaven.SG.N.NOM
- b. *Egeneto de en tōi baptisthēnai*
 happen.AOR.3.SG PTC in the.DAT.SG wash ceremonially.INF.AOR.PAS
hapanta ton laon kai Iēsou
 all.SG.M.ACC the.SG.M.ACC people.SG.M.ACC even Jesus.GEN
baptisthentos kai
 wash ceremonially.PTCP.PAS.PFV.M.GEN.SG and
proseukhomenou aneōikhthēnai ton
 pray.PTCP.IPFV.M.GEN.SG open.INF.PAS.AOR the.SG.M.ACC
ouranon
 heaven.SG.M.ACC
 ‘It came to pass that, when all the people were baptized, as Jesus was
 baptized and as he was praying, the heaven was opened’ (Luke 3:21)

Incidentally, we can also observe that in (8b) Greek does not have a finite *when*-clause equivalent (e.g. *hōte* or *hōtan*), but a nominalized accusative with infinitive, as is often the case in *egeneto*-constructions, which suggests that this usage of *jegda*, in common with English *when*, is found in Old Church Slavonic regardless of Greek.

When an aorist *jegda*-clause is followed by an imperfect matrix verb, the main eventuality may receive an inchoative reading (9) or a reading where the event described by the *jegda*-clause could be interpreted as temporally included in the main eventuality (10).

- (9) a. *i egda ṽnide ṽ dom̃ ot̃ naroda.*
 and when enter.AOR.3.SG in house.SG.M.ACC from crowd.SG.M.GEN
ṽprašaach̃ i učenic̃ ego o
 ask.IMPF.3.PL 3.SG.M.ACC disciple.PL.M.NOM 3.SG.M.GEN about
pr̃it̃č̃i.
 parable.SG.F.ACC
- b. *Kai hote eisēlthen eis ton oikon apo*
 and when enter.AOR.3.SG in the.SG.M.ACC house.SG.M.ACC from
tou okhlou epērōtōn auton
 the.SG.M.GEN crowd.SG.M.GEN ask.IMPF.3.PL 3.SG.M.ACC
hoi mathētai autou tēn
 the.PL.M.NOM disciple.PL.M.NOM 3.SG.M.GEN the.SG.F.ACC
parabolēn
 parable.SG.F.ACC
 ‘When he entered a house away from the crowd, his disciples started asking
 him about the parable.’ (Mark 7:17)

- (10) a. *Bystǔ* *že* *egda približi* *se* *isǔ* *vǔ*
 happen.AOR.3.SG PTC when approach.AOR.3.SG REFL Jesus.NOM in
erichǫ. *slěpecǔ* *eterǔ* *sěděaše* *pri*
 Jericho.ACC blind man.SG.M.NOM certain.SG.M.NOM sit.IMPV.3.SG by
poti *prose.*
 road.SG.M.LOC beg.PTCP.IPVV.M.NOM.SG
- b. *Egeneto* *de* *en tōi* *eggizein* *auton*
 happen.AOR.3.SG PTC in the.SG.M.DAT approach.INF.PRS 3.SG.M.ACC
eis Hiereikhō taphlos *tis* *ekathēto*
 in Jericho blind man.SG.M.NOM certain.SG.M.NOM sit.IMPV.3.SG
para tēn *hodon* *epaitōn*
 by the.SG.F.ACC road.SG.F.ACC beg .PTCP.IPVV.M.NOM.SG
 ‘It came to pass that, when he approached Jericho, a certain blind man was
 sitting by the road begging.’ (Luke 18:35)

If we followed Hinrichs’s (1986) theory, the sitting eventuality should include the approaching event in (10). According to Hinrichs, the temporal relation between an eventive *when*-clause and a stative main clause is the same as the one between a stative *when*-clause and an eventive main clause, namely one of inclusion of the eventive situation in the stative one. If, instead, we followed Partee’s (1984) account, the sitting eventuality should occur *just after* the approaching event. It seems, however, that the temporal extension of the main clause in (10) is irrelevant in this context, which merely appears to report that a *sitting* state holds within the time referent introduced by the *jegda*-clause, which introduces a bounded referent. This is also in line with Sandström (1993: 192–195), who argues that in such cases it is clear how a *when*-clause ‘provides a temporal referent, which the head clause says something about’ and that like a ‘temporal adverbial in general [it] serves to narrow down the section of the timeline in which the temporal location of the event is to be placed’—which is how framing adverbials and time locating adverbs are also generally described. As framing adverbials, sentence-initial *when*-clauses provide the common temporal frame for both the *when*-situation and the main clause situation, restricting the interpretation of the main eventuality to the specific domain introduced by the *when*-clause itself (cf. Charolles 2005). As observed by Declerck (1997) on English *when*-clauses, when a durative-unbounded main clause follows a punctual *when*-clause, the default is to represent the main clause, as it were, as a snapshot of a point in time—the one provided by the *when*-clause and corresponding to the common temporal frame. This is what we get in (10). If the default reading is pragmatically inaccessible, then we get what Declerck calls ‘sloppy W-simultaneity’, namely the succession of one eventuality immediately after the other. With a durative-unbounded main clause, the effect of sloppy simultaneity is that the main eventuality is interpreted as either left- or right-bounded, namely as inchoative or terminative. The coincidence between the punctual framing adverbial and the durative-unbounded main clause is, in this case, with the starting or

ending point of the main clause. Sandström (1993) also gives the inchoative reading as a possible interpretation of the temporal relation between an eventive *when*-clause and a stative *main*-clause, in which case, however, she also posits a consequentiality relation, whereby the stative main clause describes *reactions* to the *when*-clause event, which introduces a ‘change as a result of which the state predicate holds’ (Sandström 1993: 194–195). The consequentiality relation is certainly possible in (9), where the disciples start asking Jesus questions as a result of now being in the house, where he was not before the *jegda*-event.

The inchoative effect is also found with a perfective participle followed by an imperfect matrix, as shown by the dative absolute in (11) and the conjunct participle in (12).

- (11) a. *i v̄šed̄šju emu v̄ dom̄.*
 and enter.PTCP.PFV.M.DAT.SG 3.SG.M.DAT in house.SG.M.ACC
učenici ego v̄prašach̄o i
 disciple.PL.M.NOM 3.SG.M.GEN ask.IMPF.3.PL 3.SG.M.ACC
edinogo. ěko
 alone.SG.M.GEN that
- b. *kai eiselthontos autou eis oikon*
 and enter.PTCP.PFV.M.GEN.SG 3.SG.M.GEN in house.SG.M.ACC
hoi mathētai autou kat’ idian
 the.PL.M.NOM disciple.PL.M.NOM 3.SG.M.GEN in private.SG.F.ACC
epērōtōn auton
 ask.IMPF.3.PL 3.SG.M.ACC
 ‘And when He had come into the house, His disciples started to ask Him privately’ (Mark 9:28)
- (12) a. *Prīed̄š že is v̄ stran̄o*
 arrive.PTCP.PFV.M.NOM.SG PTC Jesus.NOM in region.SG.F.ACC
kesarij̄e filipovy. v̄prašaše učeniky
 Caesarea.SG.F.GEN Philippi.SG.F.GEN ask.IMPF.3.SG disciple.PL.M.ACC
svoj̄e gl̄e
 own.3.PL.ACC say.PTCP.IPFV.M.NOM.SG
- b. *Elthōn de ho Iēsous eis ta*
 come.PTCP.PFV.M.NOM.SG PTC the.NOM Jesus.NOM in the.PL.N.ACC
merē Kaisareias tēs Filippou
 region.PL.N.ACC Caesarea.SG.F.GEN the.SG.F.GEN Philippi.SG.M.GEN
ērōta tous mathētas autou
 ask.IMPF.3.SG the.PL.M.ACC disciple.PL.M.ACC 3.SG.M.GEN
legōn
 say.PTCP.IPFV.M.NOM.SG
 ‘When Jesus arrived to the region of Caesarea Philippi, he began asking his disciples’ (Matthew 16:13)

- (16) a. *i egda sěaše ovo pade pri*
 and when sow.IMP.3.SG some.SG.N.NOM fall.AOR.3.SG along
pōti i popōrano bystō. i
 way.M.LOC.SG and trample.PTCP.PAS.PFV.N.NOM.SG be.AOR.3.SG and
ptice nebskyje pozobaše e
 bird.F.NOM.PL sky.F.NOM.PL devour.AOR.3.PL 3.SG.N.ACC
- b. *kai en tōi speirein auton ho men*
 and in the.N.DAT.SG sow seed.INF.PRS 3.SG.M.ACC 3.SG.N.NOM PTC
epesen para tēn hodon kai
 fall.AOR.3.SG along the.F.ACC.SG path.F.ACC.SG and
katapatēthē kai ta peteina tou
 trample.AOR.PAS.3.SG and the.N.NOM.PL bird.N.NOM.PL the.M.GEN.SG
ouranou kataphagen auto
 air.M.GEN.SG devour.AOR.3.SG 3.SG.N.ACC
- 'And as he sowed, some fell by the wayside; and it was trampled down, and the birds of the air devoured it.' (Luke 8:5)

These three examples are to be analysed as a stative time-locating adverbial followed by an eventive main clause. Following Hinrichs's (1986) treatment, the reference time introduced by the *jegda*-clause is a subinterval of the state it describes.

When both the *jegda*-clause and the matrix clause are in the imperfect, the two situations are generally interpreted as either temporally equivalent, as in (17), or as habitual, repeated occurrences, as in (18).

- (17) a. *egda iděaše narodi ougnětaachō i*
 when go.IMP.3.SG crowd.M.NOM.PL throng.IMP.3.PL 3.SG.M.ACC
- b. *en de tōi hupagein auton hoi*
 in PTC the.M.DAT.SG go away.INF.PRS 3.SG.M.ACC the.M.NOM.PL
okhloi sunepnigon auton
 crowd.M.NOM.PL choke.IMP.3.PL 3.SG.M.ACC
- 'As he [Jesus] went, the crowds thronged him' (Luke 8:42)
- (18) a. *i egda viděachō i dsi*
 and when see.IMP.3.PL 3.SG.M.ACC spirit.M.NOM.PL
nečistii pripadaachō kō nemu. i
 unclean.M.NOM.PL fall down.IMP.3.PL to 3.SG.M.DAT and
všpiěchō glšte ěko
 cry out.IMP.3.PL say.PTCP.IPFV.NOM.PL that
- b. *kai ta pneumata ta akatharta*
 and the.N.NOM.PL spirit.N.NOM.PL the.N.NOM.PL unclean.N.NOM.PL
hotan auton etheōroun proseipton
 when 3.SG.M.ACC see.IMP.3.PL fall down before.IMP.3.PL

autōi kai ekrazon legontes hoti
 3.SG.M.DAT and cry out.IMPF.3.PL say.PTCP.IPFV.NOM.PL that
 ‘Whenever they saw him, the unclean spirits fell down before Him and cried out, saying’ (Mark 3:11)

The range of temporal relations of nonpast-tense *jegda*-clauses to their matrix clause (always nonpast-tense as well) is similar to the one we observe with past-tense clauses. Once again, what determines the interpretation can be boiled down to whether the *jegda*-clause and the matrix clause present an eventuality as durative or non-durative, and as bounded or unbounded. As already discussed, establishing the aspect of nonpast forms is not always straightforward. It is often possible for perfective nonpast indicatives in Old Church Slavonic to receive a future interpretation (cf. Dostál 1954; Eckhoff & Janda 2014; Eckhoff & Haug 2015; Kamphuis 2020), but that is not always clearly the case (Eckhoff & Janda 2014: 253). In (19), for example, the verb *sъzrěti* ‘to ripen’, arguably perfective, does not receive a future interpretation.

- (19) a. *egda že sozъrěatъ plodъ. abie posъletъ*
 when PTC ripen.PRS.3.SG grain.M.NOM.SG immediately send.PRS.3.SG
srъpъ. ěko nastoitъ žetva
 sickle.M.ACC.SG because be present.PRS.3.SG harvest.F.NOM.SG
- b. *hotan de paradoi ho karpos*
 when PTC hand over.SBJV.PFV.3.SG the.M.NOM.SG grain.SG.M.NOM
euthus apostellei to drepanon hoti
 immediately send away.PRS.3.SG the.N.ACC.SG sickle.N.ACC.SG that
parestēken ho therismos
 be present.PRF.3.SG the.M.NOM.SG harvest.M.NOM.SG
 ‘But when the grain ripens, immediately he puts in the sickle, because the harvest has come’ (Mark 4:29)

It does, however, induce narrative progression, similarly to what occurs between an eventive *jegda*-clause in the aorist followed by an aorist main clause. The *jegda*-clause in this example is habitual-repetitive, conveying a similar meaning to English *whenever*. As Declerck (1997: 232) argues, the basic temporal relation between the situation described by a *whenever*-clause and a main clause is interpreted in the same way as the non-habitual-repetitive counterpart.

Similarly to the temporal relation between an aorist *jegda*-clause and an imperfect matrix, a stative matrix clause following a perfective nonpast *jegda*-clause may receive an inchoative reading, as (20) exemplifies.

- (20) a. *a chъ egda pridetъ niktože ne*
 but Christ.NOM when come.PRS.3.SG nobody.NOM NEG
věstъ otъ kōdou bōdetъ.
 know.PRS.3.SG whence be.FUT.3.SG

- b. *ho de Xristos hotan erkhētai*
 the.M.NOM.SG PTC Christ.M.NOM.SG when come.SBJV.PRS.3.SG
oudeis ginōskei pothen estin
 nobody.NOM know.PRS.3.SG whence be.PRS.3.SG
 ‘When Christ comes, no one will know where he is from’ (John 7:27)

The state of (not) *knowing* starts *just after* the *when*-clause event. In other words, the temporal referent for the main clause, introduced by the *jegda*-clause, is the (left-bounded) interval corresponding to the consequent state of Christ *coming*.

Finally, the temporal relation between a durative nonpast *jegda*-clause followed by a stative main clause is one of overlap. In (21), the state described by the main verb *sot̃* holds at least as long as the *guarding* eventuality holds.

- (21) a. *egda krěp̃zky ouorož̃b sę*
 when strong man.M.NOM.SG arm.PTCP.PFV.M.SG.NOM REFL
chranit̃ svoi dvor̃. ṽ mirě
 guard.PRS.3.SG his.M.ACC.SG house.M.ACC.SG in peace.M.LOC.SG
sot̃ iměniě ego
 be.PRS.3.PL possession.N.NOM.PL 3.SG.M.GEN
- b. *hótan ho iskhuros*
 when the.M.NOM.SG strong man.M.NOM.SG
kathōplismenos phulassēi tēn
 arm.PTCP.PAS.PRF.M.NOM.SG guard.SBJV.PRS.3.SG.ACT the.F.ACC.SG
heautou aulēn en eirēnēi estin
 3.SG.M.GEN house.F.ACC.SG in peace.F.DAT.SG be.PRS.3.SG
ta huparkhonta autou
 the.N.NOM.PL possession.N.NOM.PL 3.SG.M.GEN
 ‘When a strong man, fully armed, guards his own house, his possessions are safe’ (Luke 11:21)

4.1.1 Summary

This section presented an overview of the main temporal relations holding between *jegda*-clauses and their matrix clause, on the basis of the different attested combinations of tense-aspect in each of them. We have observed that the range of readings inferred for English *when*-clauses, as gathered from some of the major studies on the subject (e.g. Partee 1984; Hinrichs 1986; Sandström 1993; Declerck 1997), can generally be observed to hold in Early Slavic *jegda*-clauses as well. In particular, we have observed that accounts of temporal relations taking into account not only boundedness *and* durativity of eventualities, but the larger discourse context and pragmatic incompatibility (e.g. Moens 1987; Sandström 1993; Declerck 1997), seem more suitable to explain a wider range of readings beyond the default interpretations based on tense-aspect only (e.g. Partee 1984;

Hinrichs 1986). This is the case, for instance, of what Declerck (1997) calls *occasion-referring when*-clauses and, similarly, what Sandström (1993) refers to as *subevent* relation (the ‘constitutive part’ relation of Moens 1987). This type of temporal relation can be observed in Early Slavic between an aorist *jegda*-clause followed by an *aorist* matrix clause, which can *also* (perhaps typically, but certainly not exclusively) invite a reading whereby the matrix event immediately follows the *jegda*-clause event (i.e. Partee’s (1984) and Dowty’s (1986) ‘time just after’).

We have also seen how imperfect *jegda*-clauses *do* occur relatively often with an aorist matrix clause, unlike Greek imperfect *hóte-/hótan*-clauses, for which an aorist matrix is exceedingly uncommon. Such configuration in Early Slavic normally involves a reading whereby the matrix event is included in the *jegda*-clause (stative) eventuality, similarly to the reading obtained from an imperfective participle followed by an aorist matrix clause.

4.2 Rhetorical relations and the discourse structure of *jegda*- and participle clauses

4.2.1 Competing motivations, competing relations

Behrens et al.’s (2012) analysis of competing structure, while mentioning finite subordinate clauses as one of the possible competitors to non-finite adjuncts, focuses on the semantic-pragmatic parameters affecting the choice between non-finite adjunction and syntactic coordination.³ Their main observation is that fronted adjuncts typically trigger an Accompanying Circumstance⁴ or SDRT’s *Background* relation to the matrix clause, often with a (weakly) causal meaning (Behrens et al. 2012: 207) and that their prototypical uses are generally incompatible with SDRT coordinating discourse relations such as *Narration*, which imply temporal succession of events, whereas syntactic coordination is instead taken to signal discourse coordination.⁵ They also make a difference in ‘discourse potential’ (i.e. presumably the set of compatible discourse relations) between VP coordination, where the conjuncts share the subject, and clause coordination, where they do not, as in (22a), from Blakemore & Carston (2005: 570) and (22b), from Levinson (2000: 121), respectively. They suggest that VP coordination, given the closer tie between the first and the second conjunct, is more likely to establish a causal background for the

³The authors, unfortunately, largely dismiss the competition between finite subordinates and non-finite adjuncts to focus on the competition of the latter with coordinated and juxtaposed structure, since, according to them, finite subordinate ‘differ from non-finite adjuncts mainly by expressing a more precise semantic relation to the host’ and are therefore ‘less interesting competitors than are the non-subordinate alternatives, namely coordination and juxtaposition’ (Behrens et al. 2012: 201).

⁴Accompanying circumstance (AC), as a clause-linking relation, is informally defined as relating two events or states to each other, which, “from the point of view of the speaker/writer, form a unit mentally” (Kortmann 1991: 122)’ (Behrens & Fabricius-Hansen 2009). Note that, as the authors also point out, *Accompanying circumstance* does not correspond to any SDRT relation in particular, while being compatible with several of them (e.g. Background, Elaboration, Result).

⁵From an SDRT standpoint, the equivalence between syntactic coordination and discourse coordination is not as straightforward: see for example Asher & Vieu (2005: 598-599) on this issue.

next event in the sequence, encoded by the second conjunct, which makes VP coordination closer to adjunction than clausal coordination.

- (22) a. She jumped on the horse *and* rode into the sunset.
 b. John turned the key *and* the engine started.

They also add that some languages may have VP coordination as a felicitous alternative to clause-final non-finite adjuncts that describe an unbounded state or activity and that introduce an *Accompanying Circumstance* to the matrix, in which case they ‘contribute to the main storyline on a par with their host’ (Behrens et al. 2012: 216). However, we know that INDEPENDENT RHEMES are discourse-structurally very similar to independent clauses and, in many cases, a *Background* or *Accompanying Circumstance* reading is unambiguously not available, as in (23) (repeated from Chapter 1) and (24).

- (23) a. *i abie tekъ edinъ otъ*
 and immediately run.PTCP.PFV.M.NOM.SG one.M.NOM.SG from
nichъ. i priemъ gobъ.
 he.GEN.PL and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG
isplъnъ ocъta. i vъznezъ
 fill.PTCP.PFV.M.NOM.SG vinegar.GEN.SG and put.PTCP.PFV.M.NOM.SG
na trъstv. napaěše i
 on reed.ACC.SG give.to.drink.IMPF.3.SG he.ACC.SG
- b. *kai eutheōs dramōn heis ex*
 and immediately run.PTCP.PFV.M.NOM.SG one.NOM.SG from
autōn kai labōn spongon
 he.GEN.PL and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG
plēsas te oxous kai
 fill.PTCP.PFV.M.NOM.SG with vinegar.GEN.SG and
peritheis kalamō epotizen auton
 put.PTCP.PFV.M.NOM.SG reed.DAT.SG give.to.drink.IMPF.3.SG he.ACC.SG
 ‘Immediately one of them ran and took a sponge, filled it with sour wine and put it on a reed, and offered it to him to drink’ (Matthew 27:48)

- (24) *i viděvъ ja bžbъstvъnyj*
 and see.PTCP.PFV.M.SG.NOM 3.PL.M.ACC divine.M.NOM.SG
unoša i radъ byvъ
 young man.M.NOM.SG and happy.M.NOM.SG be.PTCP.PFV.M.SG.NOM
tekъ pokloni sja imъ
 run.PTCP.PFV.M.SG.NOM bow.AOR.3.SG REFL 3.PL.M.DAT
 ‘And when he saw them, the divine young man rejoiced, ran up to them and bowed’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 28v)

Indeed, the sentence in (22a) intuitively corresponds to the typical contexts in which INDEPENDENT RHEMES occur, whereas (22b), involving subject switch-reference, may

be realistically expected as a sequence of absolute construction, or *jegda*-clause, and main clause, even if this may trigger somewhat different discourse interpretations than a sequence of main clauses. Compare (22) with (25).

- (25) a. *Jumping on the horse*, she rode into the sunset.
 b. *When John turned the key*, the engine started.

Even from the English non-finite adjunct in (25a), it is clear that the relation between the adjunct and the main clause is not *Background*, since it involves two events, whereas to license *Background*, in SDRT terms, one of the two eventualities should be a state (cf. Asher & Lascarides 2003; Asher et al. 2007): $Event(e_\alpha) \wedge State(e_\beta) \wedge ?(\alpha, \beta) > Background(\alpha, \beta)$, as well as $State(e_\beta) \wedge Event(e_\alpha) \wedge ?(\alpha, \beta) > Background(\alpha, \beta)$.⁶ But it is not *Accompanying Circumstance* either, since the two eventualities are understood as occurring in sequence, with the *jumping* event almost enabling the *riding* one, which goes against the co-temporality requirement of Accompanying Circumstances (Behrens & Fabricius-Hansen 2009). While both (22b) and (25b) invite a causal/enablement reading, in (25b) the consequentiality relation seems stronger. A possible explanation for this is that *when*, as highlighted by Behrens et al. (2012) among the potential differences between coordination and adjunction/subordination, can have a connective function by referring back to the preceding discourse and link a known or inferrable eventuality to the matrix clause, whereas, according to them, a first conjunct in a coordinated construction cannot because ‘it has to introduce a new event referent’ (Behrens et al. 2012: 221). Moreover, besides their ‘connective function’, which is in common with non-finite adjuncts, Behrens et al. (2012: 219) also argue that finite adverbial clauses ‘usually highlight one particular (temporal, causal, concessive) aspect of the full co-eventuality relation at the cost of others, thus narrowing down the interpretation as compared to a non-finite adjunct’. In (25b), the intuition is that the use of a *when*-clause would be maximally coherent if, for instance, the engine started *only* when John turned the key (e.g. after trying other ways) or when John turned the key *this time* (e.g. as opposed to other times), perhaps after fixing something meanwhile, and so on. Behrens et al.’s (2012) observations thus seem useful for explaining some of the differences between coordinated independent clauses and a non-finite adjunct followed by the main clause when the adjunct functions as a FRAME, as well as between participle adjuncts and a *jegda*-clause. However, they are still not helpful for accounting for the choice between coordinated independent clauses and an INDEPENDENT RHEME participle followed by the matrix. In particular, unlike what Behrens et al. (2012) observe regarding non-finite adjuncts, INDEPENDENT RHEMES introduce new events in the discourse, thus inducing narrative progression (cf. Bary & Haug 2011: 13-16), and they are not only *compatible* with *Narration*, but they *typically*

⁶The question mark in the SDRT notation simply indicates that two discourse constituent are linked by a semantically underspecified relation.

license it. In fact, as argued by Bary & Haug (2011), they *always* induce narrative progression, which is instead only the default between two eventive main clauses and is among the rationales for referring to Ancient Greek participles as *grammaticalized* versions of SDRT's rhetorical relations (Bary & Haug 2011: 5).

4.2.2 INDEPENDENT RHEMES as triggers for Strong Narration

As argued by Asher & Lascarides (2003: 20), the coherence of a discourse is scalar, in the sense that its quality can vary from minimally to maximally coherent (or 'relevant', broadly in the now classical sense of Sperber & Wilson 1986) depending on several factors, including how many rhetorical relation hold between two discourse segments (the more relations, the more coherent the interpretation), the number of anaphoric expressions with a resolved antecedent (the more expressions, the greater the quality of the coherence), and the specific rhetorical relation(s) involved in the interpretation. Similarly, Altshuler & Truswell (2022) argue that the acceptability of a discourse with a given assigned structure is a gradient matter, which can be boiled down to two major components, namely the 'gradient quality' of some discourse relations, such as *Parallel* or *Contrast* (cf. Altshuler & Truswell 2022: 245-248), and the 'more slippery notion' of how strongly an inferred discourse structure is supported by world knowledge (ibi: 215-216). Already Asher & Lascarides (2003) had argued that some rhetorical relations are inherently scalar. The 'quality' of a *Narration* relation, for instance, depends on the specificity of the common topic that summarizes the content of the 'story' within which the units linked by *Narration* occur.

Bras, Draoulec & Vieu (2001), for instance, analysed the different ways certain temporal connectives in French, such as *puis* 'then' and *un peu plus tard* 'a bit later', affect the temporal structure of a discourse and license somewhat different versions of the same rhetorical relation *Narration*. *Puis* triggers *Strong Narration*, which has a stronger requirement on the presence of a common topic summarizing the eventuality linked by the relation and signals that no relevant intermediate event e_γ can occur between the two discourse units e_α and e_β connected by *Narration* without the post-state of e_α being incompatible with the pre-state of e_β . *Un peu plus tard* instead licenses *Weak Narration*, which merely signals temporal succession and can link two eventualities with a much weaker common topic. The temporal sequentiality (or *abutment*, Kamp & Reyle 1993: 573) entailed by *Strong Narration* is therefore not present with the *Weak Narration* relation.⁷ Moreover, there are different non-temporal effects typically allowed by *Strong* and *Weak Narration*. Both signal temporal succession of two events, but the former does not seem to allow *Result* to additionally hold between the two, while the latter does; conversely, the former allows some form of *Enablement* relation, whereas the

⁷See also Kehler's (2002) relation *Occasion*, which is largely equivalent to Bras et al.'s (2001) *Strong Narration*, and its discussion in Altshuler & Truswell (2022).

latter does not.⁸

Parallels can be made between the discourse properties of *puis* observed by Bras et al. (2001) and the temporal and non-temporal semantic effects of INDEPENDENT RHEMES. First, the typical usage of INDEPENDENT RHEMES is in contexts where one event immediately follows another, often with the non-temporal effect that the first loosely enables the second. This can occur in relatively long chains of conjunct participles leading up to a finite main verb, inviting a reading in which the events follow one another without interruptions. Second, sequences of conjunct participles are necessarily interpreted as part of the same storyline as the event described by the matrix clauses which they lead up to. INDEPENDENT RHEMES may thus be considered as explicitly signalling that the relationship between the events described is one of *Strong Narration*, namely that the matrix clause event and any event encoded by an INDEPENDENT RHEME syntactically dependent on it are to be interpreted strictly as ‘telling the “same story” (topic)’ (Bras et al. 2001: 8) and as temporally *abutted*. The ‘same story’ requirement (i.e. the presence of a clearly identifiable FT), in Early Slavic (as in Greek) is often fulfilled by an explicit sentence-initial FRAME—i.e. another conjunct participle, dative absolute, or finite temporal subordinate, if not several of these together. (26)-(28) are examples of this discourse structure.

- (26) *i prozvuteru tomu suštju*_[FRAME]. *i*
 and priest.SG.M.DAT that.SG.M.DAT be.PTCP.IPFV.M.SG.DAT and
čvrnorizbcju iskusnu. iže i
 monk.SG.M.DAT skillful.SG.M.DAT which.SG.M.NOM and
*poimъ*_[INDRHEME] *blaženago feodosia.* *i*
 take.PTCP.PFV.M.SG.NOM blessed.SG.M.GEN Feodosij.SG.M.GEN and
po obyčaju svętyichъ ocъ
 according to custom.SG.M.DAT holy.PL.M.GEN father.PL.M.GEN
*ostrigy*_[INDRHEME] *i obleče i*
 tonsure.PTCP.PFV.M.SG.NOM 3.SG.M.ACC clothe.3.SG.AOR.ACT 3.SG.M.ACC
vъ mъnišbkuju odežju
 in monastic.SG.F.ACC clothes.SG.F.ACC

‘As he was a priest himself and a skillful monk, he tonsured the Blessed Theodosius according to the custom of the Holy Fathers and put him in monastic clothes’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 31v)

- (27) *andrěi že poklonšju*_[FRAME] *sę ocju i*
 Andrej.NOM PTC bow.PTCP.PFV.M.SG.DAT REFL father.M.DAT.SG and
*šedъ*_[INDRHEME] *sěde v peresopnici*
 go.PTCP.PFV.M.SG.NOM sit.AOR.3.SG in Peresopnitsa.LOC
 ‘Andrej, after bowing to his father, went and sat in Peresopnitsa’ (*Suzdal Chronicle*, Codex Laurentianus f. 109v)

⁸Note that *Enablement*, as Bras et al. (2001) already pointed out, has not been formalized within SDRT, but it is informally borrowed from Sandström (1993), who posited it as a non-temporal effect holding between a *when*-clause and its matrix in several contexts.

- (28) *i žena edina srošiti_[FRAME] vto točenie*
 and woman.SG.NOM some.SG.NOM be.PTCP.IPFV.F.SG.NOM in flow.LOC
krave lěto dvě na desěte. i mnogo
 blood.GEN year.GEN.PL twelve and much
postradavšii_[FRAME] ot mnogobalii. i
 suffer.PTCP.PFV.F.SG.NOM from much.GEN.PL doctor.GEN.PL and
iždivšii_[FRAME] vse svoe. i ni
 spend.PTCP.PFV.F.SG.NOM all.N.ACC.SG her.N.ACC.SG and NEG
edinoje polbdze obrětsii_[FRAME]. nto pače vto
 one.F.GEN.SG benefit.F.GEN.SG receive.PTCP.PFV.F.SG.NOM but even in
gore prišedšii_[FRAME]. slyšavšii_[FRAME] o
 worse.N.ACC.SG come.PTCP.PFV.F.SG.NOM listen.PTCP.PFV.F.SG.NOM about
isě. prišedšii_[INDRHEME] vto narodě. sto zad
 Jesus.LOC come.PTCP.PFV.F.SG.NOM in crowd.M.LOC.SG from behind
prikosno se rizě ego.
 touch.AOR.3.SG REFL garment.F.LOC.SG 3.SG.M.GEN

‘Now a certain woman had a flow of blood for twelve years, and had suffered many things from many physicians. She had spent all that she had and was no better, but rather grew worse. When she heard about Jesus, she came behind Him in the crowd and touched His garment’ (Mark 5:25-27)

In (26), the independent rhemes *poimto* ‘took’ and *ostrigy* ‘tonsured’ are preceded by the dative absolute *suštju*, functioning as a FRAME. In (27), the INDEPENDENT RHEME *sedto* ‘went’ is also preceded by a dative absolute *polkonšju* ‘after bowing’, which is in fact co-referential with the INDEPENDENT RHEME (and the matrix verb), clearly signalling the different discourse status of the first participle compared to the following one. In (28), we see a much more complex FRAME, made of several conjunct participles, leading up to what seems to be the only conjunct participle in the sentence potentially functioning as an INDEPENDENT RHEME, namely *prišedšii* ‘came’.

4.2.3 Formalizing FRAMES as Background-triggers

All the examples so far involved participle clauses in pre-matrix position, which, as we have repeatedly seen, is where participles are typically either FRAMES or INDEPENDENT RHEMES. In SDRT terms, we have seen how, most typically, INDEPENDENT RHEMES are likely to trigger a *Narration* relation. When preceded by framing constructions, the latter are most likely to license *Background*. The formalization in Asher et al. (2007) of the SDRT relation *Background* in its *Background_{Forward}* version, namely when the backgrounded eventuality occurs linearly before the foregrounded eventuality, in fact, seems to capture the *typical* discourse relations between pre-matrix participle clauses and their matrix particularly well. The schematized structure in Figure (4.2), from Asher et al. (2007: 14), could in principle be applied to all the examples seen so far.

As Figure 4.2 shows, *Background_{Forward}* does not directly attach to the foregrounded constituent, but to a constructed framing topic (FT), π_1 in Figure 4.2, to which the

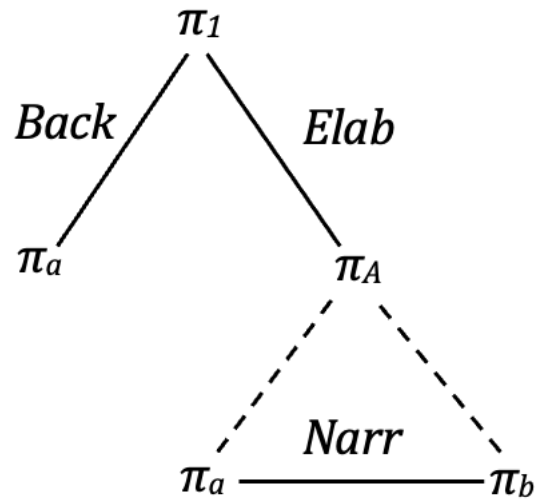


Figure 4.2: Basic discourse structure of a discourse unit involving a *Background_{forward}* relation in SDRT according to Asher et al. (2007: 14)

foreground is then attached via *Elaboration*. The foregrounded constituent can then be complex and contain, for example, several discourse units attached to each other via *Narration*.

The FT is constructed by ‘percolating’ up a discourse referent from the backgrounded constituent (π_a) to make it available for the foreground (π_A). Different factors are considered to affect the potential of a referent to percolate up to the topic, including information structure, the position of the referent in the givenness scale, the syntactic and thematic role of a referent, and, more generally, the salience of the referents (Asher et al. 2007: 17). What is relevant to our discussion here is that the FT required by the INDEPENDENT RHEMES attached to the matrix by (*Strong*) *Narration* is the same as the one which is constructed when a background constituent is available in the discourse structure. When a FRAME is available, then the framing participles (or *jegda*-clause) constitute the backgrounded constituent, whose referents percolate up to construct an FT within which the foregrounded constituent(s) (i.e. typically INDEPENDENT RHEMES, if present, plus the main verb) are located. Basic structures like the one in Figure 4.2 reflect the very common case in which one or several participles can be stacked up before the matrix verb and that only the first is, potentially, a FRAME, as Bary & Haug (2011) had already observed. This also fits the observation that very often the first clauses in such series are dative absolutes or *jegda*-clauses, followed by one or more conjunct participles. Note that the backgrounded constituent can certainly be a complex one, as Declerck (1997: 114, 130-131, i.a.) observed about multiple temporal adverbials, which

are normally understood as including one another and all eventually contribute to the specification of a single time. See, for example, (29)-(31).

- (29) { *i egda bystь*_[FRAME] *dvoju na desete lětou vžsedžšemь*_[FRAME]
 and when be.AOR.3.SG twelve year.GEN go in.PTCP.PFV.DAT.PL
imъ vъ imъ po obyčajo prazdnika
 3.PL.M.DAT in Jerusalem.ACC according to custom.DAT.SG festival.GEN.SG
*i konьčavžšemь*_[FRAME] *dni vžvraštajoštemь*_[FRAME] *se*
 and finish.PTCP.PFV.M.DAT.SG day.M.ACC.PL return.PTCP.IPFV.DAT.PL REFL
imъ} osta otrokъ is vъ imě
 3.PL.M.DAT leave.AOR.3.SG boy.M.NOM.SG Jesus.NOM in Jerusalem.LOC
i ne čju iosifъ i mati ego
 and NEG realize.AOR.3.SG Joseph.NOM and mother.NOM.SG 3.SG.M.GEN
 ‘When he was twelve years old, they went up to the festival according to the custom and, after the festival was over, while they were returning home, the boy Jesus stayed behind in Jerusalem and Mary and Joseph did not realize’ (Luke 2:42–43)

- (30) { *i egda vorotiša*_[FRAME] *se opetъ polovci knezъ*
 and when return.AOR.3.PL REFL again Polovtsians.NOM.PL king.NOM.SG
*že jaropolkъ ukrepiвь*_[FRAME] *se bveju*
 PTC Yaropolk.NOM strengthen.PTCP.PFV.M.SG.NOM REFL God’s.F.INST.SG
*pomoštъju ne žda*_[FRAME] *inoe*
 help.F.INST.SG NEG expect.PTCP.IPFV.M.SG.NOM other.F.GEN.SG
pomošti ni brata ni drugago} tokmo s
 help.F.GEN.SG neither brother.M.GEN.SG nor other.M.GEN.SG only with
perejaslavci svoimi ispostize ja u
 Pereyaslavs.INST his.INST.PL overtake.AOR.3.SG 3.PL.M.ACC at
polkъstěne
 Polkosten.GEN

‘And when the Polovtsians returned again, Prince Yaropolk strengthened himself with God’s help and, expecting no other help, neither from a brother nor from others, only with his Pereyaslavs he overtook them at Polkosten’ (*Suzdal Chronicle*, Codex Laurentianus f. 98r)

- (31) { *egda že priide*_[FRAME] *noštъ drugaja po*
 when PTC come.AOR.3.SG night.F.NOM.SG other.F.NOM.SG after
*pravilě vozlegšu*_[FRAME] *mi i ne*
 Prayer Rules.LOC lay down.PTCP.PFV.M.SG.DAT 1.SG.M.DAT and NEG
*spja*_[FRAME] } *molitvy govorju*
 sleep.PTCP.IPFV.M.SG.NOM prayer.F.ACC.PL say.PRS.1.SG

‘When another night came, as I lay down after my Prayer Rules and I could not sleep, I pray’ (*The Life of Avvakum*, Pustozerskij Sbornik f. 99r)

All these examples have a complex frame, made of multiple adverbial clauses incrementally building a common FT for the foregrounded eventuality. If we break down these

examples and construct a discourse structure (SDRS) within an SDRT framework, we can see how, in fact, the construction of a common FT for the main eventuality occurs with every incoming discourse unit. Each discourse unit adding to the top-level FT (i.e. the one serving as FT for the main eventuality) can, in turn, introduce different discourse relations to the other units that contribute to the main FT. Let us look more closely at one of the examples, (29), to illustrate this. Let (a)-(f) be more literal renditions of each discourse unit in (29):

- (a) *i egda byst̃ d̃voju na des̃ete l̃t̃ou* [And when he was twelve]
- (b) *ṽšed̃š̃em̃ im̃ ṽ im̃ po obyč̃aj̃o prazd̃bñika* [they having gone up to Jerusalem according to the custom of the feast]
- (c) *i koñčaṽš̃em̃ d̃ni* [and having completed the days]
- (d) *ṽzvrašt̃aj̃ošt̃em̃ s̃e im̃* [while they were returning]
- (e) *osta otrok̃ is ṽ im̃* [the boy Jesus remained behind in Jerusalem]
- (f) *i ne č̃ju iosif̃o i mati ego* [and Mary and Joseph did not realize]

All discourse units from (a) to (d) eventually contribute to the FT, which the foregrounded eventuality, also a complex one and including both (e) and (f), elaborate on. A possible SDRS for (29) would look like (4.3). As is conventional in SDRT, π is any discourse unit, elementary (simple) or complex. Subscript lowercase refers to elementary discourse unit (EDU), namely each of the units listed as (a) to (f). Subscript numbers refer to FTs, whereas uppercase refer to complex discourse units, namely discourse-coordinated units (both EDUs and FTs) dominated by the same node.

First, π_a introduces an FT, which the complex unit π_B , including all EDUs from (b) to (d), elaborate on. The EDUs inside π_B are not all on the same level: π_b and π_c are linked via *Narration*, whereas π_d is linked to both π_b and π_c (i.e. π_A) via *Continuation*, which signals that the second constituent (π_d) elaborates on the same topic as the first. This allows π_d to be stative, and π_b and π_c to be eventive. Now, not only (π_1) but also referents from (π_B) can percolate up and become available for constructing an FT for the whole discourse (π_3). The FT (π_2) for the foregrounded eventuality, drawing discourse material from (π_3), is linked to (π_1) via *Continuation*, since (π_2) elaborates on the same topic as (π_1).

We have thus established that complex frames all contribute to the construction of a discourse topic to which foregrounded narrative sequences may then be attached. The structure (4.3) is, of course, one specific scenario among the ways in which a framing topic can be constructed and in which a foreground can attach to a background. In particular, (π_C) can, in principle, be made of a chain of conjunct participles functioning as

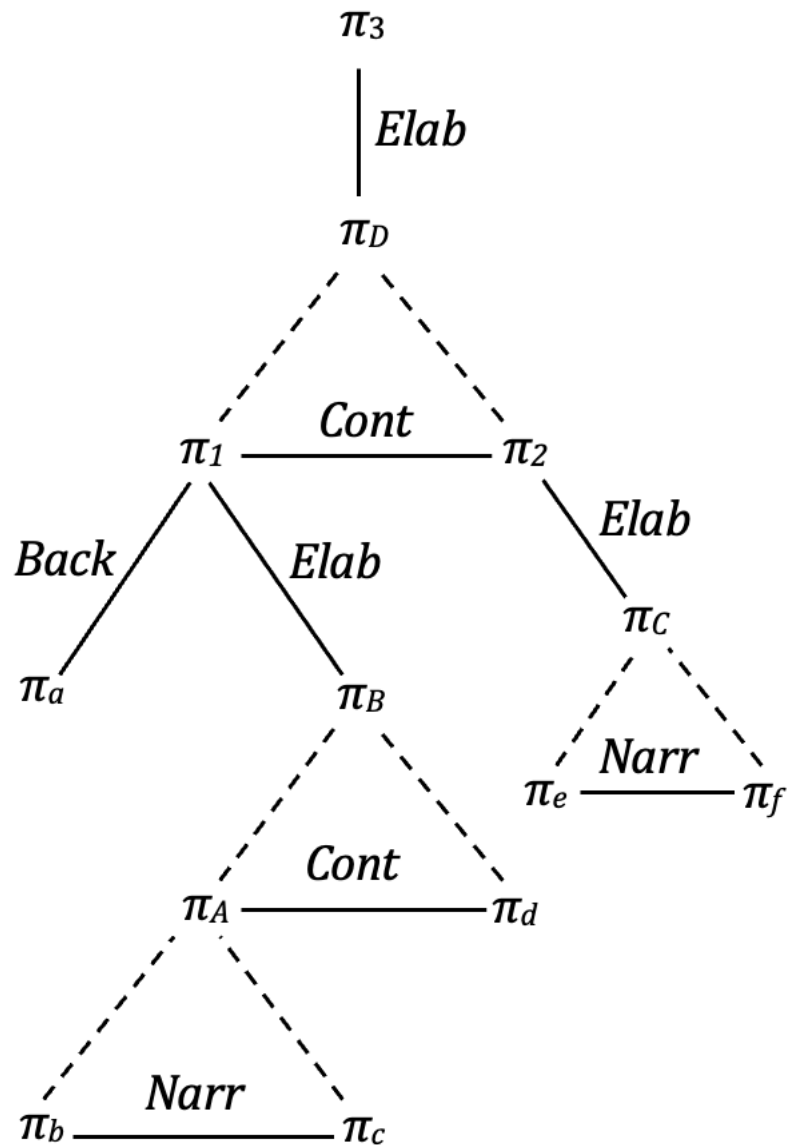


Figure 4.3: Discourse representation structure for (29), involving a complex discourse unit consisting of several FRAMES contributing to a common Framing Topic for a foreground constituent

INDEPENDENT RHEMES. In other cases, the construction of the framing topic may not be explicit, i.e. it might not be achieved with framing constructions, but merely by reference to the discourse at large—in fact, the accommodation of presupposition is one of the main functions of the SDRT relation Background (specifically in its *Background_{forward}* ‘flavour’) (Asher & Lascarides 1998; Asher et al. 2007). A simpler discourse structure showing both the presence of INDEPENDENT RHEMES in the foreground part of the structure in (4.2) and implicit background is the one for (23) above, whose structure might look as shown in Figure (4.4). In Figure 4.4 each discourse segment (π_a) to

(π_d) corresponds to each conjunct participle in (23), while (π_e) corresponds to the main verb. The (π_1) constituent is structurally the same as (π_2) in Figure 4.3, but instead of attaching to some discourse constituent (π_a) in the sentence, it will attach to the preceding discourse (via *Background*) by solving the presupposition $\delta(\pi, \pi_1)$.

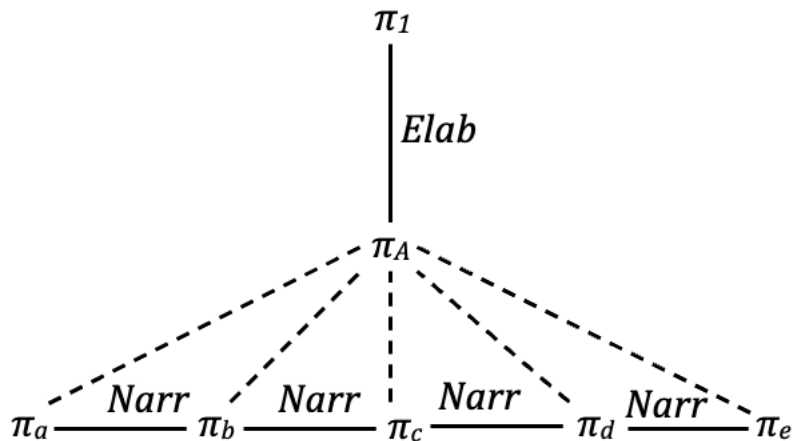


Figure 4.4: Discourse representation structure for (23), involving an implicit Framing Topic

4.3 Summary

This chapter provided a tentative account of the typical functions of *jegda*-clauses and participle constructions by drawing from frameworks for discourse representation (SDRT), particularly from accounts of temporal adverbials and time-locating adverbs.

We have seen that previous work (such as Behrens et al. 2012) on the competition between finite clauses and non-finite adjuncts (loosely) based on discourse representation theories started from the assumption that the main difference between finite subordinates and non-finite adjuncts is that the former expresses a more precise semantic relation to the matrix. This observation potentially holds when comparing clearly framing *perfective* participle constructions with *perfective jegda*-clauses. In that case, it is possible that greater semantic ‘preciseness’ brought up by Behrens et al. (2012) can be seen in the fact that framing perfective participles always cause narrative progression, which often has the effect that the matrix eventuality is located *at some point soon* after the event referred to by the participle (within the consequent state). With perfective *jegda*-clauses, this is not always necessarily the case, and we can, in fact, get an interpretation whereby the perfective *jegda*-event includes a perfective matrix event (as in (8) above, which Declerck (1997) would define as ‘occasion’-referring *when*-clauses), that is, it is a more precise contribution only insofar as the *jegda*-clause introduces a bounded temporal referent within which the matrix event is set. On the other hand, perfective framing participles

seem to always introduce a consequent state and set the matrix event within it.

We have then observed that, unlike claimed by Behrens et al. (2012), not only do INDEPENDENT RHEMES seem to typically trigger the SDRT relation *Narration*, but they may specifically be introducing a subtype of it, which Bras et al. (2001) call *Strong Narration*. Strong Narration involves the pragmatic inference that the temporal relation between the event expressed by an INDEPENDENT RHEME and the following event in line (another INDEPENDENT RHEME or its matrix) is one of *abutment*, namely of strict sequentiality with a stronger discourse topic requirement, which fits the idea that participle clauses as a whole are particularly well-suited as devices for narrative cohesion.

Finally, we formalized a typical Early Slavic example containing multiple framing clauses on the basis of the SDRT accounts of the rhetorical relation *Background*. The formalization suggested that the discourse properties of this relation and related observations on *Narration* are consistent with the observations made on both framing constructions (*jegda*-clauses and participles) and independent rhemes, particularly the discourse topic requirement and its attachment to the subsequent discourse via *Narration* and to the previous discourse via *Background*.

Part II

The constructions in cross-linguistic perspective

Chapter 5

The semantic map of WHEN and its typological parallels

5.1 Massively parallel Bible corpora: beyond ancient Indo-European translations

The previous chapters have shown that the discourse functions of Early Slavic predicative participle can partly be inferred compositionally from a combination of syntactic configuration, tense-aspect of both matrix and participle clause, and lexical fillers.¹ Following Bary & Haug's (2011) treatment of Ancient Greek predicative participles, Early Slavic participle clauses were treated as grammaticalized discourse relations, with a main tripartite distinction between FRAMES, INDEPENDENT RHEMES, and ELABORATIONS. We have seen that conjunct participles and absolute constructions have overall quite different distributions and discourse functions in the Early Slavic corpus, despite their partial competition as FRAMES at the beginning of a discourse segment. We have also seen that finite subordinates introduced by *jegda* 'when' compete with both constructions as frame setters, but overall appear to be distributionally more similar to absolute clauses.

To allow us to make generalizations about core semantic-pragmatic properties of different competing temporal constructions, this chapter takes a temporary step back from Early Slavic and Greek with the goal of investigating whether the division of labour between different participle clauses and finite temporal subordinates in Early Slavic and Greek corresponds to well-attested cross-linguistic patterns. English *when* is used as the new point of departure, for two main reasons. First, the scope of *when*-clauses spans both that of *jegda*-clauses and that of conjunct participles and absolute constructions, as can be evinced by the counterparts to each of these constructions in different English versions of the Bible, that is, all three OCS constructions can correspond to a *when*-clause in English. In fact, any of the three Early Slavic competitors would be considered under the generic label of WHEN-clauses in some typological frameworks to temporal subordination (cf. Cristofaro 2013b).² Second, their semantic properties have been more

¹This chapter is the result of a collaboration with Dag Haug (Haug & Pedrazzini 2023).

²Small caps WHEN will be used throughout the chapter to refer to the semantic concept, and italicized *when* for the English lexical item.

widely studied and are relatively well understood, despite their complexity, as we have seen in Chapter 3.

To explore the expression of WHEN cross-linguistically and see how the ground covered by English *when* is expressed across languages, we use Mayer & Cysouw’s (2014) massively parallel corpus, which contains the New Testament in more than 1400 languages. The data is described in detail in Section 5.2.

To explore the data, probabilistic semantic maps are employed, as they are now a well-established tool in language typology for capturing universal correspondences between classes of forms and ranges of highly similar situational meanings across ‘massively cross-linguistic’ datasets (Wälchli & Cysouw 2012). Probabilistic semantic maps can deal with very large datasets containing great degrees of variation within and across languages (Croft & Poole 2008), and unlike traditional implicational semantic maps (Haspelmath 2003) they do not rely on a limited set of posited abstract functions and translational equivalents. The premises and central aim of this thesis—to investigate the competing motivations governing the choice of similar temporal constructions—strikes a clear chord with the theoretical bases of probabilistic semantic maps, namely:

- *similarity semantics*, whereby similarity is considered to be a more basic notion than identity (see Wälchli & Cysouw 2012);
- *exemplar semantics* (Croft 2003), whereby the abstract functional domains assumed by traditional semantic-map approaches (Haspelmath 2003) are in fact clusters of ‘contextually embedded’ situations.

Wälchli & Cysouw (2012: 677) specifically argue for probabilistic semantic maps as a good fit for analyses in which usage-related factors, including discourse structure, are particularly relevant, which is a welcome observation in this study, since *jegda-* and participle clauses play an important role in discourse-structuring, as seen in the previous chapters. The methods used to generate and analyse the semantic maps for the analysis are described in detail in Section 5.3.

5.2 Data

Mayer & Cysouw (2014) massively-parallel Bible corpus³ comprises translations representing 1465 ISO 639-3 language codes.⁴ As noted in Good & Cysouw (2013), an ISO 639-3 code should be understood as referring to a LANGUOID, a generalization of the term *language* referring to the grouping of varieties as represented in specific resources (DOCULECTS) without the common constraints associated with the definition

³I am grateful to Michael Cysouw for sharing the repository with the full parallel corpus with me and Dag Haug.

⁴As of January 2023.

of language, dialect or family. This is crucial for avoiding creating the misconception that the ‘languages’ represented in the dataset are defined as such by virtue of their sociolinguistic status. Rather, each of them can be considered as sets of DOCULECTS at some level of hierarchical grouping. For practical purposes, the varieties represented by each Bible translation in the parallel corpus will be referred to as ‘languages’, with the caveat in mind that not all the varieties with an ISO 639-3 code will equally correspond to what is generally considered a ‘language’.⁵

Several of the languages in Mayer & Cysouw’s 2014 parallel corpus have multiple translations and a few contain only (or predominantly) the Old Testament. To obtain the best textual coverage for the largest number of varieties possible, only languages with a version of the New Testament were considered. For languages with multiple translations, the New Testament version with the widest coverage in terms of verses was selected. If the difference in coverage between versions was less than 2000 verses, the different versions were considered as having the same coverage, in which case the most recent one was chosen.

Although Mayer & Cysouw’s corpus already contains versions for some historical languages, for Ancient Greek, Church Slavonic, Latin, Gothic, and Classical Armenian, their versions from the PROIEL Treebank (Haug & Jøhndal, 2008) were used to facilitate the potential integration of their several layers of linguistic annotations in the semantic maps.

The final dataset comprises 1444 languages (around 19% of the world’s languages), representing, following the Glottolog classification, 121 families and 16 language isolates. In comparison, the world’s languages are currently classified into 233 families and 167 isolates.⁶ Tables 5.1-5.2 give an overview of the language families most represented in our dataset compared to their frequency in the world’s languages according to the Glottolog database. The top three families among the world’s languages, the Atlantic-Congo, Austronesian, and Indo-European occupy the same position in our dataset and show very similar relative frequencies to those found in Glottolog. By comparing Table 5.1 with Table 5.2 we can see that the Nuclear Trans New Guinea, Sino-Tibetan, Otomanguean are slightly overrepresented in our dataset compared to the world’s languages, while

⁵The Glottolog database (<https://glottolog.org>; Nordhoff & Hammarström 2011, Hammarström, Forkel, Haspelmath & Bank 2023), for example, which adopts a DOCULECT-based approach while also grouping languoids into successively larger ‘levels’ (such as subdialects, dialects, languages, subfamilies and families) classifies 15 of the languages in our dataset as dialects. Norwegian Bokmål (NOB) and Norwegian Nynorsk (NNO), for example, are considered ‘dialects’ of Norwegian (NOR), even though the latter is in fact defined collectively by the combination of the former two (among other ‘dialects’). ‘Norwegian’ (NOR), then, could therefore be considered as a languoid at a higher hierarchical level than the languoids Norwegian Bokmål and Norwegian Nynorsk.

⁶These numbers do not include some of the ‘non-genealogical trees’ to which some languages are assigned by Glottolog, specifically UNCLASSIFIABLE, UNATTESTED, and SPEECH REGISTER. SIGN LANGUAGES, MIXED LANGUAGES, and PIDGINS are instead considered in the numbers and they are therefore counted in the frequencies in Tables 5.1-5.2. So-called BOOKKEEPING LANGUOIDS are also excluded from the counts. These exclusions explain why the figures reported here are slightly different from those reported on the Glottolog webpage (<https://glottolog.org/glottolog/glottologinformation>).

	family	bible_raw	bible_rel	world_raw	world_rel
	Atlantic-Congo	249	17.2%	1380	18.1%
	Austronesian	246	17.0%	1289	16.9%
	Indo-European	110	7.6%	595	7.8%
Nuclear Trans New Guinea		94	6.5%	313	4.1%
	Sino-Tibetan	90	6.2%	441	5.8%
	Otomanguean	79	5.5%	180	2.4%
	Afro-Asiatic	47	3.3%	371	4.9%
	Quechuan	27	1.9%	45	0.6%
	Uto-Aztecan	26	1.8%	64	0.8%
	Mayan	25	1.7%	35	0.5%

Table 5.1: The 10 most frequent language families in our dataset compared to the 10 most frequent families among the world’s languages according to the Glottolog classification. *bible_* refers to the former, *world_* to the latter. *raw* are raw number of languages belonging to the relevant family, *rel* is the relative frequency of these in relation to the total number of languages in the respective dataset (the parallel Bible dataset for *bible_*, the whole Glottolog language database for *world_*).

	family	world_raw	world_rel	bible_raw	bible_rel
	Atlantic-Congo	1380	18.1%	249	17.2%
	Austronesian	1289	16.9%	246	17.0%
	Indo-European	595	7.8%	110	7.6%
	Sino-Tibetan	441	5.8%	90	6.2%
	Afro-Asiatic	371	4.9%	47	3.3%
Nuclear Trans New Guinea		313	4.1%	94	6.5%
	Pama-Nyungan	268	3.5%	13	0.9%
	Otomanguean	180	2.4%	79	5.5%
	Austroasiatic	160	2.1%	10	0.7%
	Tai-Kadai	91	1.2%	3	0.2%

Table 5.2: The 10 most frequent families among the world’s languages according to the Glottolog classification compared to the 10 most frequent language families in our dataset. The same labelling as Table 5.1 applies.

this is certainly the case for the Quechuan, Uto-Aztecan, and Mayan families. On the other hand, the Afro-Asiatic, Pama-Nyungan, Austroasiatic, and Tai-Kadai families are rather heavily underrepresented in our dataset. The families not represented at all in our dataset constitute around 48% of the world’s families and comprise, for the most part, fewer than 10 languages.⁷

In terms of areal distribution, following the Glottolog classification into six main macro-areas (Africa, Australia, Eurasia, North America, South America, Papunesia),

⁷The complete list of the world’s language families used to extract the counts reported here, their frequency according to Glottolog and in our dataset can be found in the Appendix.

macro-areas	bible_raw	bible_rel	world_raw	world_rel
Papunesia	415	28.7%	2136	28.1%
Eurasia	336	23.3%	1743	22.9%
Africa	335	23.2%	2196	28.9%
North America	181	12.5%	674	8.9%
South America	157	10.9%	488	6.4%
Australia	20	1.4%	371	4.9%

Table 5.3: Areal distribution of the languages in our dataset compared to the world’s languages, following the classification into macro-areas from Glottolog.

as Table 5.3 shows languages from Africa and Australia are underrepresented in our dataset, particularly from the former, while languages from the Americas are somewhat overrepresented. Figure 5.1 maps the distribution of the languages in our dataset among the world’s languages.⁸

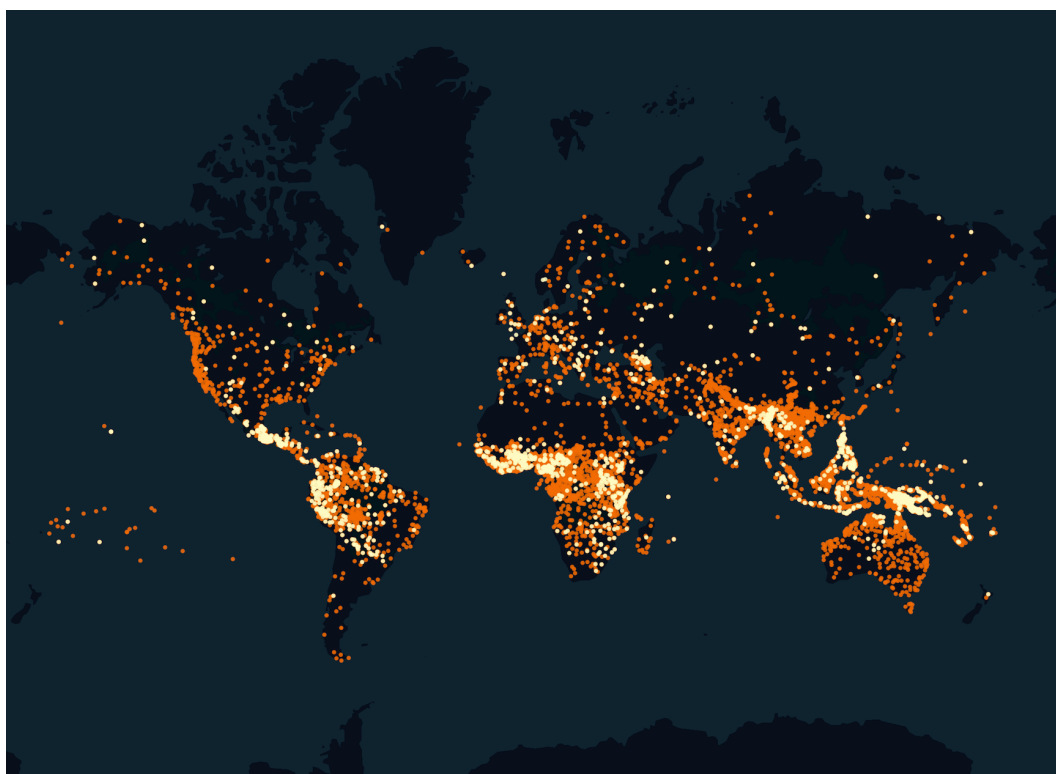


Figure 5.1: Approximate areal distribution of the languages in our dataset (light yellow) among the world’s languages (orange)

⁸The points in the maps are obviously an approximation of where a particular language is used. The coordinates for the map in Figure 5.1 are from Glottolog.

5.3 Methods

The texts in the target languages were aligned to the English one at the word level, using SyMGIZA++ (Junczys-Dowmunt & Szał 2012), a modification of the well-known GIZA++ program (Och & Ney 2003) that allows to train two-directed word alignment models in parallel. The results are one-to-one alignment models, namely one token in the source language corresponding to only one token, or no token at all, in the target language (as opposed to a one-to-many or many-to-one alignment).

SyMGIZA++ was chosen over the popular and much faster FastAlign model (Dyer, Chahuneau & Smith 2013) after applying some heuristics to gauge the quality of their results. Because of the sheer number of languages in the parallel corpus, some bias in the evaluation method will necessarily be introduced, since it requires familiarity with both source and target language. Rather than evaluating the alignment across the board, we checked a randomly selected subsample (10%) of all the sentences (= 876) containing *when*-clauses in English and calculated the accuracy of the alignment between the token *when* and its respective forms, or lack thereof, in the Norwegian, Russian and Italian versions. SyMGIZA++ yielded 96.5% accuracy on the Norwegian test set, 90% on the Russian, and 77.9% on the Italian one, whereas FastAlign only yielded 77.9%, 75.9% and 59.3%, respectively. Overall, SyMGIZA++ and FastAlign performed similarly at identifying the correct parallel when the target language uses a subordinator, but FastAlign generally aligned *when* to some other token in the absence of a direct parallel (e.g. to a conjunction or an auxiliary verb), whereas SyMGIZA++ more often explicitly indicated the lack of a parallel with a ‘NULL’ alignment, which intuitively means that the target language uses a construction with no subordinator (e.g. an independent clause) or a construction where the subordination is expressed morphologically (e.g. a converb).

Before training the final models with SyMGIZA++, minimal preprocessing (lower-casing and punctuation removal) was applied. The occurrences of English *when* and its parallels in all the target languages were then extracted. Each instance of *when* and its parallels was treated as one usage point for the hypothesized semantic concept WHEN, and Hamming distance was used as a measure of similarity between pairs of contexts, leveraging the number of times a language uses two different linguistic means where English uses one (*when*). For example, based on the six languages shown in Table 5.4, the distance between the two contexts is 3, because they differ in the word choice in Maori (*mri*), Finnish (*fin*) and Kazakh (*kaz*).

	eng	mri	por	...	fin	kaz	kor
1	when	no	quando	...	kun	қашан	때에
2	when	ka	quando	...	jolloin	кейін	때에
<i>n</i>

Table 5.4: Matrix of *when* and aligned tokens (sample).

After calculating the Hamming distance between all pairs of contexts, we obtain a $n \times n$ Hamming-distance matrix, where both column and rows correspond to the contexts, and each cell indicates the Hamming distance between the respective two contexts.

Multidimensional scaling (MDS), as implemented by the `cmdscale` function from the R package *stats* (Team 2013), is then applied to the resulting matrix to visualize the distance between each occurrence of *when* two-dimensionally. This output is shown in Figure 5.2. Each dot represents a usage point for *WHEN* (i.e., a Bible verse). If two dots are far apart, they tend to be expressed with different lexical items across the languages in the corpus.

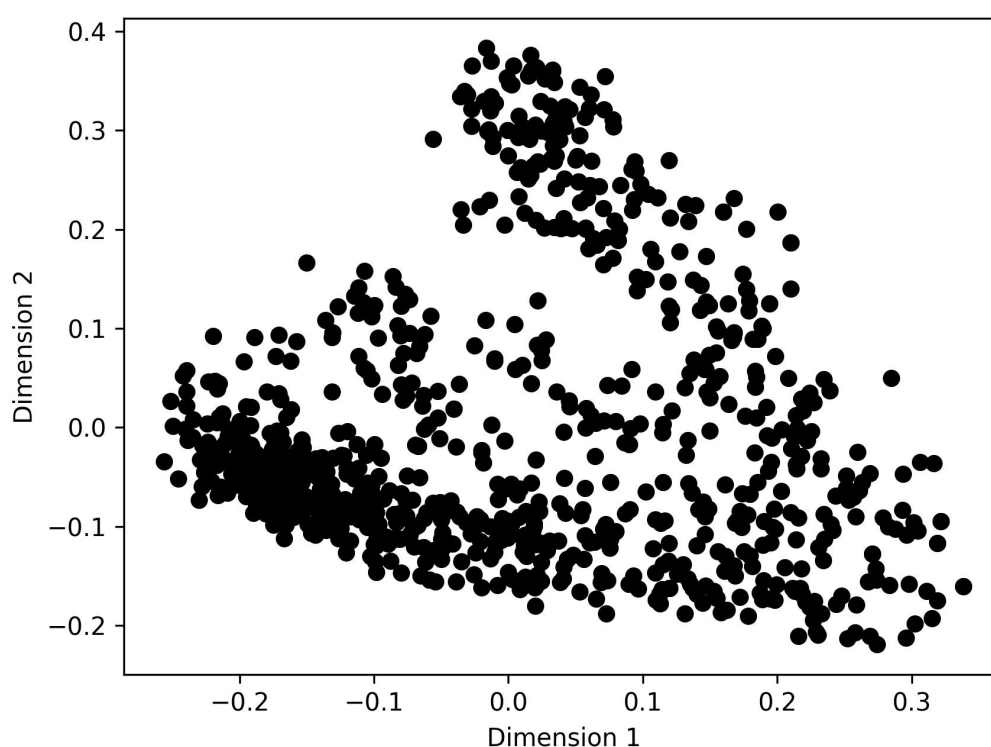


Figure 5.2: Basic semantic map of *WHEN* resulting from plotting the first two dimensions of the MDS matrix

Clusters of semantically similar observations are identified and analysed in two main ways. First, similarly to Hartmann et al. (2014), starting from the MDS matrix, Kriging is applied as an interpolation method that uses a limited set of sampled data points (each observation in the target languages) to estimate the value of a variable in an unsampled location.

As an example, Figure 5.3 shows the resulting semantic map for Doyayo (Atlantic-Congo, Africa) after applying Kriging to the MDS matrix by using the parallels to English *when* in the language to interpolate the areas shown in green, red, and yellow in the

Doyayo (Atlantic-Congo, Africa)

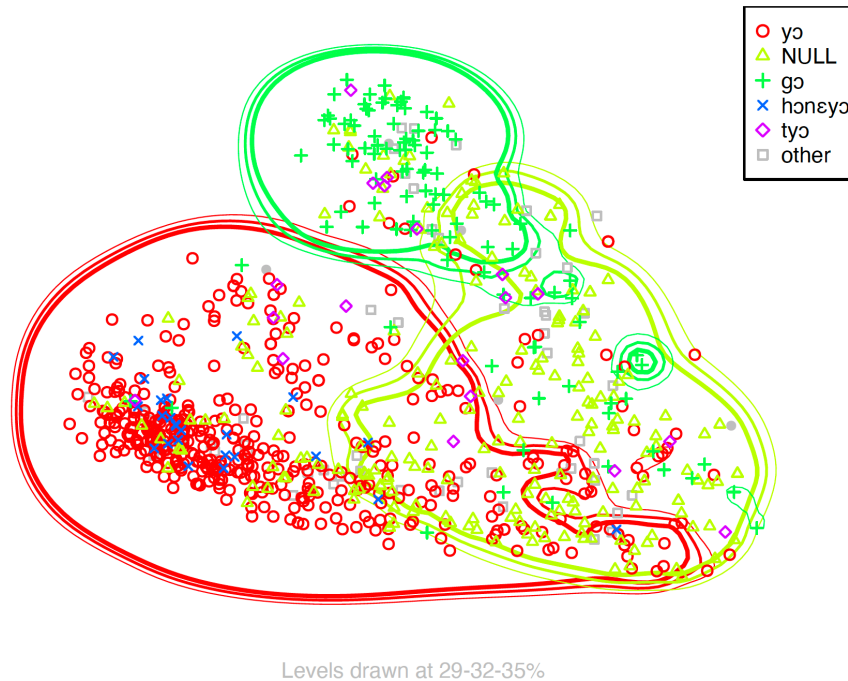


Figure 5.3: Kriging map for *when*-parallels in Doyayo (Atlantic-Congo, Africa).

figure. Unlike Hartmann et al. (2014), we start from one single means (*when*), without pre-emptively assigning a semantic label to the different *when*-situations in English, so that the discernible Kriging-areas in the semantic maps of the target languages must be interpreted on the basis of comparison between similar cross-linguistic patterns. Like Hartmann et al. (2014), the `Krig` function from the R package *fields* (Nychka, Furrer, Paige & Sain 2021) was used to draw lines at different levels of probability distributions (35%, 32% and 29%). Unlike traditional semantic maps, where boundaries are drawn around all observations of the same type (i.e. the same means in a given language), the lines in the Kriging map in Figure 5.3 represent the probability for a particular means to be dominant within their boundaries. This is why, for instance, red points in Figure 5.3 can also be found outside the red area identified by Kriging, but it also explains why relatively large areas can overlap, as the points between the red and yellow areas in the figure show. The latter is a welcome feature of this approach, albeit more challenging to account for than non-probabilistic ones, since it reflects the intuition that situations encoded by any two given competing constructions may be envisioned as the union of the sets of all situations which can be encoded by either of those constructions.

Second, we fit a Gaussian Mixture Model (GMM) to the MDS matrix, to identify clusters which are more likely to correspond to separate universal functions of *WHEN*,

regardless of how much variation a particular language shows within any of the clusters (which could go from no variation across the whole map or across one cluster, to several linguistic means in a single cluster). The result is shown in Figure 5.4.

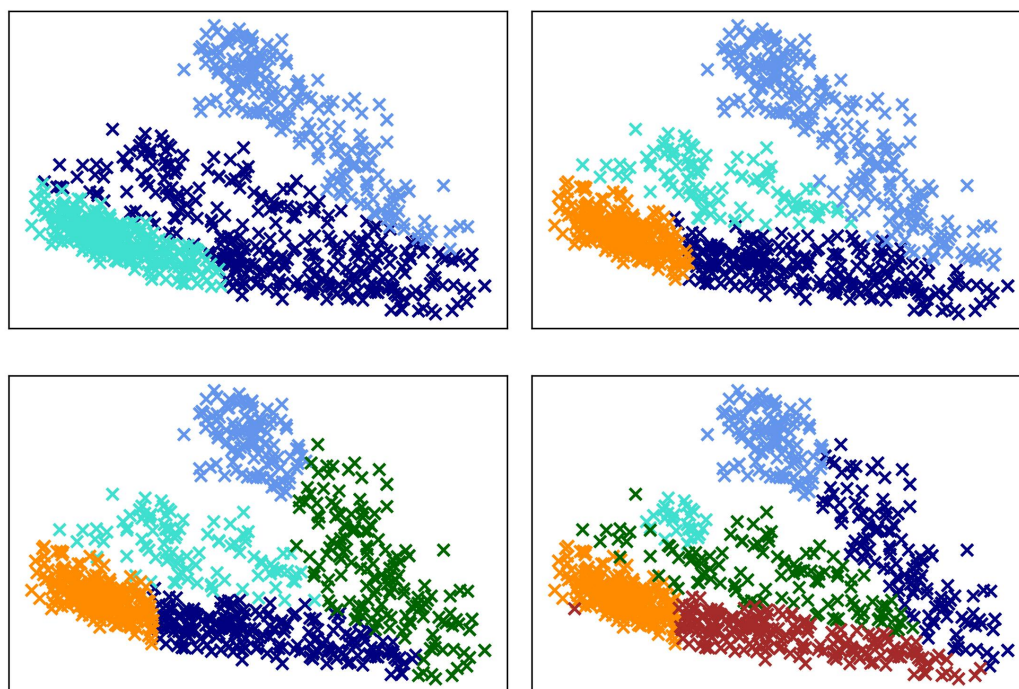


Figure 5.4: Clusters identified via Gaussian Mixture Modelling and k -means initialization, based on the first two dimensions of the MDS matrix: comparison between 3, 4, 5 and 6 clusters.

GMM assigns data to a given number of clusters based on probability distributions rather than on the distance from a centroid, as in other well-established clustering algorithms (e.g. k -means). This means that GMMs also allow for elliptical clusters, unlike k -means, which will always result in roughly spherical shapes, since they are based on mere Euclidean distance and thus cannot sensibly approximate the semantic map of competing constructions, which are, by definition, more of a continuum than a set of clearly separate areas.

The number of components (clusters) for the GMM models is chosen with the help of different heuristics. In Figure 5.5 (left), we can see that the optimal number (based on the first two dimensions of the MDS matrix) is 6 clusters. Figure 5.5 (right) indicates the best number of clusters based on the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978). Both AIC and BIC agree with the Silhouette score that 6 is the number of clusters for the best trade-off between model fit and complexity.

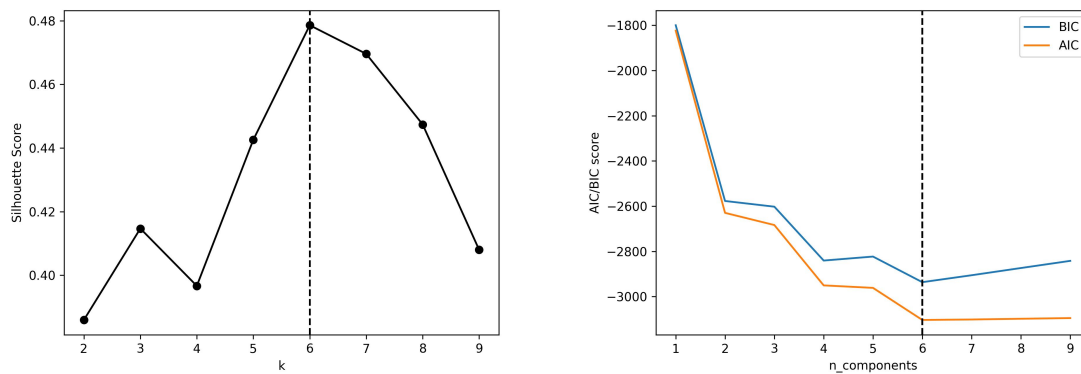


Figure 5.5: Evaluation of the optimal number of clusters for a Gaussian Mixture Model of the two-dimensional MDS matrix. **(Left)** Silhouette score: higher value is better. **(Right)** Akaike information criterion (AIC) and Bayesian information criterion (BIC): lower value is better.

These methods are meant to indicate how many clusters can be considered maximally separate from each other. However, empirically, we know that the temporal constructions under consideration are often competing and that their scopes are not at all clear-cut. While taking this into account, the GMM model consisting of 6 clusters (Figure 5.6) was selected, given the agreement between the Silhouette and the BIC/AIC scores.

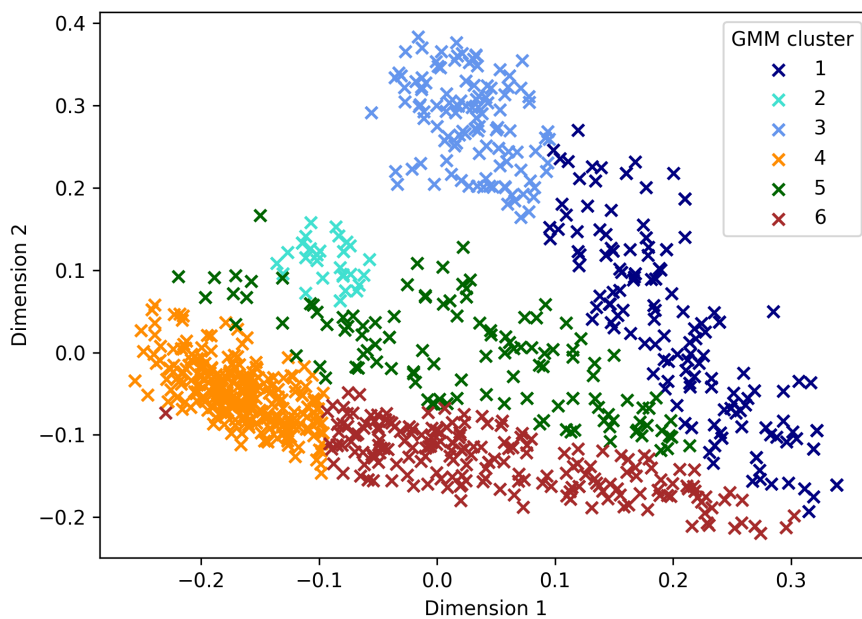


Figure 5.6: Gaussian Mixture Model of the two-dimensional MDS matrix (6 clusters)

Each cluster in Figure 5.6 *potentially* corresponds to a specific functional domain of

WHEN. To test whether this is in fact the case, we check to what extent the languages have lexical items that align well with each GMM cluster. Concretely, all the attested means used by each language are first extracted for the individual clusters. For each attested means, its occurrences in that cluster are counted as *true positives*, its occurrences outside that cluster as *false positives*, and the occurrences of other means in that cluster as *false negatives*. The *precision* of a means as a rendering of the *when*-clauses corresponding to that cluster, then, is the number of true positives divided by the sum of true positives and false positives; the *recall* is the number of true positives divided by the sum of true positives and false negatives. We can then compute the *F1 score* (harmonic mean of precision and recall) for each means and, for each language, plot the precision and recall of the word with the highest F1 scores. A means with a high F1 score will correspond to a likely lexification of the relevant GMM cluster in its particular language. A frequent high F1 score across several languages may instead indicate a common lexification pattern, as we will see in Section 5.4.

Finally, leveraging the advantages of both the Kriging and the GMM methods, the Kriging areas that best correspond to each GMM cluster in each language can be identified. As we will see in Section 5.4, this allows us to study colexification patterns across languages. The alignment of Kriging areas and GMM clusters runs as follows.

1. For each of n number clusters, across which cross-linguistic variation in co-lexification is to be investigated, calculate its centroid. This is the sum of the coordinates of the points belonging to each cluster, divided by the number of points in the cluster, namely:

$$\left(\frac{1}{j} \sum_{i=1}^j x_i, \frac{1}{j} \sum_{i=1}^j y_i\right)$$

where j is the number of points in a GMM cluster, x are the x-coordinates (i.e. dimension 1 of the MDS matrix) and y the y-coordinates (i.e. dimension 2 of the MDS matrix). The centroid of a GMM cluster is preliminarily assumed to be the best representation of that cluster. Note that it is unlikely to correspond to an actual observation in the target languages.⁹

2. For each GMM cluster, extract k actual observations corresponding to the k -nearest neighbours of the centroid of that cluster. The value k should be adjusted in a trial-and-error fashion; we set ours to 30, i.e. 30 points are extracted for each cluster. The nearest neighbours were identified using the balltree approach (Omohundro 1989), a space partitioning system which can be applied to multi-dimensional space

⁹The procedure can, in principle, have groups of observations drawn from clusters obtained with any method as a starting point. A group can also be made of one individual observation.

for nearest neighbour search.¹⁰ The result of the search is a group of ‘core’ points surrounding the centroid of the GMM cluster. Figure 5.7 shows the three groups from our experiment.

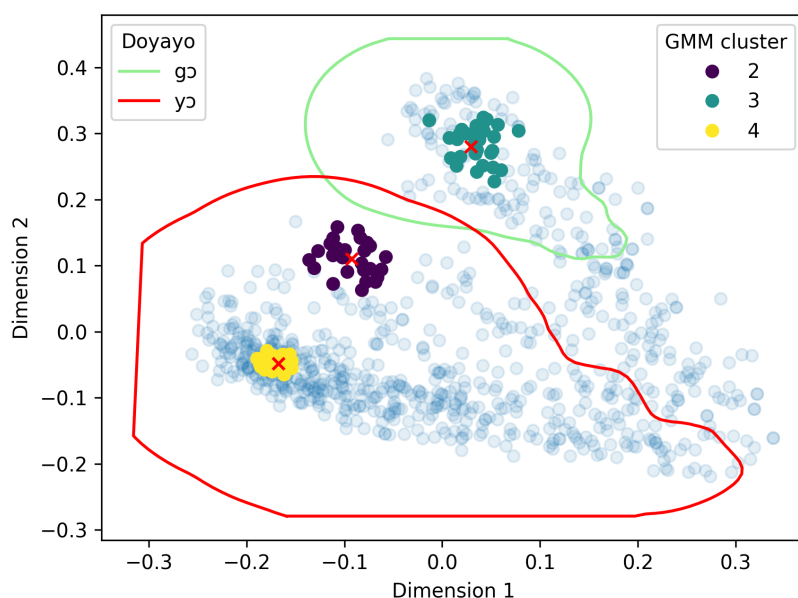


Figure 5.7: Result of the 30-nearest-neighbour search using the balltree method, with an example of its application to Doyayo (Atlantic-Congo, Africa). The red marks are the centroid of the respective GMM clusters (as represented in Figure 5.6), while the points in which they are embedded are their 30 nearest neighbours. The contour lines in green and red correspond to the Kriging areas for Doyayo $g\text{ɔ}$ and $y\text{ɔ}$ at 29% probability.

3. For each language, check which Kriging area, if any, contains each of the groups in Figure 5.7 and, for each language, create a dictionary to take note of the mapping between groups and Kriging areas.¹¹ For example, the group of points corresponding to GMM cluster 3 in Figure 5.7 are contained within the Kriging area for Doyayo $g\text{ɔ}$, those corresponding to GMM cluster 2 and 4 are contained within the Kriging area for $y\text{ɔ}$. The resulting dictionary for Doyayo, then, is $\{\text{group-3: } g\text{ɔ}, \text{group-2: } y\text{ɔ}, \text{group-4: } y\text{ɔ}\}$, meaning that all the points of each group are contained within *one* Kriging area only. This is the simpler scenario.

The more complex scenario is one in which more than one Kriging area includes points from the same group.

For example, the dictionary for Patep (Austronesian, Papunesia) is $\{\text{group-3: } [\text{ob}\acute{e}c, \text{buc}], \text{group-2: } \text{buc}, \text{group-4: } \text{NULL}\}$, meaning that the Kriging area

¹⁰We used the implementation of balltree by Scikit-Learn (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Perrot & Duchesnay 2011).

¹¹To obtain this information, we used Kriging areas at 29% of probability.

for *buc* contains points from groups 2 and group 3, but points from group 3 are also found in the Kriging area for *obêc*. In such cases, we apply the following heuristics to infer whether more than one Kriging area should be considered meaningful in that group for the purpose of looking at patterns of coexpression.

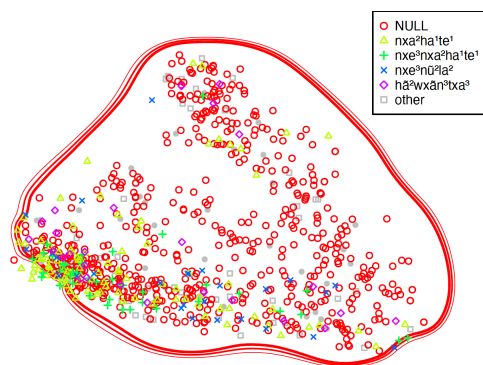
- a. If one of the two Kriging areas is unique to a given group (e.g. *obêc* in the Patep example), while the other is not (e.g. *buc* in the Patep example), consider the former as meaningful, regardless of how many points from that group it contains. Instead, consider the latter as meaningful only after running a test of proportion with the competing Kriging area. If it contains significantly more points than the competing area, or if the difference in proportion is not statistically significant, then both Kriging areas are kept in the dictionary. To determine this, we use Fisher's exact test with $\alpha = 0.01$. In the Patep example above, *obêc* is considered a meaningful Kriging area for group 3 because it is only found there. On the other hand, to decide whether to also keep the Kriging area for *buc*, we run a Fisher's test, which indicates that the difference in proportion is not significant (26 out of 30 points are found in the Kriging area for *obêc*, 21 out of 30 in the one for *buc*; $p = 0.32$), so both Kriging areas are considered meaningful lexicalization for group 3.
 - b. If neither of any two competing Kriging areas is unique to a particular group, then a Fisher's test is used to establish which one to consider meaningful. For example, the dictionary for Yucatec Maya (Mayan, North America) is {group-3: *ken*, group-2: [*ken*, *ka*], group-4: *ka*}. A Fisher's test indicates that the Kriging area for *ken* contains significantly more points from group 2 than the Kriging area for *ka* ($p < 0.01$), so the dictionary is modified to {group-3: *ken*, group-2: *ken*, group-4: *ka*}.
 - c. Give lexical items a greater weight than NULLs. Only consider a NULL Kriging area as meaningful if it is the only one containing a particular group. For example, the dictionary for Manam (Austronesian, Papunesia) is {group-3: [*bong*, NULL], group-2: *bong*, group-4: [*bong*, NULL]}, which, for the purpose of looking at lexicalization patterns is then modified to {group-3: *bong*, group-2: *bong*, group-4: *bong*}. On the other hand, the dictionary for Hills Karbi (Sino-Tibetan, Eurasia) is {group-3: *ahut*, group-2: *ahut*, group-4: NULL}, in which case the only Kriging area containing points from group 4 is a NULL area.
5. Assign patterns of lexicalization based on the Kriging areas considered meaningful for each group.

As shown in Figure 5.7, for example, the points in group 3 all fall within the Kriging area for Doyayo ɔɔ, while those in both group 2 and 4 are all contained within the Kriging area for yɔ. On the basis of this, we can assign languages behaving like Doyayo to a pattern in which the top left area (cluster 3) is the domain of one word, whereas the mid and bottom left areas are colexicalized by a different word. This—which we will call ‘pattern C’ in Section 5.4—is one of five basic patterns which can be observed on the basis of the three groups of core points represented in Figure 5.7 (one for each possible logical combinations between the groups). An overview of the patterns will be given in Section 5.4.2.

5.4 Analysis

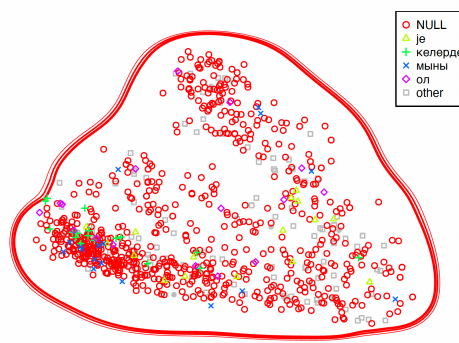
The very first, perhaps obvious, observation to make is that no target language has a one-to-one match with English. All target languages show at least some variation, but a common feature in the maps for all the languages is that if the interpolation method assigned all observations to one common area, then that area is a ‘null’ one, rather than a specific lexical item (i.e. a subordinating conjunction). Figures 5.8a and 5.8b are two examples of null-only maps from Southern Nambikuára (Nambiquaran) and Southern Altai (Turkic).

Southern Nambikuára (Nambiquaran, South America)



(a)

Southern Altai (Turkic, Eurasia)



(b)

Figure 5.8: Kriging maps for Southern Nambikuára (Nambiquaran, South America) and Southern Altai (Turkic, Eurasia)

As already mentioned in the Method section, *null* refers to either no parallel in the target language (e.g. the token *when* has no parallel in a typical participle construction), or items that have been reassigned a *null* alignment because their low frequency in the relevant languages likely suggested a misalignment (e.g. in a participle construction with an auxiliary, such as *having seen*, the auxiliary may have been considered by the model as a *when*-counterpart in that language). *Null* is also always one of any number of areas detected by Kriging for any language. That is, if Kriging detected two areas, for example,

one of these is a *null* area.

Throughout the analysis, null areas will not simply be regarded as the *lack* of a construction, but as corresponding to different means of juxtaposing clauses to achieve what the English version we are using achieved with a *when*-clause. This approach is in line with previous typological literature (Cristofaro, 2013b), which defines WHEN-clauses in functional terms, classifying as such not only those introduced by specific temporal conjunctions (e.g., English *when X did Y* or *when doing Y*), but also clauses that are simply juxtaposed and whose function must be contextually inferred, as in 1. In some other cases, languages may use specific verb forms to mark adverbial subordination, without, however, specifying their semantic relation to the main clause. This is the case of cross-linguistically well-attested converbs and predicative participles (Haspelmath & König, 1995), as in examples 2-6 from our data.¹²

- (1) Canela-Krahô (Macro-Gê; Cristofaro 2013b)

pê wa i-pým, pê inxê ty
 PST 1 1-fall PST mother die
 ‘My mother died when I was born’.

- (2) Kumyk (Turkic)

Etti èkmekni de, balyk'ny da al-yp, bir bašl-ap šûkûr dua
 seven bread PTC fish PTC take-CVB first start-CVB thanks prayer
ète, song syndyr-yp âk'čylaryna bere, olar da halk'g'a
 do.INF then break-CVB disciples.DAT give.INF they PTC people.DAT
pajlajlar.
 distribute.PST
 ‘Then he took the seven loaves and the fish, and when he had given thanks, he broke them and gave them to the disciples, and they in turn to the people’
 (Matthew 15:36)

- (3) Ukrainian (Indo-European)

A pobačivši, rozpovily pro vse te, ščo pro
 and see.PFV.CVB.PST recount.PFV.PST.PL about all that which about
Cju Dytynu bulo jim zviščeno.
 this child was 3.DAT reported
 ‘And when they saw it, they made known the saying that had been told them concerning this child’ (Luke 2:17).

- (4) Korean (Language isolate)

¹²All glossing my own unless otherwise stated.

geudeureun i malsseumeul deut-ko keuge nollamyeo geubunegeseo
 they this word hear-CVB greatly surprised from-him
tteonagassda
 went-away

‘When they heard it, they marveled. And they left him and went away.’ (Matthew 22:22).

(5) Italian (Indo-European)

a. *ma essendosi i discepoli radunati intorno a lui*
 but be.GER.REFL the disciples gather.PST.PTCP.PL around to him
egli si alzò ed entrò in città
 he refl rise.3SG.PST.PFV in city

‘But when the disciples gathered about him, he rose up and entered the city’ (Acts 14:20).

b. *e il centurione avendo sentito parlare di Gesù*
 and the centurion have.GER hear.PST.PTCP speak about Jesus
gli mandò alcuni anziani dei giudei
 3.DAT send.3SG.PST.PFV some elders of-the Jews

‘When the centurion heard about Jesus, he sent to him elders of the Jews, asking him to come and heal his servant’ (Luke 7.3).

(6) Ancient Greek (Indo-European)

eparantes de tous ophthalmous autōn oudena
 lift-up.PL.PFV.PST.PTCP.M.NOM PTC the eyes their nobody
eidon ei mē ton Iēsoun monon
 see.3PL.PFV.PST if not the Jesus alone

‘And when they lifted up their eyes, they saw no one but Jesus only.’ (Matthew 17.8)

Examples such as 2-6 above are captured as null alignments by the alignment models. This is because the verb form itself, be it a converb, a participle, or a finite (main) verb, is aligned to the verb form in the English *when*-clause. Since we chose a one-to-one alignment between source and target language, the connector *when* has no parallel in these examples and is therefore aligned to a ‘null’ token. Despite some inevitable noise in the alignment, due, for instance, to particularly frequent juxtaposed verb forms that may be interpreted as a recurring parallel to the token *when* itself (e.g. the auxiliary in *having done*), the overall reliability of the alignments is clear from the clear-cut pattern across the world’s languages regarding where in the semantic map of WHEN null constructions tend to cluster, as we will see in Section 5.4.1. Null alignments thus have significance in themselves—recall, in fact, that the abundance of null alignments was the very reason for choosing the specific model we are using, as opposed to others which, instead, aligned *when* to some other token in the absence of a direct parallel. Since our focus is on

the competition between *jegda*-clauses and participle constructions, understanding where null constructions cluster and what kinds of meanings they encode compared to lexified connectors (such as *jegda*) is crucial.

In the overview that follows, it is important to keep in mind what the Kriging maps and the GMM clustering are meant to capture, which are, in a way, complementary perspectives on the data. Kriging maps, such as those in Figures 5.3 and 5.8b-5.8a, are used to identify the means dedicated to expressing a particular function or meaning of *WHEN* in a specific language. Conversely, the aim of GMM clustering is to abstract from language-specific observations in order to understand which of those functions or meanings (for which individual languages may have dedicated means) tend to have dedicated means cross-linguistically.

In other words, our application of Kriging approximates the search for potential *grammatical morphemes* (GRAMS), while the GMM method is used to look for potential GRAM TYPES. A GRAM, according to Bybee & Dahl (1989), is a linguistic item—a bound morpheme, a lexical item or a complex construction—with a specific function or meaning. Bybee & Dahl's (1989) definition thus encompasses both specific lexical items aligned to English *when* (e.g. a subordinator) and null constructions. The concept of *gram* is more abstract than that of *morph* and is in many ways analogous to the concept of *construction* in Construction Grammar. It is language-specific and does not necessarily correspond to a crosslinguistic GRAM TYPE, which Dahl & Wälchli (2016) defines as 'a cluster of language-specific grams whose closeness in meanings and functions is reflected in similar distributions in a parallel corpus'. Together, the gram types make up the semantics atoms in the grammatical space that English *when* covers.

The morphosyntactic diversity exemplified by (2)-(6) echoes Dahl & Wälchli's (2016) remark that grams differ in how *transparent* they are, namely in how constant and isolable their form is, which has bearing on how easily they can be automatically identified via methods such as ours. English *when* is maximally transparent, as it is a single word with a constant form and can therefore be automatically identified with little obstacle. The Italian past participle clauses in (5), on the other hand, are less transparent, as the auxiliary, though marked as a GERUND and isolable from the participle form, can be either *essendo* (lit. 'being') or *avendo* (lit. 'having'), depending on the chosen participle verb. The Kumyk converb markers *-ap/-yp* in (2) are much less transparent since they are not easily isolable and are only two of several allomorphs. Cases like the Ancient Greek predicative participle in (6) are maximally opaque since their form depends on the constituent in the matrix clause with which they agree. Moreover, while participles in their predicative function are similar to converbs, they can often also occur in other, e.g. attributive contexts.

5.4.1 Overtly-subordinated versus ‘null’ constructions

As mentioned above, null constructions tend to cluster in a specific area of the semantic maps. We can capture this pattern across languages by counting how many languages in the dataset express a particular observation in the semantic map with a null construction as opposed to some other means. The heatmap in Figure 5.9 shows the concentration of languages expressing data points in the map with a null construction: the warmer the colour of a point, the more languages use a null construction, and the colder the colour, the more languages use a lexified means.¹³

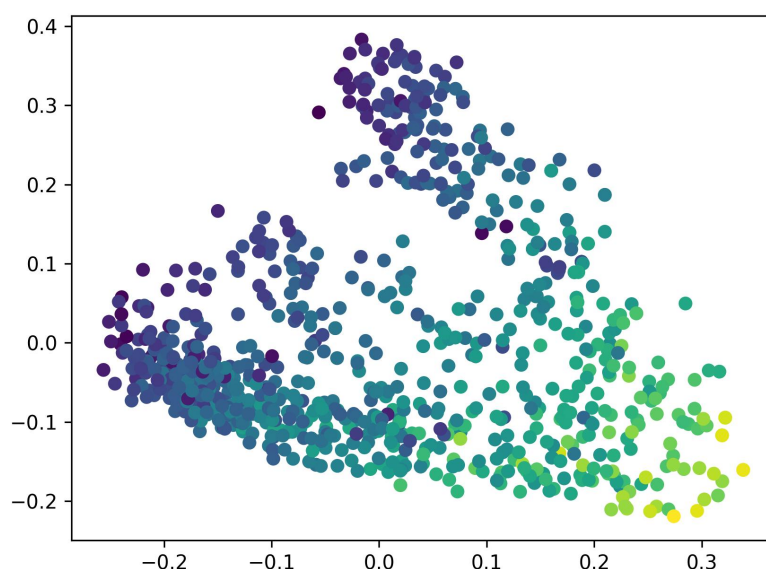


Figure 5.9: Heatmap showing the concentration of languages expressing a data point with a null construction as opposed to lexified means. The warmer the colour of a point, the more languages use a null construction.

The closer we get to the lower right corner of the map, the more likely it is that a language will use a null construction. 869 languages, for example, use a null construction for the verse in 7 from the very bottom right corner of the map.

- (7) And he took bread, and **when he had given thanks**, he broke it and gave it to them (Luke 22:19)

In fact, several other English translations also use a null construction for the same verse: (e.g. King James Version, New King James Version and New International Version: *and he took bread, **gave thanks** and broke it, and gave it to them, saying*). Note how Old Church Slavonic (8a) and Ancient Greek (8b) use an INDEPENDENT RHEME participle instead.

¹³An interactive version of the heatmap is available in the Observable notebook at the following address: <https://observablehq.com/@npedrazzini/3d-mds-when-map>.

- (8) a. *i priimō chlěbō chvalo*
 and take.PTCP.PFV.M.NOM.SG bread.ACC blessing.ACC
vъzdavъ prělo mi. i dastō
 give.PTCP.PFV.M.NOM.SG break.AOR.3.SG and give.AOR.3.SG
imō glę.
 3.PL.M.DAT say.PTCP.IPFV.M.NOM.SG
- b. *Kai labōn artōn eukharistēsas*
 and take.PTCP.PFV.M.NOM.SG bread.ACC thank.PTCP.PFV.M.NOM.SG
eklasen kai edōken autois
 break.AOR.3.SG and give.AOR.3.SG 3.PL.M.DAT
legōn
 say.PTCP.IPFV.M.NOM.SG
 ‘Then he took the bread, said the blessing, broke it, and gave it to them, saying’ (Luke 22:19)

At the other end of the spectrum, from the top left of the semantic map, we find examples like 9, which is expressed with a null construction by only 120 languages.

- (9) **When all things are subjected to him**, then the Son himself will also be subjected to him who put all things in subjection under him (1 Corinthians 15:28)

A *hōtan*-clause is also found in the Ancient Greek version (the Old Church Slavonic version does not have a translation for Bible verses outside of the Gospels), as (10) shows.

- (10) *hotan de hupotagēi autōi ta*
 when PTC subordinate.SBJV.PAS.PFV.3.SG 3.SG.M.DAT the.N.NOM.PL
panta tote kai autos ho huīos
 all things.N.NOM.PL then even self.M.NOM.SG the.M.NOM.SG son.M.NOM.SG
hupotagēsetai tōi hupotaksanti
 subordinate.FUT.PAS.3.SG the.DAT.SG subordinate.PTCP.PFV.PST.M.DAT.SG
autōi ta panta hina ēi
 3.SG.M.DAT the.N.ACC.PL all things.N.ACC.PL that be.SBJV.PRS.3.SG
ho theos ta panta en pasin
 the.M.NOM.SG god.NOM the.N.NOM.SG all things.N.NOM.PL in all.N.DAT.PL

It seems relatively common for languages to show an overall split between null- and non-null situations roughly to the right and to the left of the map, respectively, as is the case, for example, in Serbian (Figure 5.10a) and Moose Cree (Figure 5.10b), among several others.

After-clauses, if present in a language, appear to generally cluster at the bottom of the map, where competition between different linguistic means also seems to be more intense than in other areas. Romance languages are an example of maps with a dedicated *after*-area detected by Kriging, as in Romanian (Figure 5.11a), but several other languages show a similar pattern, like Lahu (Sino-Tibetan; Figure 5.11b).

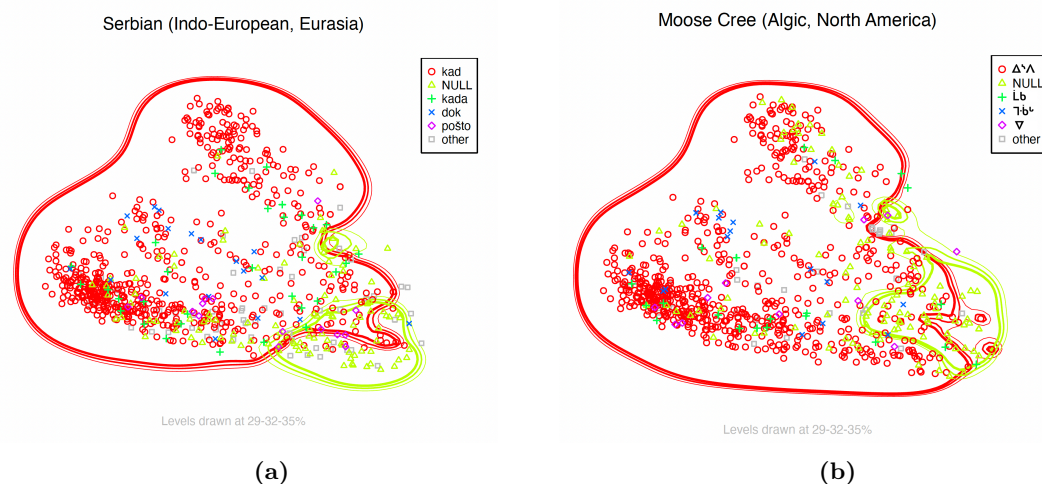


Figure 5.10: Kriging maps for Serbian (Indo-European, Eurasia) and Moose Cree (Algic, North America)

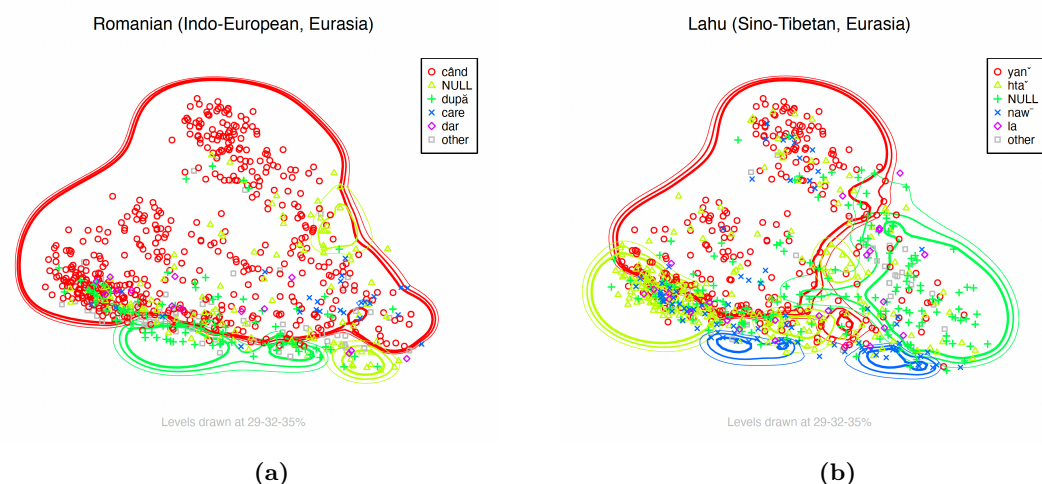


Figure 5.11: Kriging maps for Romanian (Indo-European, Eurasia) and Lahu (Sino-Tibetan, Eurasia)

In Figures 5.11a and 5.11a, respectively, Romanian *după* ‘after’ and Lahu *lhū* (part of the phrase *hk’aw naw* ‘after’; Matisoff 2020) occupy similar positions in the map, and in both cases there is a clear overlap between different linguistic means on or just around the area of *after*-clauses. It is, in fact, quite clear that the competition is particularly high across all languages roughly in the whole area highlighted in Figure 5.12.

The maps for several of the target languages even show two distinct overlapping areas (or areas within areas) in the bottom left of the map, as in Banggai (Figure 5.13a) and Gen (Figure 5.13b).

In fact, by adding a third dimension to the map, we can see that the observations in that area are distributed across the length of the *z*-axis, so that any two points that

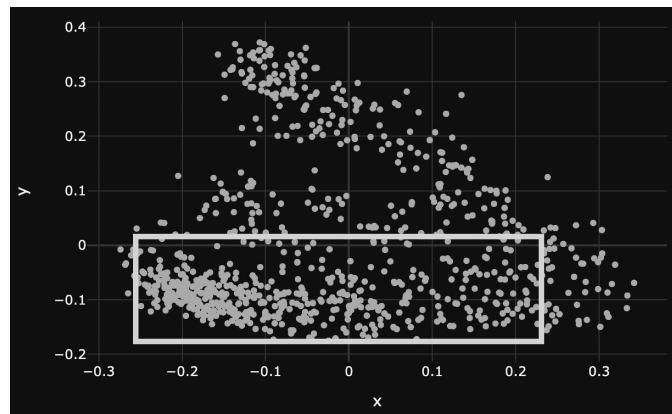


Figure 5.12: Approximate area in the maps where competition appears to be particularly high across all languages

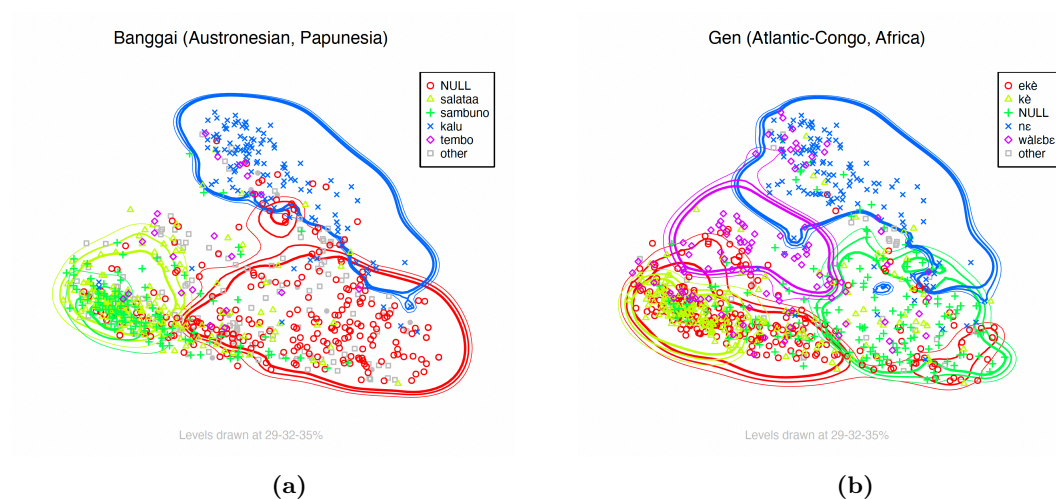


Figure 5.13: Kriging maps for Banggai (Austronesian, Papunesia) and Gen (Atlantic-Congo, Africa)

are close to each other in a two-dimensional plot may actually be very distant on a third dimension. Figure 5.14 shows snapshots of the three-dimensional MDS map from different angles, with observations coloured by how distant the observations are from the two-dimensional plane formed by the x and y axes only: the warmer the colour (red), the closer to the plane, the colder the colour (blue), the farther from the plane. (5.14a) correspond to the view with the x and y axes in the foreground and the z -axis on the background, namely the same perspective as the one obtained from the two-dimensional map in Figure 5.2 and elsewhere. (5.14b) shows the map from the perspective of the x - z plane, namely the one obtained by rotating (5.14a) 90° clockwise around the x -axis (i.e. by rotating (5.14a) upwards/backwards). (5.14c) shows the map from the perspective of the y - z plane, namely the one obtained by rotating (5.14b) 90° clockwise around the

z -axis (i.e. by rotating (5.14b) rightwards). Finally, (5.14d) shows the three-dimensional map as if facing the observations from the bottom right corner of Figure 5.14b or 5.2.¹⁴

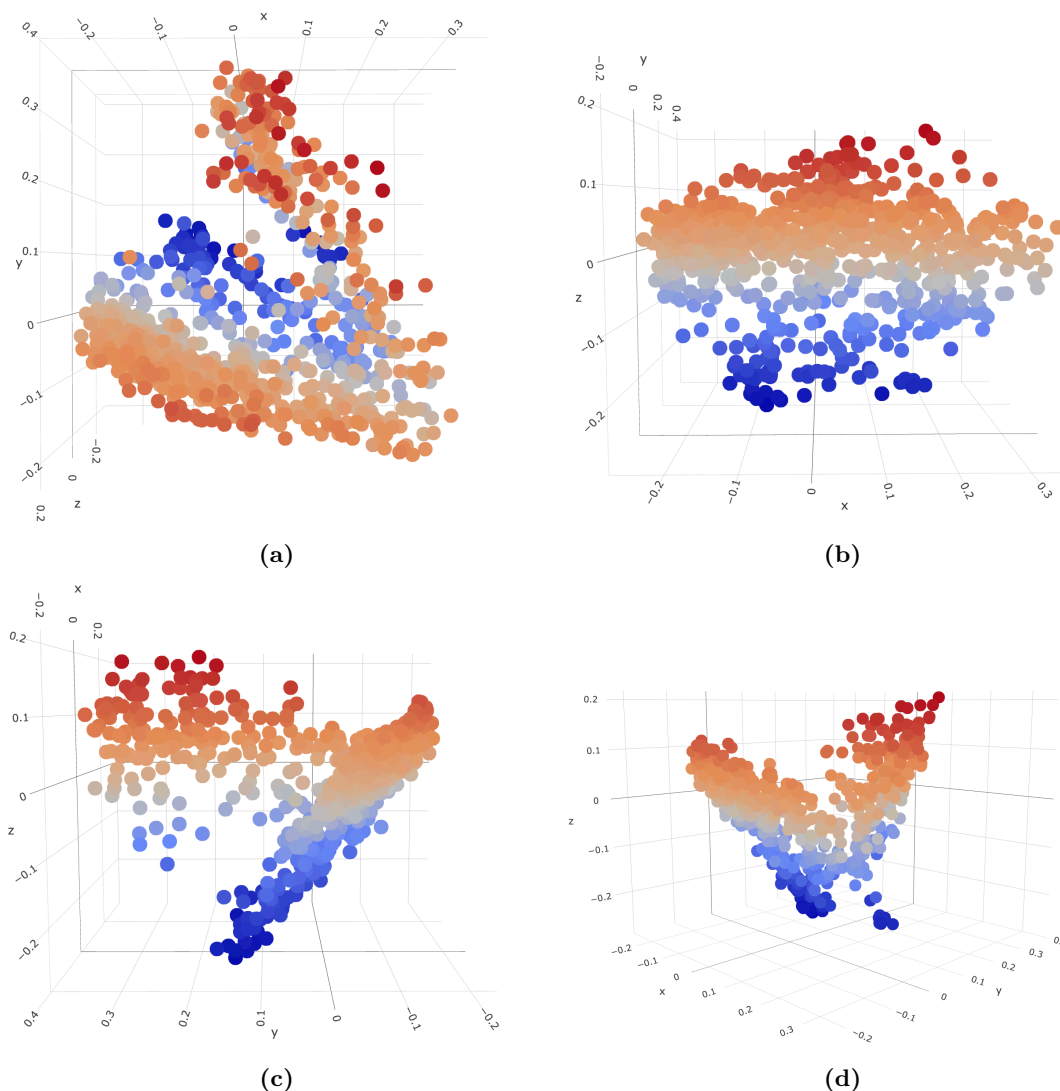


Figure 5.14: Three-dimensional unlabelled MDS map from the perspective of different axes

Abstracting from the variation observed on the left side of all the maps, we can identify a main three-way split in the possible patterns of lexification (Figure 5.15). Let us informally refer to the three areas in this split as TL (top-left), ML (mid-left), and BL (bottom-left).

We can speculate that area BL could actually also contain an area BL-2, since that is a configuration found in several languages, as Figures 5.13a and 5.13b showed.

On the basis of a two-dimensional map, it seems safe to use the three-way split shown in Figure 5.15 as a template that is potentially applicable to roughly the left half of the map for most of the languages in the corpus. The cores of the endpoints of

¹⁴An interactive version of the 3D plot can be found in this Observable notebook: <https://observablehq.com/@npedrazzini/3d-mds-when-map>.

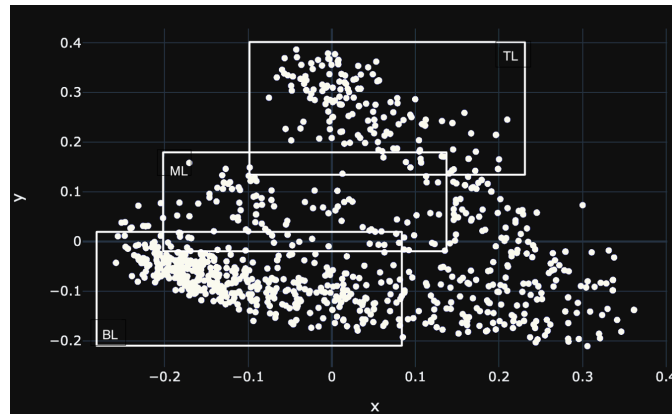


Figure 5.15: Potential three-way split of non-null situations.

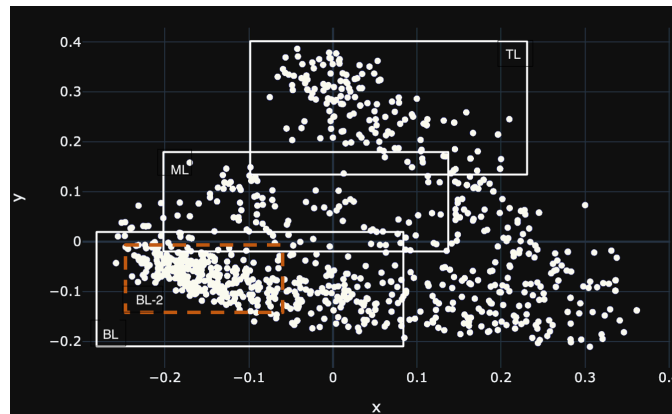


Figure 5.16: Potential four-way split of non-null situations.

these three areas approximately correspond to the GMM cluster centroids identified in Figure 5.7 above. Virtually all languages possessing non-null areas seem to group the three clusters in Figure 5.15 together following (approximately) one of their boundaries, either individually or by merging two or all three of them. The rightmost area (Figure 5.17) is instead always predominantly occupied by ‘null’ situations across all languages.

Looking more closely at the attested configurations to the left of the maps, we observe several different patterns of colexification for areas TL, ML, and BL (following the lettering from Figure 5.15). Using the points within GMM clusters 3, 2 and 5 identified by the balltree approach as the core points for areas TL, ML, and BL, respectively, following the method described in Section 5.3, we can now automatically classify languages based on how their Kriging maps fit those GMM clusters.

5.4.2 Classification

In the majority of languages (1165 out of 1452), there is one Kriging area that is the best correspondence to each GMM area. This does not mean that the *same* Kriging area

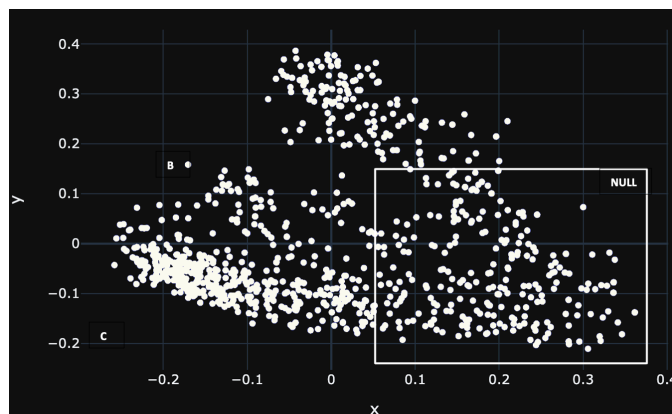


Figure 5.17: Approximate null area across all target languages

cannot span two or all three of the GMM clusters under consideration, but only that the competition around the centroid of each of those clusters is low enough for 1165/1452 languages to have one predominant means. For such languages, then, there are five logically possible colexification patterns. Table 5.5 shows these with their frequencies. Note that a null Kriging area is considered as a separate means in its own right. The table thus also indicates how many languages (‘freq null’) have a null Kriging area rather than a lexified one for any of the three clusters (‘null in’ indicates the cluster which the counts in ‘freq null’ refer to).

	pattern	freq	null in	freq null
A	TL = ML = BL	636	all	250
B	(TL = ML) \neq BL	146	TL,ML BL	23 84
C	TL \neq (ML = BL)	198	TL ML,BL	14 53
D	TL \neq ML \neq BL	110	TL ML BL	6 21 29
E	(TL = BL) \neq ML	75	ML TL,BL	6 47

Table 5.5: Frequency of lexification patterns

An additional 222 languages have significant competition with at least one more Kriging area within one of the three GMM clusters, but these could be subsumed to one of the five main patterns in Table 5.5. If, for example, the TL area has two different competing Kriging areas, while the ML and BL areas are together under the scope of yet another Kriging area, then the main pattern could still be considered C (TL (ML = BL)), even though the TL area has two, rather than one prominent means. The

updated frequencies with the addition of these 222 languages to the respective patterns are shown in Table 5.6.¹⁵

	pattern	freq	null in	freq null
A	TL = ML = BL	639	all	250
B	(TL = ML) ≠ BL	171	TL,ML BL	24 84
C	TL ≠ (ML = BL)	277	TL ML,BL	14 59
D	TL ≠ ML ≠ BL	195	TL ML BL	9 26 40
E	(TL = BL) ≠ ML	105	ML TL,BL	6 47

Table 5.6: Frequency of lexification patterns, including subpatterns within a main pattern

40 languages in the dataset have at least one GMM cluster in which there is no Kriging area (i.e. there is not one particular means that is significantly more prominent than others), so we are not able to assign them to any of the five main patterns. This is the case, for example, of Hawaiian (Figure 5.18a), Djambarrpuyngu (Figure 5.18b), Poqomchi' (Figure 5.18c), and Somali (Figure 5.18d).

A small number of languages (17) have at least one main Kriging area per GMM cluster, but their pattern cannot be easily subsumed to any of the 5 main patterns. This is the case, for example, in Bete-Bendi (Figure 5.19a) and Sepik Iwam (Figure 5.19b).

Using Table 5.6 as our baseline, we see that the most common pattern among the world's languages is A, where one main Kriging area encompasses all three GMM clusters, as is the case of Serbian and Moose Cree seen above (Figures 5.10a and 5.10b). For some of these languages, such as Mi'kmaq (Figure 5.20a) and most Romance languages (e.g. Italian; Figure 5.20b), the competition in the BL area is more evident from the Kriging map despite being assigned to pattern A. Note, for example, how the scope of *quando* 'when' in Italian does not cover the whole BL area (and similarly in the other Romance languages). For 250 out of the 639 languages (around 40%) with pattern A, the Kriging area is a null one. Given the higher likelihood of using null constructions on the right-hand side of the map, it is very likely that, for these languages, this is the case for the whole map. Examples of the latter are the maps of Southern Nambikuára and Southern Altai seen above (Figures 5.8a and 5.8b).

The least common pattern is E, consisting of the colexification of the TL and BL areas, with area ML under the scope of a different lexical item. The relative infrequency of this pattern is expected, given that the TL and BL areas are not contiguous in the semantic map. Languages with this pattern are, for example, Nobonob (Figure 5.21a),

¹⁵I have included both classifications for all languages in the Appendix, indicating the main pattern and the subpattern (if any), as well as a breakdown of how each subpattern was assigned.

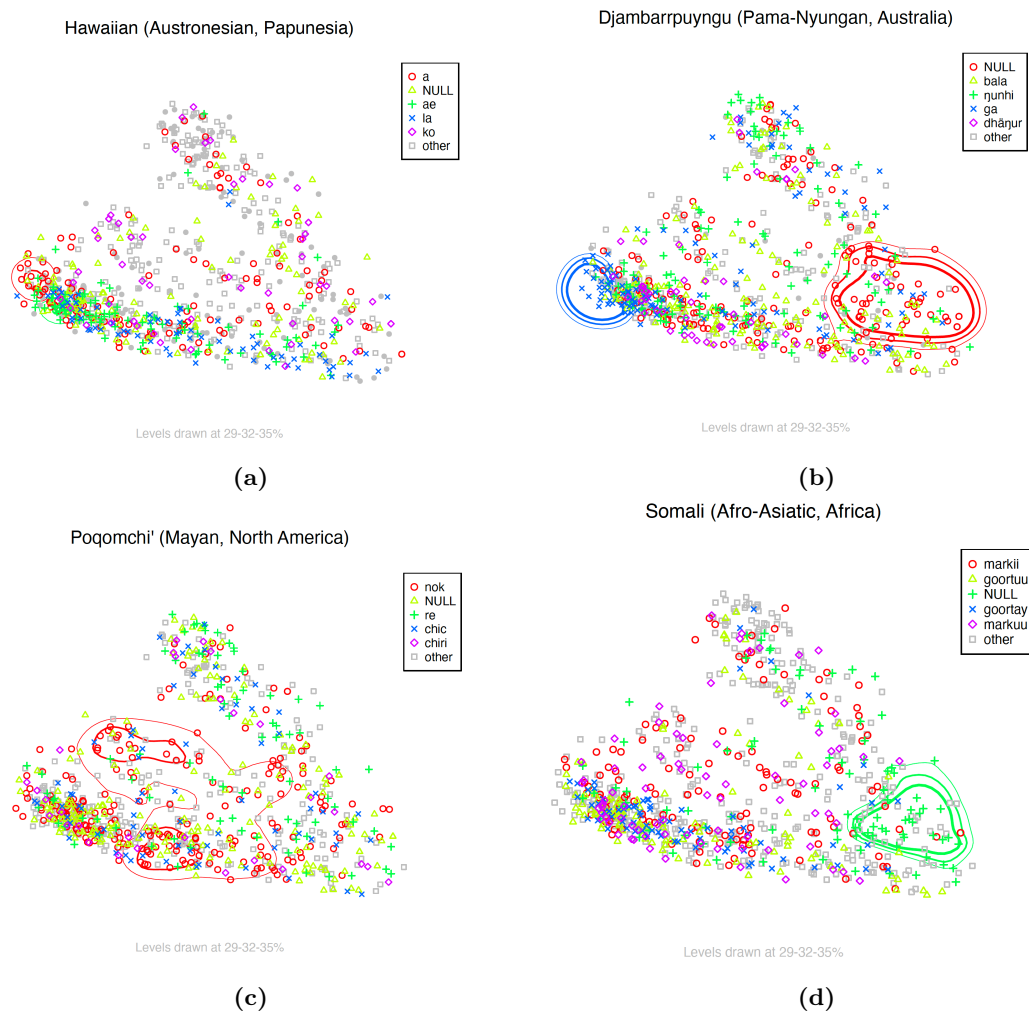


Figure 5.18: Kriging maps for Hawaiian (Austronesian, Papunesia), Djambarrpuyngu (Pama-Nyungan, Australia), Poqomchi' (Mayan, North America), Somali (Afro-Asiatic, Africa)

where TL and BL are expressed by null constructions, and Eastern Oromo (Figure 5.21b), where TL and BL are colexified.

In patterns B, C and D, the TL and BL clusters are under the scope of different Kriging areas, and they differ in whether the ML area is lexified separately or together with either the TL cluster or the BL cluster.

Pattern-D languages have a separate Kriging area for the ML cluster, as in Tuvuli (Figure 5.22a) and Kako (Figure 5.22b).

Similarly to what we noticed with A-pattern languages, among pattern D-languages, the scope of the ML Kriging area is quite varied. For several of these, the ML area has scope on part of the BL area as well, where competition among different means is, as already observed, generally more intense cross-linguistically. Ancient Greek (Figure

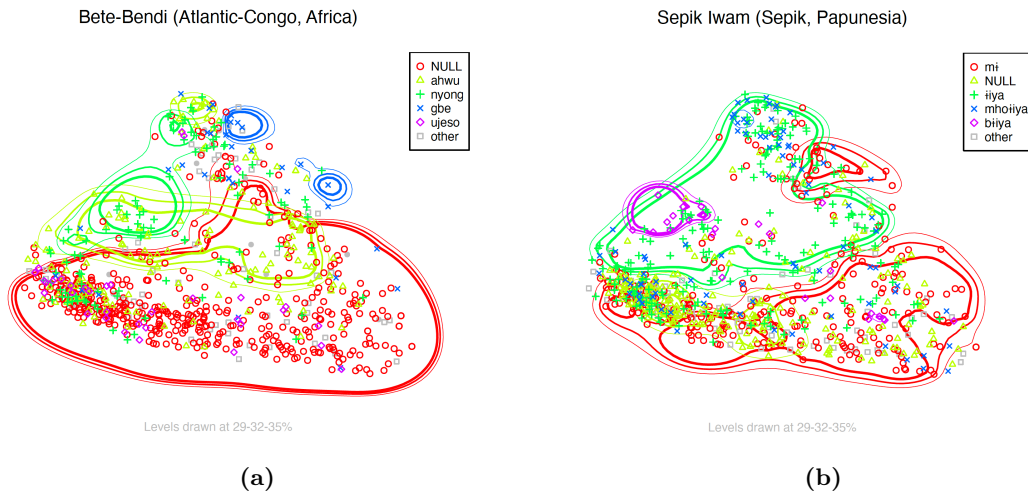


Figure 5.19: Kriging maps for Bete-Bendi (Atlantic-Congo, Africa) and Sepik Iwam (Sepik, Papunesia)

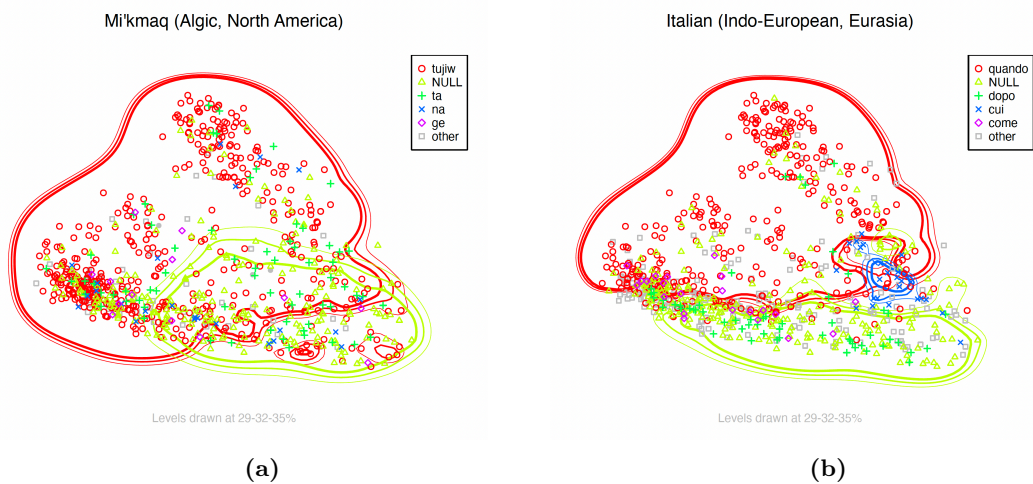


Figure 5.20: Kriging maps for Mi'kmaq (Algic, North America) and Italian (Indo-European, Eurasia)

5.23a) is one such language, as well as, for example, Wandala (Figure 5.23b), Chumburung (Figure 5.23c), and Tii (Figure 5.23d), among several others.

Patterns B and C are, in a way, asymmetrical. That is, they differ in whether the ML area is colexified with the TL or the BL area. In pattern-B languages, the TL area is colexified with the ML area. Old Church Slavonic (Figure 5.24a) belongs to this pattern, with the TL and ML areas having *jegda* as the predominant Kriging area, while the BL area is dominated by null constructions.¹⁶ Other examples are Eigham (Figure 5.24b),

¹⁶At the top-right of the Old Church Slavonic map, we can see a small Kriging area (13 occurrences) for *kogda*, which is ‘when’ as a question word, as in: *kogda že te viděchomъ bolešta ili vъ temъnici i pridomъ kъ tebě* ‘and when did we see you sick or in prison and visit you?’ (Matthew 25:39). *When* as a question

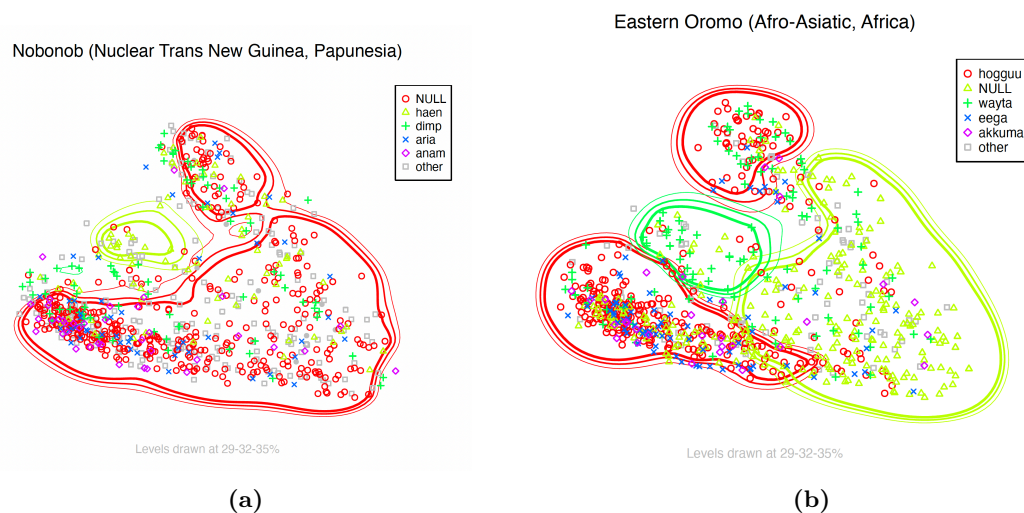


Figure 5.21: Kriging maps for Nobonob (Nuclear Trans New Guinea, Papunesia) and Eastern Oromo (Afro-Asiatic, Africa)

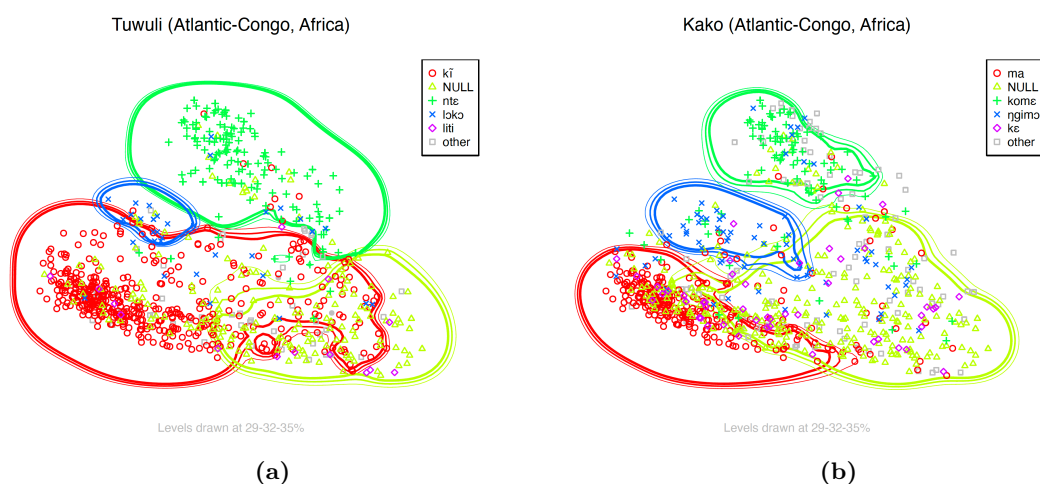


Figure 5.22: Kriging maps for Tuwuli (Atlantic-Congo, Africa) and Kako (Atlantic-Congo, Africa)

Xavánte (Figure 5.24c), and Goan Konkani (Figure 5.24d).

Patter C is the second most-frequent pattern among the world's languages. Some examples from this very common pattern are Lomeriano-Ignaciano Chiquitano (Figure 5.25a), Kiribati (Figure 5.25b) and Siwu (Figure 5.25d). It also includes most Germanic languages, for example Afrikaans (Figure 5.25c).

Like the Germanic languages in this group, pattern-C languages are likely to have one means for expressing the meaning of WHEN to refer to repeated events in past, present or future, as well as for singular events in the future, and one means for expressing

word is not analyzed separately in this thesis.

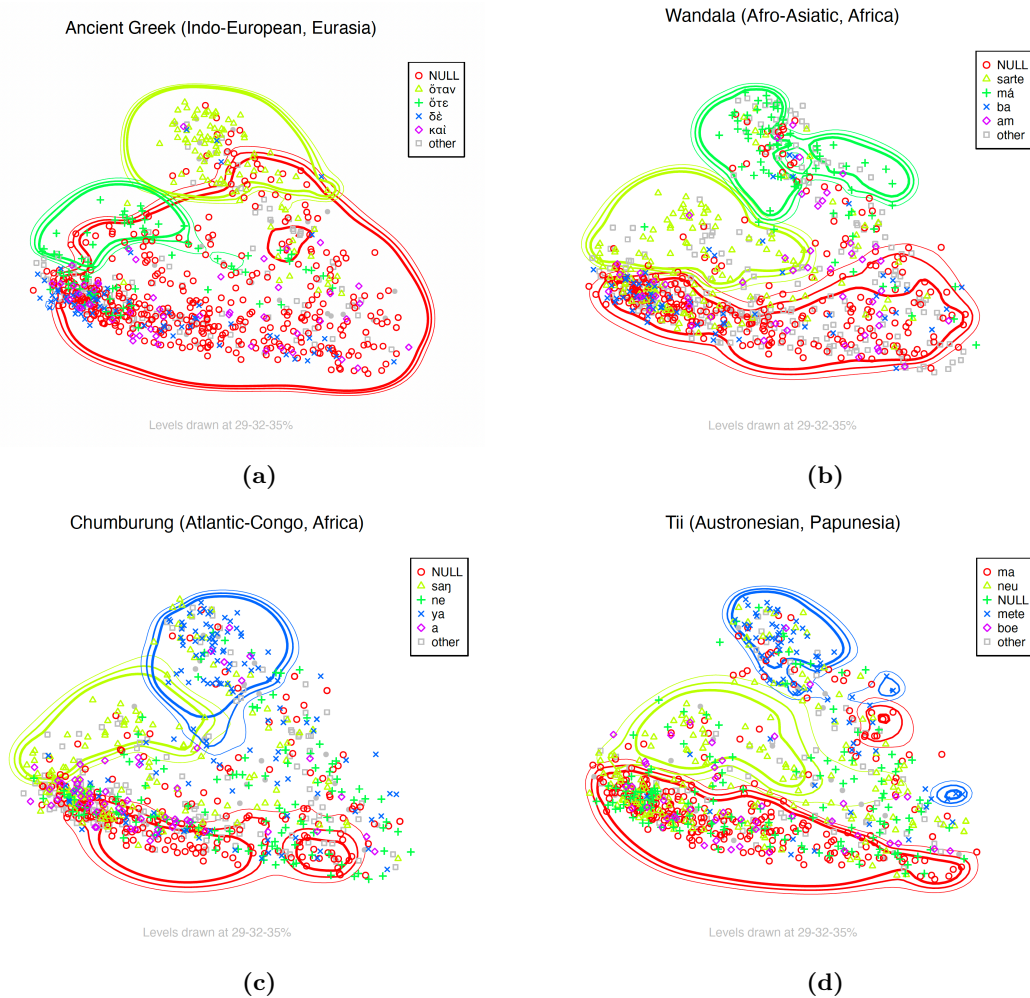


Figure 5.23: Kriging maps for Ancient Greek (Indo-European, Eurasia), Wandala (Afro-Asiatic, Africa), Chumburung (Atlantic-Congo, Africa), Tii (Austronesian, Papunesia)

singular events in the past. In other words, they have one means for *universal* WHEN irrespective of tense, which is also used for *existential* WHEN in the future, and one means for *existential* WHEN in the past (cf. Sæbø 2011b; Fabricius-Hansen & Sæbø 1983: 2). This contrast can be exemplified by the minimal pair in the German example in (11).

- (11) a. **Als** *ich ins Bett ging, konnte ich nicht einschlafen*
 when.EXIST I in bed went, could I not sleep
- b. **Wenn** *ich ins Bett ging, konnte ich nicht einschlafen*
 when.UNIV I in bed went, could I not sleep

In the Afrikaans example in Figure 5.25c, the TL area is under the scope of *wanneer* ‘when’ (universal and nonpast existential), whereas the ML/BL area is under the scope

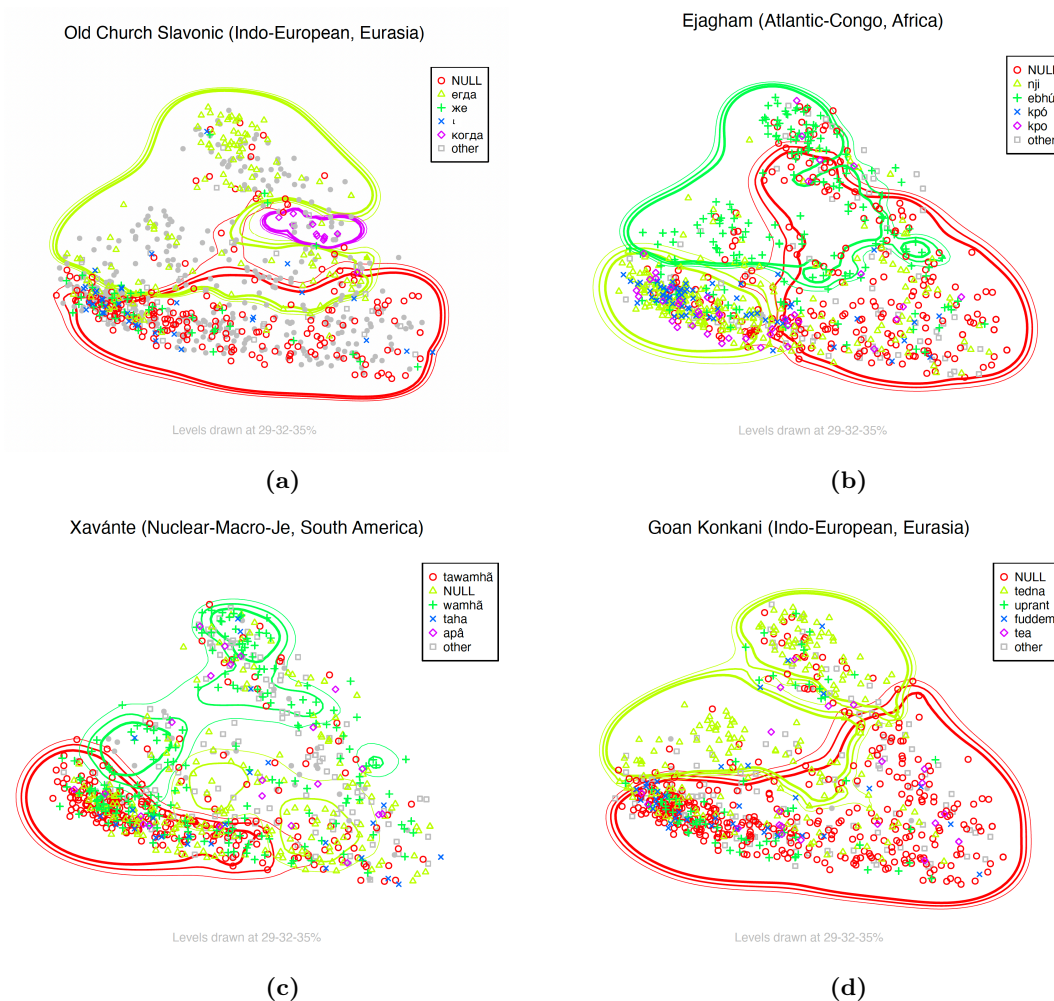


Figure 5.24: Kriging maps for Old Church Slavonic (Indo-European, Eurasia), Eigham (Atlantic-Congo, Africa), Xavante (Nuclear-Macro-Je, South America), Goan Konkani (Indo-European, Eurasia)

of *toe* ‘when’ (past existential).

It is relatively uncommon for pattern C languages to have a null Kriging area in the TL cluster (59 out of 277 languages), compared to pattern-B languages, where that is the case in about half of them (84 out of 171 languages).

Overall, it seems exceedingly rare for languages with only one Kriging area besides nulls to use it exclusively to encode the situations in the BL cluster. On the other hand, it is very common to find languages with a dedicated Kriging area for the TL cluster only. As a hypothetical generalization, if languages make extensive use of null constructions and if they have one subjunctive that is equivalent to *when*, then it seems very likely that it will be used most typically for the situations in the TL area without much competition with null constructions. This is what we observe in the Old Church

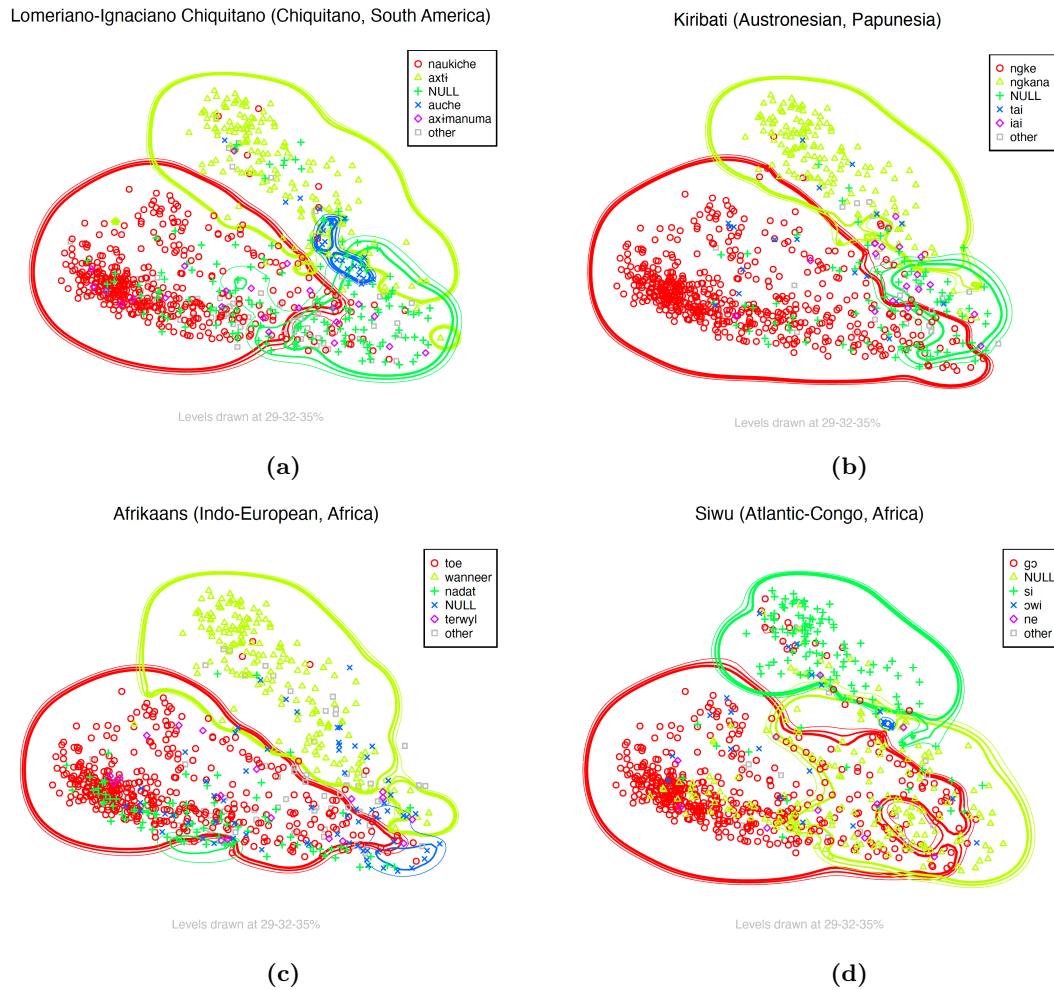


Figure 5.25: Kriging maps for Lomeriano-Ignaciano Chiquitano (Chiquitano, South America), Kiribati (Austronesian, Papunesia), Afrikaans (Indo-European, Africa), Siwu (Atlantic-Congo, Africa)

Slavonic and, albeit with two rather than one lexified means, in the Ancient Greek map. The observations in the TL area appear to be more clearly under the domain of the *when*-equivalents available in the language (i.e. *jegda* and *hótan*) than the ML and BL areas are. In Ancient Greek and Old Church Slavonic, the Kriging area for *hóte/jegda* and the one for null constructions show a high degree of overlap, suggesting that the competition between finite temporal subordinators and participle constructions, which we can expect to be predominant within their null areas, is particularly clear in between the ML and the BL areas.

5.4.3 Overall precision and recall for the GMM clusters

As explained in the Methods section, we can look for potential cross-linguistic lexification patterns by checking how well each Kriging area in each language fit each of the GMM cluster. Recall that, to do so, we count how many observations within a Kriging area are found inside a given GMM cluster, how many outside of the same GMM cluster, as well as how many observations not from other Kriging areas (or not belonging to any Kriging area) are found inside that same GMM cluster. These numbers are then used as true positives, true negatives and false positives, respectively, which we can then use to calculate precision and recall for each Kriging area. The highest-scoring Kriging area is then considered the best match for that GMM cluster in the language. That is, if a language has one Kriging area yielding a high harmonic mean (F1 score) with respect to a particular GMM cluster, that will indicate that it has a particular means which is relatively specialized for whichever meaning or function of *WHEN* that GMM cluster corresponds to. The higher the F1 score, the more specialized the means.

The results of this method are shown in Figure 5.26. Note that a high precision score will not necessarily correspond to a likely lexification of that GMM cluster if the recall is very low (i.e. close to 0)—most likely, these are rare items (possibly due to misalignment by the model), most or all of which happen to be distributed inside the GMM cluster under consideration. High recall with low precision is instead likely to indicate that the scope of the means in question is broader than the GMM cluster under consideration. English *when*, for example, will show a recall of 1 for all clusters, since the *whole* semantic map is under its scope. What we are looking for is, therefore, both high precision and high recall (i.e. a high F1 score).

The potential generalization regarding common patterns of usage of overtly subordinated constructions as opposed to null constructions finds some degree of confirmation in the different F1 scores for the different GMM clusters.

The results for GMM clusters 1, 5 and 6 do not immediately suggest any pattern. Languages in these plots cluster quite densely, with little variation between languages, and both mean precision and recall, indicated in the plots by a red marker *x*, are low. Both the lack of visible differences between languages and the very low average precision and recall suggest that our method is not capturing cross-linguistic differences nor a clearly well-defined function from the observations in this cluster.

Cluster 2, which corresponds to the left extremity of the ML cluster commented on in the previous section (cf. Figure 5.15), also has both low mean precision and recall, but this time we see a handful of languages with high recall (≥ 0.6) and slightly higher precision (≥ 0.4) than most other languages with the same recall. A closer look at some of them confirms that these are mostly languages with a relatively well-defined Kriging area for the ML cluster, clearer than languages such as Ancient Greek (Figure 5.23a) or Chumburung (Figure 5.23c), for example, which also have a separate Kriging

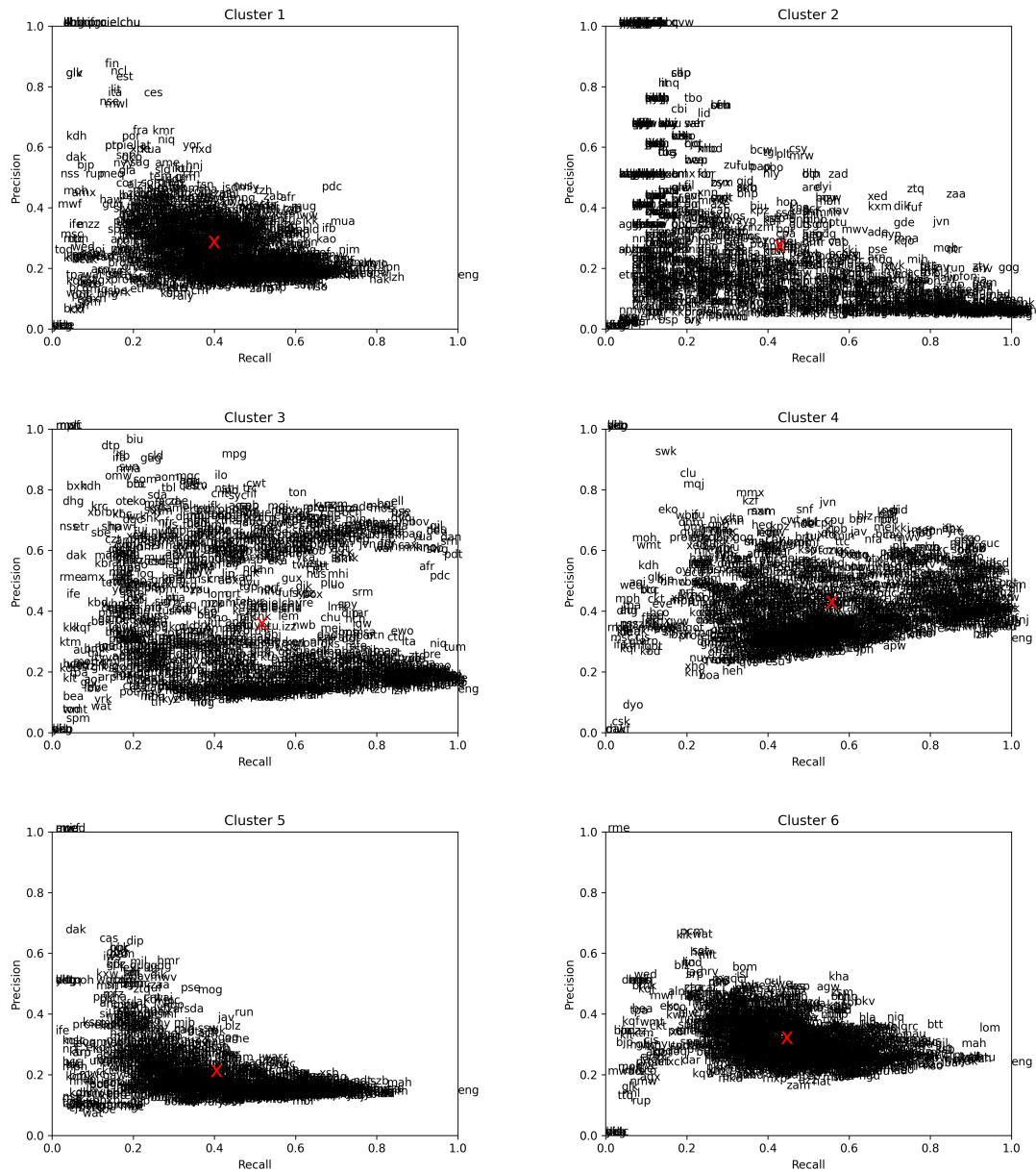


Figure 5.26: Precision and recall plots for the six GMM clusters shown in Figure 5.6. The x and y values (recall and precision, respectively) for each language in each of the subplots correspond to the recall and precision of the item with the highest F1 score in that language for that cluster (compared to all other items in that language occurring at least once within that cluster). The red marker x indicates the mean precision and recall value.

area for the ML cluster but perhaps more intense competition with other means, or a wider scope extending to the BL area as well (e.g. as we saw in the Ancient Greek map, among others). Compare the clearer boundaries of the Kriging area for the ML cluster in Figure 5.27a and 5.27b, for Pular and Sierra de Juárez Zapotec, with those

in Figures 5.23a-5.23d seen earlier on.

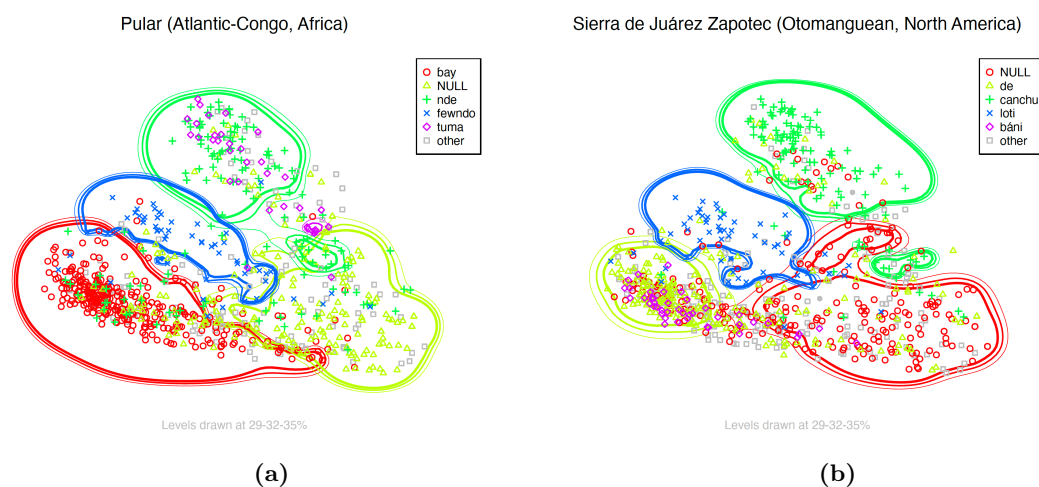


Figure 5.27: Kriging maps for Pular (Atlantic-Congo, Africa) and Sierra de Juárez Zapotec (Otomanguean, North America)

The clearest pattern comes from the plot for GMM cluster 3, corresponding to the TL area commented on in the previous section. Languages divide into two bands, one with lower precision (between 0.1 and 0.3) and one with higher precision (≥ 0.5), both stretching across the whole range of recall from 0 to 1. The clear pattern for this cluster suggests that GMM cluster 3 is likely to approximate a real gram type and that it is very common for this gram type to be expressed by some specialized means cross-linguistically.

Figure 5.28 shows the best matches to cluster 3, with a precision ranging between 0.5 to 0.75, and a recall between 0.6 all the way to nearly 1.

We see that some of the best matches are found, among others, across a range of Germanic languages, in particular Danish [dan] and Norwegian [nob/nno] *når*, and German [deu] *wenn* (also cf. Kiribati [gil], represented in Figure 5.25b as well). As we mentioned above, these languages have lexified the distinction between universal and existential readings in the past, with one subordinator generally serving for universal readings in any tense, as well as nonpast existential meanings, and another subordinator for past existential readings only.

Another interesting pattern is the one obtained from GMM cluster 4 (i.e. the BL cluster). Here we do not get any clear division into bands among the world's languages. Instead, most languages show overall relatively high precision, but few as high as the upper band or as low as the lower band observed for cluster 3. Both mean precision and recall are, in fact, slightly higher than the one for cluster 3. On the one hand, this suggests that, with this method, we are not capturing cross-linguistic differences in the way languages express the situations in cluster 4, as is the case with cluster 3. On the other hand, however, it also points to the fact that it is very common for individual languages to express the situations in this area with one main means, but even in the

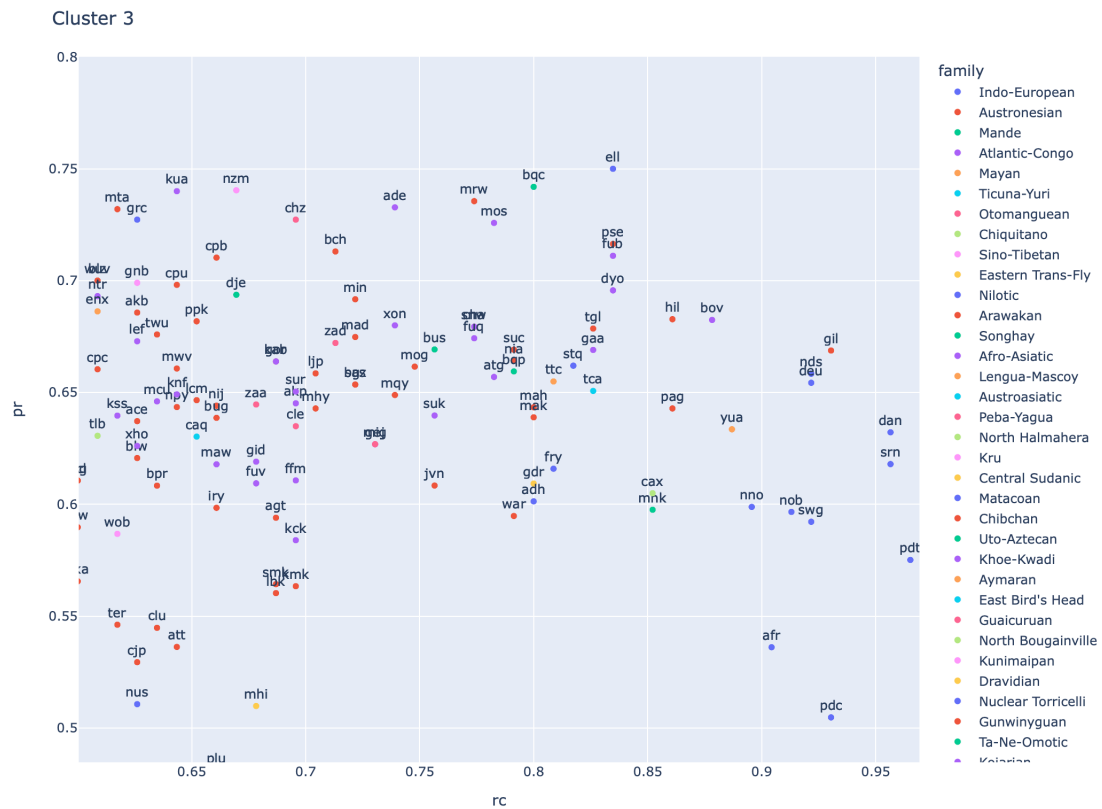


Figure 5.28: The top-scoring languages (by precision and recall) for GMM cluster 3

presence of such means, there might be more competition with other constructions. Given what we have seen above regarding the distribution of null constructions across the world's languages, namely that they are more and more likely to occur the more we look at the situations to the bottom right of the map, a possibility is that cluster 4 is more likely to have null constructions as competitors to some other means.

Figure 5.29 shows the best matches to cluster 4, with a precision ranging between 0.45 to 0.7, and a recall between 0.7 all the way to nearly 1. Languages at the higher range of precision have quite well-defined Kriging areas for the BL area, with relatively limited competition, considering some of the more extreme cases seen earlier on (cf. Sierra de Juárez Zapotek, Figure 5.27b, as well as all the maps in Figures 5.23a-5.23d and Figures 5.24a-5.24d). Some of these examples are, once again, Pular (Figure 5.27a), as well as, for example, Batad Ifugao (Figure 5.30a) and Konobo-Eastern Krahn (Figure 5.30b).

At the lower precision range (0.45-0.5) from Figure 5.29, we find languages that have some relatively well-defined Kriging area for cluster 4, but much more intense competition with other constructions. Some examples are Nawuri (Figure 5.31a) and Tagalog (Figure 5.31b).

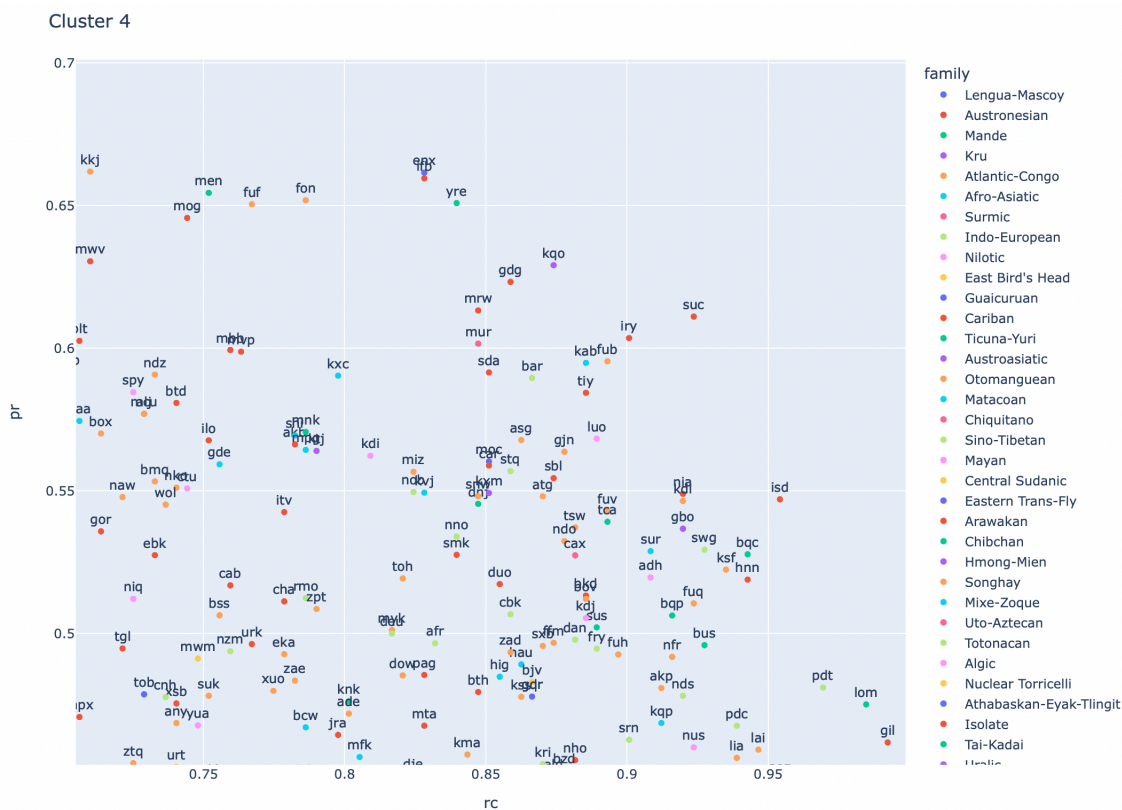


Figure 5.29: The top-scoring languages (by precision and recall) for GMM cluster 4

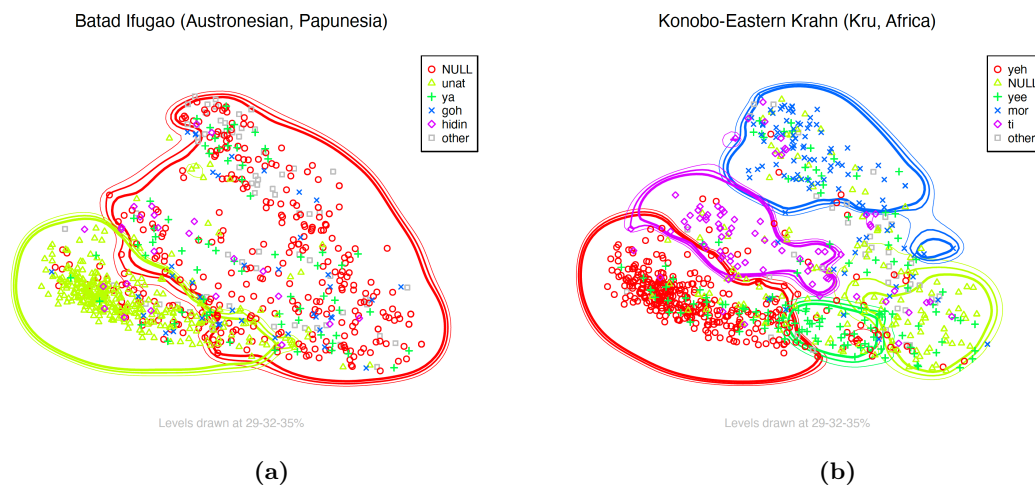


Figure 5.30: Kriging maps for Batad Ifugao (Austronesian, Papunesia) and Konobo-Eastern Krahn (Kru, Africa)

At the higher recall range from Figure 5.29, we find several languages that have a relatively good match for cluster 4, but the same Kriging area covers more ground, typically including cluster 2 as well (i.e. the left extremity of the ML cluster). Most

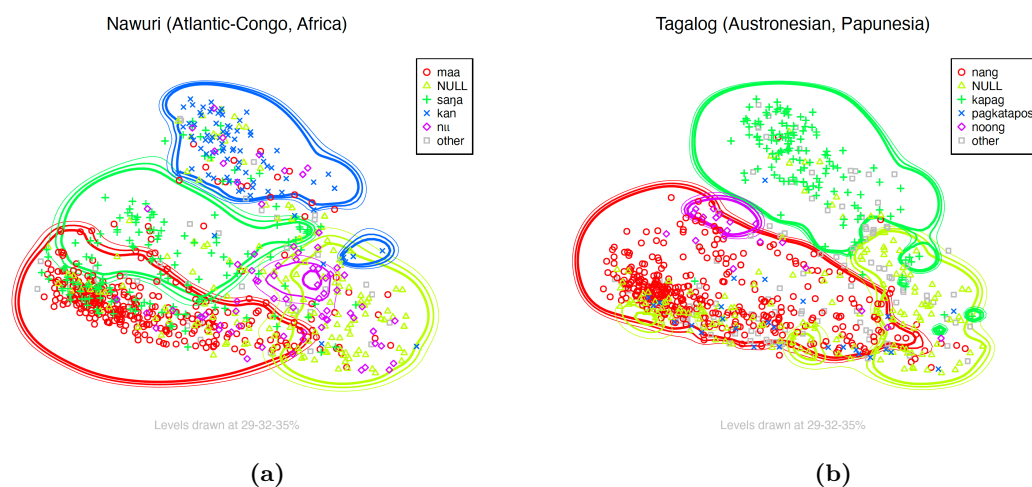


Figure 5.31: Kriging maps for Nawuri (Atlantic-Congo, Africa) and Tagalog (Austronesian, Papunesia)

Germanic languages are one such example, like Afrikaans (Figure 5.25c; *recall* = 0.83) and German (Figure 5.32a; *recall* = 0.81), or Adhola [adh] (Figure 5.32b; *recall* = 0.90).

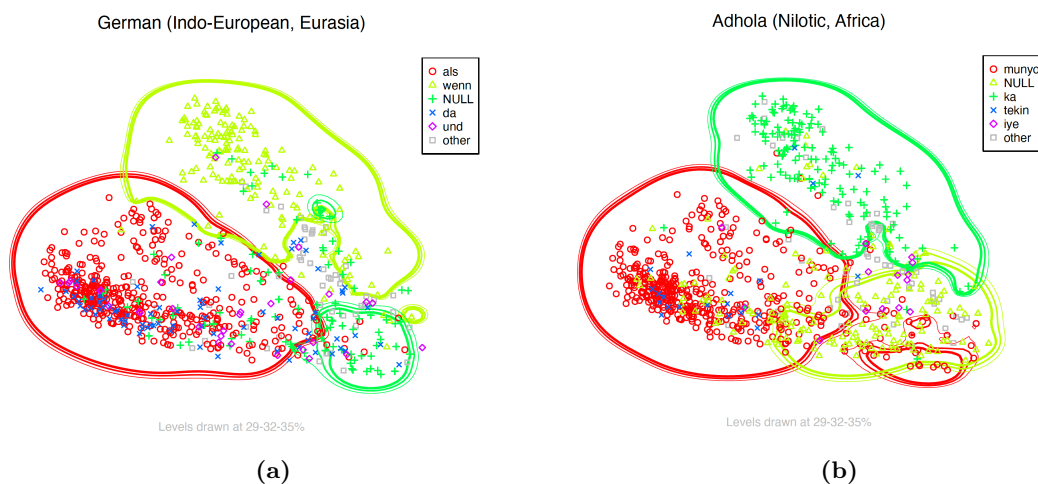


Figure 5.32: Kriging maps for German (Indo-European, Eurasia) and Adhola (Nilotic, Africa)

In Section 5.4.2, we saw that the observations in the TL area (identified on a close-reading basis, as shown in Figure 5.15) in the Old Church Slavonic and Ancient Greek maps appeared to be more neatly under the domain of the *when*-equivalents available in the language (i.e. *jegda* and *hótan*) than those in the ML area and even less so in the BL area, where finite subordinators are instead clearly present but more heavily competing with null constructions, as the Kriging results also indicated. The temporal usage of *wenn*, then, as exemplified in (11b), may approximate some of the functions of WHEN that are, overall, the *least* likely to be rendered with null constructions in the two languages.

Since we do not have access to a comparative concept (Haspelmath, 2010) that could tell us whether to include a particular usage point in a gram type, a possible route to finding a core or prototypical meaning for a particular GMM cluster could be to count how many languages express each observation in that cluster using their top-ranking lexical item (by F1 score, as described above) for that cluster. For instance, the top-ranking item for GMM cluster 3 for German, Old Church Slavonic, and Italian are *wenn*, *jegda*, and *quando*. If an observation from inside cluster 3 is expressed by *wenn*, *jegda*, and *quando*, its ‘prototypicality’ score would be 3; if an observation from cluster 3 is expressed by *wenn* and *jegda* but not *quando*, then the score would be 2; and so on for each sentence from cluster 3. After scoring each data point inside cluster 3 for prototypicality, the highest-ranking data points may then best approximate the prototypical usage of the gram type, if any, corresponding to that cluster.

Among the prototypical examples for cluster 3, we find both universal *when* in the present tense 12 and existential *when* in the future tense 13.

- (12) And he said to them, “**When you pray**, say: “Father, hallowed be your name. Your kingdom come. (Luke 11:2)
- (13) And he said, “Jesus, remember me **when you come** into your kingdom.” (Luke 23:42)

This shows that the colexification of existential *when* in the future with the universal *when*, which could appear to be an idiosyncrasy of German, is actually found across the languages that make some distinction between existential and universal **when**. Both these examples are expressed by *jegda*-clauses in Old Church Slavonic and *hótan*-clauses in Ancient Greek and are quite representative of the overall type of temporal relations found among the sentences in this cluster.

When we extract the most prototypical sentences from cluster 2, we see that the top-ranking sentences are all stative when-clauses of the type ‘when I was a child’, ‘when you were young’, ‘when we were with you’, ‘when you were slaves’, ‘when you were pagans’, and similar. Note that these stative clauses are in fact *existential*, not *universal*, when-clauses, since there is effectively only one eventuality of the type, for example, *I be child* or *you be pagans*. The matrix clause of these *when*-clauses, however, seem to be mostly repeated occurrences of an eventuality, as in (14) or (15).

- (14) **When you were** slaves to sin, you were free from the control of righteousness (Romans 6:20)
- (15) You know that **when you were** pagans you were led astray to mute idols, however you were led (1 Corinthians 12:2)

When the matrix clause is stative instead, then it seems more common for *when*-clauses in this cluster to be eventive, as in (16).

- (16) Now Thomas, one of the twelve, called the Twin, was not with them **when Jesus came** (John 20:24)

Note that this example is not fundamentally at odds with those in (14) and (15) from the temporal-relation perspective. In both cases, the *when*-clause works as an existential determiner over times. Since one of the two eventualities is stative, the temporal relation between the *when*-clause eventuality and the matrix eventuality is symmetrical, in the sense that, even if the order in which they are presented is swapped, their temporal interpretation remains very similar and what may change is rather their information-structural status and discourse relation, as observed, among others, by Sæbø (2011b: 3) and as we discussed in Chapter 1.

Finally, a prototypical example from cluster 3 involves two eventive clauses, as in (17a)-(17c)

- (17) a. But **when Jesus heard** it he said, “This illness does not lead to death. It is for the glory of God, so that the Son of God may be glorified through it.” (John 11:4)
- b. **When Jesus saw** her, he called her over and said to her, “Woman, you are freed from your disability.” (Luke 13:12)
- c. So **when Martha heard** that Jesus was coming, she went and met him, but Mary remained seated in the house (John 11:20)

Old Church Slavonic translates (17a)-(17b) with a conjunct participle and (17c) with a *jegda*-clause. Intuitively, all three sentences involve the same temporal relation between clauses, namely one of temporal *abutment*, whereby the matrix event, in this case, immediately follows the event of the adverbial clause. What seems to differ between (17c) and the other two examples is the fact that the matrix clause in (17c) continues with an (adversatively) coordinated clause with a different subject. Among the languages in which null constructions are predominant in GMM cluster 4, while clusters 2 and 3 are lexified (i.e. pattern-B or -D languages in our classification), we find that languages with converbs (or converb-like forms) or known for allowing serial constructions are particularly frequent.¹⁷ Among pattern-D languages, for instance, we find numerous West African languages, where extensive use of serial verb constructions is a well-known prominent feature (cf. Givón 2015; Stahlke 1970; Awoyale 1987; Bamgboṣe 1974; Lord 1973), as well as Yabem (cf. Bisang 1995a) and several other Austronesian and Papuan languages (cf. Senft 2004; Conrad & Wogiga 1991), also oft-cited for their use of verb

¹⁷see what reviewer said about this

serialization. Among pattern-B languages, we find several North and South American language families, such as Arawakan, Aymaran, Chibchan and Tupian, all of which have also been studied with respect to their use of serial constructions (cf. Aikhenvald & Muysken 2010). Languages that make extensive use of converbs, such as Korean and Kumyk, or predicative participles functionally very similar to converbs, such as Ancient Greek, are also among pattern-B or -D languages in which null constructions are predominant in GMM cluster 4.

The intuition is that the situations found at the bottom half of the semantic map are more likely to be found as part of sequential events, which can be expressed with serial verb constructions or clause chaining by the languages where these are possible.

As already discussed, the way in which the data was sampled (i.e. starting from a single, albeit relatively underspecified, lexified means, namely English *when*) does not allow us to say much about cross-linguistic correspondences between different types of null constructions using the semantic map without further layers of annotation. What is important to highlight here, however, is that both the analysis of the Kriging maps and the precision and recall analysis of the GMM clusters indicate that there are clear cross-linguistic patterns as to which situations are more likely to be encoded by overtly subordinated, finite *when*-clauses and which are more likely expressed with juxtaposed constructions. This means that the division of labour between *jegda*-clauses and participle constructions in Early Slavic is likely to follow broader principles followed by many languages cross-linguistically.

5.5 Summary

In this chapter, the semantic space of the temporal connective *when* was explored by leveraging a massively parallel corpus of Bible translations and starting from the semantically underspecified subordinator *when* in English. The methods presented in the chapter provided tentative, but promising results, particularly thanks to clear evidence for a cross-linguistic gram type expressing *universal* WHEN.

Another clear finding is that null constructions, including Early Slavic participle constructions, cluster in particular regions of the semantic map. This means that they are not equally viable as alternatives to any use of WHEN, but carry particular meanings that make them less suitable for some functions of WHEN. The way the data was sampled so far, though ideal for the exploration of lexified WHEN-clauses, was not well suited to further investigations in this area, but rather to single out the broad domains where overtly subordinated *when*-clauses and null constructions are respectively predominant. What we observed is that the former have the lowest competition with null constructions in correspondence to *universal* WHEN-clauses, as well as *existential* WHEN-clauses in the future, followed by existential, but stative WHEN-clauses in the past tense (i.e. GMM cluster 2 in our analysis). Competition seems stronger (both options are viable) with

two eventive eventualities, but the Kriging analysis suggested that this is most typically the domain of null constructions.

The next chapter will add layers of annotation to the semantic map for Old Church Slavonic and Ancient Greek, and look for potential finer-grained observations at the level of null constructions. Similarly, I will leverage the extensive literature and data on languages which showed similar patterns to Old Church Slavonic to corroborate the intuitions we had on the latter regarding the usage of null constructions in contexts where the sequentiality of eventualities is stressed.

Chapter 6

Early Slavic participle clauses and their typological parallels

6.1 Participle constructions in the Old Church Slavonic WHEN-map

A clear finding of the previous chapter was that null constructions cluster, cross-linguistically, in particular regions of the semantic map of WHEN, indicating that they are not equally viable as alternatives to any use of WHEN, but carry particular meanings that make them less suitable for some of its functions. We also observed that the Old Church Slavonic map for WHEN (here repeated in Figure 6.1a) belongs to a configuration (which we called pattern B) in which the top-left (TL) and middle-left (ML) areas are colexified, while area the bottom-left (BL) area forms a separate cluster where null constructions are predominant. In comparison, Ancient Greek (map repeated in Figure 6.1b), which possesses two main generic *when*-counterparts, belongs to a slightly different configuration, in which the TL and ML areas are encoded by two separate means (*hótan* and *hóte*, respectively), while the BL area is under the scope of the Kriging cluster for null constructions, as in Old Church Slavonic.

The Old Church Slavonic map contains multiple unattested observations (all the dots in grey), due to the several gaps in the Gospels from the Codex Marianus. Mayer & Cysouw's (2014) corpus contains a more recent Church Slavonic translation of the Bible, possibly a version of the *Elizavetinskaja Biblija* (Elizabeth Bible), published in 1751.¹ Since the syntax of this version seems to be faithfully close to the oldest attestations,² we can check whether the missing parallels in the Codex Marianus might have caused MDS and Kriging to skew the results. The map for the more recent Church Slavonic version is, in fact, very close to the Old Church Slavonic one from the Codex Marianus:

¹The source of the text indicated by Mayer & Cysouw (2014) does not report which translation it is based on.

²All occurrences of the dative absolute and of *jegda*-clauses found in the Codex Marianus were also found in this version after a manual check.

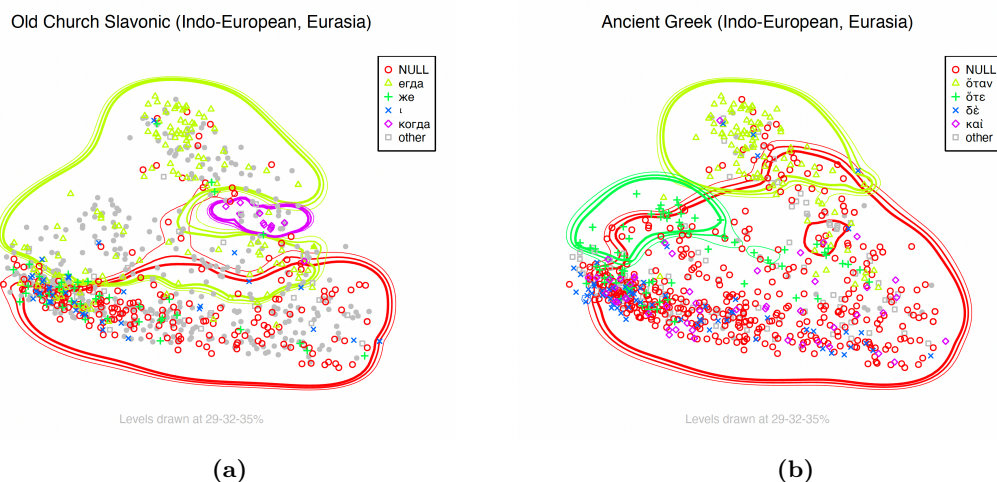


Figure 6.1: Kriging maps for Old Church Slavonic (Indo-European, Eurasia) and Ancient Greek (Indo-European, Eurasia), repeated

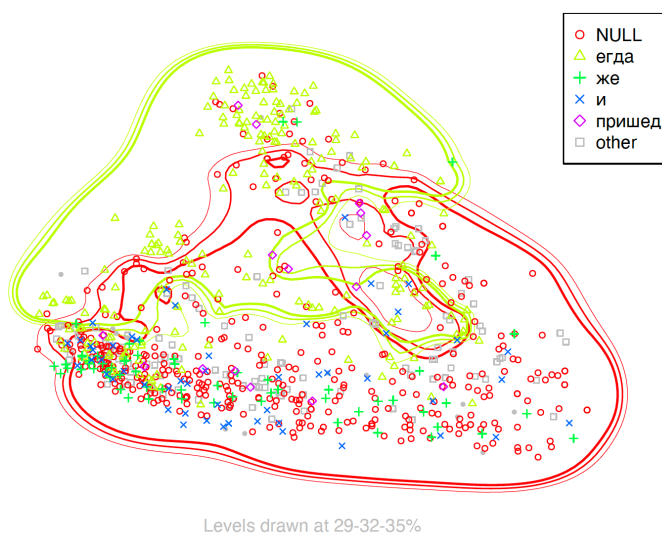


Figure 6.2: Kriging map for Later Church Slavonic

The grouping of the TL and ML areas together, as well as the predominance of null constructions in the BL area, are the same in the two maps, which suggests that, despite the missing occurrences, the results from Kriging on Old Church Slavonic are quite reliable.

As we can see from Figure 6.1a (and equally in Figure 6.2), the alignment model from English to Old Church Slavonic incorrectly identified some tokens (i.e. *že* PTC and *i* ‘and’) among the most-common *when*-parallels because they occur frequently and consistently just before certain null constructions, particularly participles. The same is true of *prišedō* ‘having arrived, when (he) arrived’, which is a very common conjunct participle throughout the Gospels. Moreover, there are occurrences of *jegda* appearing in

the less-frequent Old Church Slavonic spelling *jegda* (which is, in fact, the spelling of the dictionary form used in TOROT), which are thus not counted together with the more frequent *jegda*.

As explained in the previous chapter, the maps for WHEN were expected to be informative as far as *when* equivalents in the target languages are concerned, as well as regarding the overall split between when-situations (i.e. constructions with an overt temporal subordinator) and null-situations (i.e. generally either non-finite constructions or specific multi-word expressions not directly a counterpart of *when*), but not when it comes to distinguishing participles from other null constructions (or between conjunct participles and dative absolutes). We can thus improve the Old Church Slavonic map by re-processing the source text in the following way:

1. lemmatized all forms of *jegda* to *jegda* (out of consistency with TOROT);
2. remove *že* and *i* as stopwords;
3. insert a placeholder before the relevant participle constructions (*xadv* for conjunct participles and *absoluteadv* for absolutes), as a way of levelling out the different surface realizations of conjunct participles and dative absolutes, thus allowing the models to capture the placeholders as dummy subordinating conjunctions.

After realigning the English version to the processed Old Church Slavonic version, we are able to capture some finer-grained differences in distribution between participles and *jegda*-clauses, and between conjunct participles and absolute constructions, as Figure 6.3a shows. Despite the lemmatization of *jegda*, the area of the latter did not change substantially. The former null/*jegda* division is now a much more clear-cut conjunct participles-*jegda* split, with a smaller island of the same conjunct-participle cluster roughly between the TL and ML areas. Although dative absolutes are all mostly concentrated between conjunct participles and *jegda*-clauses, they were not detected by Kriging as a separate area in the map. This is unlike Ancient Greek, after adding the same placeholders and preprocessing the text by removing *dé* PTC and *kaí* ‘and’, as Figure 6.3b shows.

Absolutes in Greek form a Kriging area of their own, which seems to overlap almost entirely with the conjunct participle area. We should still remain wary of not over-interpreting the differences in boundaries between Old Church Slavonic and Ancient Greek, since the latter has virtually complete coverage of the New Testament, which may affect the boundaries of the Kriging areas. Nevertheless, the differences in the overall number and position of the clusters can be considered relatively reliable despite the missing occurrences in the Old Church Slavonic dataset, as Figure 6.2 from the later Church Slavonic version suggested.

In Chapter 1, we saw that, although not in a statistically significant way, dative absolutes are more frequent in the imperfective, which means that ‘while *x*-ing’ may

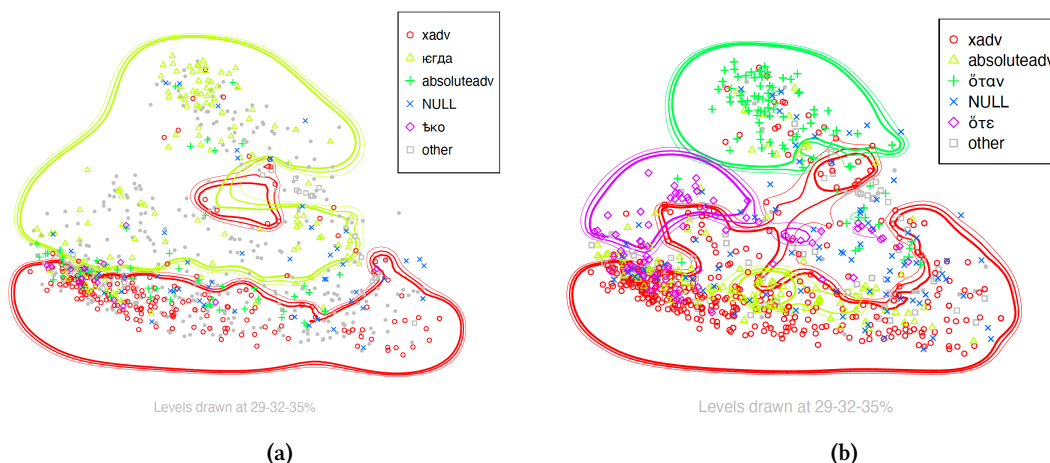


Figure 6.3: Kriging maps for (a.) Old Church Slavonic and (b) Ancient Greek after preprocessing and with syntactic placeholders for absolute constructions and conjunct participles

be a more likely English translation than ‘when *x*-ing’ for the 108 imperfective dative absolutes in the dataset.

Adding *while*-clauses to the *when*-maps affects the Old Church Slavonic map in a useful way, since this time dative absolutes occupy a clearly defined Kriging area (Figure 6.4a).

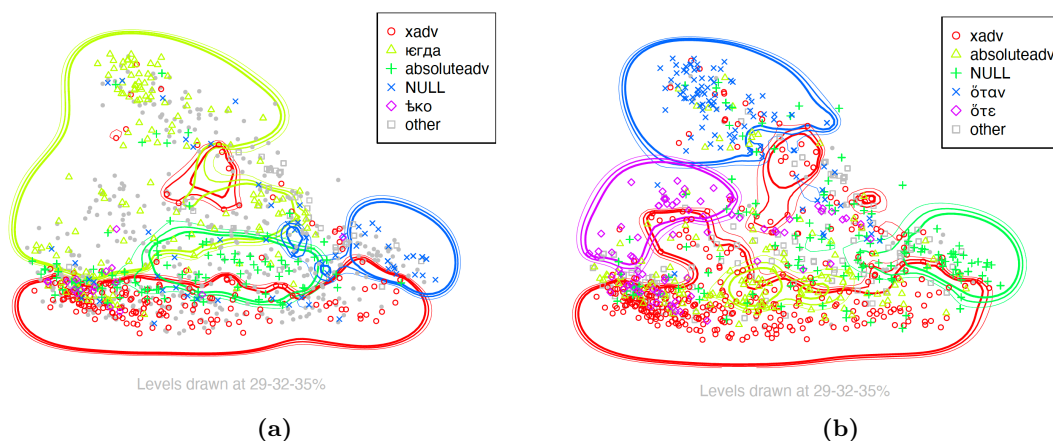


Figure 6.4: Kriging map for (a.) Old Church Slavonic and (b) Ancient Greek after adding the parallels to English *while* (in addition to *when*) and syntactic placeholders for absolute constructions and conjunct participles

If we compare these maps with the English map for *when* and *while* (Figure 6.5), we clearly see that Old Church Slavonic absolute constructions largely overlap with the WHILE area, suggesting that English *while* and dative absolutes are at least partially used in very similar contexts. Since a cluster is ultimately the result of the cross-linguistic comparison which was carried out via Hamming distance, the fact that the

data points corresponding to absolutes in Old Church Slavonic are close together in an MDS map in itself indicates that those points are also likely often encoded by the same means cross-linguistically.

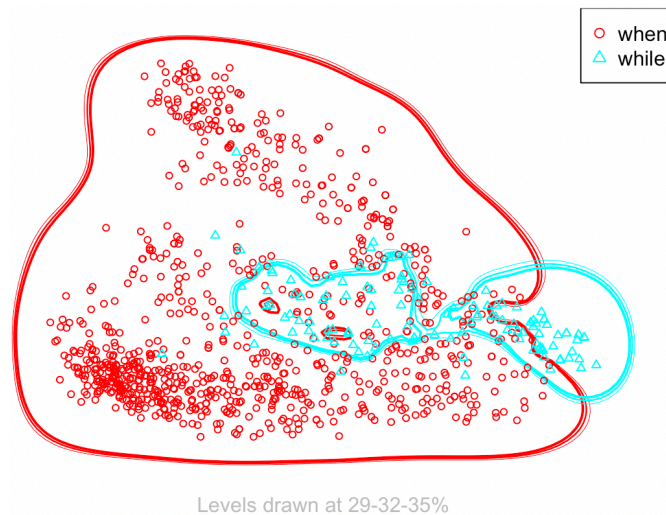


Figure 6.5: Kriging map for English (*when* and *while* situation).

The Ancient Greek map (Figure 6.4b) remains similar to the previous one, with absolutes occupying the same overall portion of the map as Old Church Slavonic dative absolutes. The Kriging areas for the construction in the two languages are, however, not quite identical. In both languages, they can be seen as forming a continuum with conjunct participles and are in particularly intense competition with both *when*-equivalents and conjunct participles in the BL area of the map, as already noted for other languages in the previous chapter. In Old Church Slavonic, the Kriging area overlaps with part of the *while* area in English (Figure 6.6a), whereas this is not as neatly the case in Ancient Greek (Figure 6.6b).

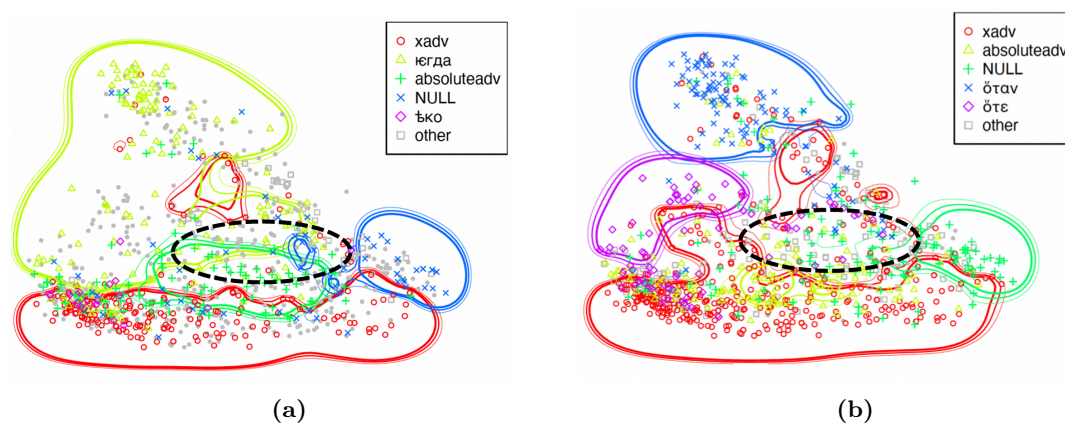


Figure 6.6: Kriging map for (a.) Old Church Slavonic and (b) Ancient Greek WHEN and WHILE situations. The black dashed circle corresponds to part of the English *while* area overlapping with absolutes in the two languages.

This mismatch between Old Church Slavonic and Greek may reflect the observation made in Chapter 2 that, although in the majority of cases Old Church Slavonic absolutes correspond to an Ancient Greek absolute, there is a small but consistent subset of Old Church Slavonic occurrences that do not correspond to an absolute in Greek, notably Greek *en tō* with infinitive ('during, while *x*-ing'). Overall, however, the fact that several *while*-clauses in English cluster to the very right corner of the semantic map, where all languages have some null constructions, suggests that they *may* not correspond to lexified counterparts as frequently as *when*-clauses. Whether this reflects some cross-linguistic regularity (i.e. WHILE-clauses are less likely to be separately lexified) or a lexical feature of the corpus is an open question which, evidently, cannot be investigated at scale with New Testament data.

6.2 Lexified counterparts to Early Slavic participle constructions in the world's languages

In the analysis in Chapter 5, we observed that few languages have relatively clear boundaries for *both* the middle-left (ML) cluster (GMM clusters 2) and the bottom-left (BL) cluster (GMM cluster 4), since in most cases the competition between different constructions is quite evident in one or both of those areas. Among these languages, we saw Pular, which had lexicalized means for both the ML and the BL clusters, as well as relatively low competition compared to what we observed in several other languages. In our precision and recall analysis in the previous chapter, this language always figured among those with a relatively high F1 score for the different clusters. An interesting characteristic of Pular is that it possesses several counterparts to *when*, several more, in fact, than the ones suggested by the Kriging areas in the Pular map (Figure 6.7, repeated from Chapter 5).

We can attempt to use existing descriptions of Pular to understand the meanings of each subordinator, potentially using them as a means of confirming (or refining) the observations made so far about the semantics of the different clusters in the semantic map of WHEN

According to Evans (2017), each connective conveys a particular temporal meaning and can generally only be found in combination with specific temporal-aspectual morphology on the clause it introduces. Figure 6.8 schematizes the range of their functions.

By crossing Evans's (2017) overview with the Kriging map in Figure 6.7, we see that four of the several means in Figure 6.8 are captured as Kriging areas. The TL cluster is mostly under the domain of *nde* but competes with *tuma*, which, in fact, has a small Kriging area in between the Kriging area for *nde* and the one for null constructions. *Nde*, according to the overview of Evans (2017), is used to introduce both future and past eventualities. However, a number of occurrences of *nde* are found in the bottom area, so it is possible that the latter reflect the use of this connective for past-tense clauses, whereas

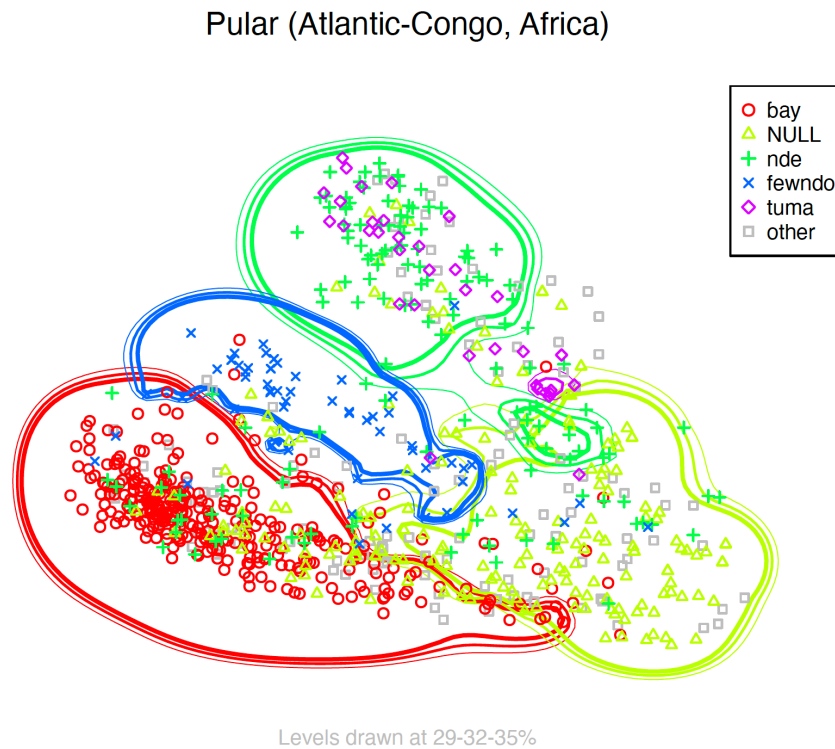


Figure 6.7: Kriging map for Pular (Atlantic-Congo, Africa), repeated

those under the Kriging area at the top-left of the map are, presumably, those introducing future-tense clauses, similarly to this usage of German *wenn*, Norwegian/Danish *når* or Afrikaans *wanneer* seen in the previous chapter. This seems to be confirmed by its competition with *tuma*, which, according to the overview, has a similar function.

The ML cluster is under the scope of *fewndo*, which the overview labels as PAST CIRCUMSTANCE. According to Evans (2017: 62), when introducing past events, the main difference among all the connectives is between those that are used to express sequentiality (or ‘main event line’) and those that introduce background information. The CIRCUMSTANCE category is considered as an intermediate category since ‘it can be used in the story line, but can also be used to provide a setting for what follows’ (Evans 2017: 65).

The BL cluster is instead under the domain of *bay*. This connective is one of several that are used to express sequentiality between the events they connect, differing in the level of ‘suddenness’ they convey. With *bay*, ‘there is no sense of suddenness; rather one of expected development’ (Evans 2017: 63), whereas the verbal noun (*-ugol*), *tun*, and *no...* *-irnoo*, convey a ‘greater sense of suddenness’, ‘a fair degree of suddenness’, and ‘the greatest sense of urgency’, respectively. (1a), (1b) and (1c) are examples containing

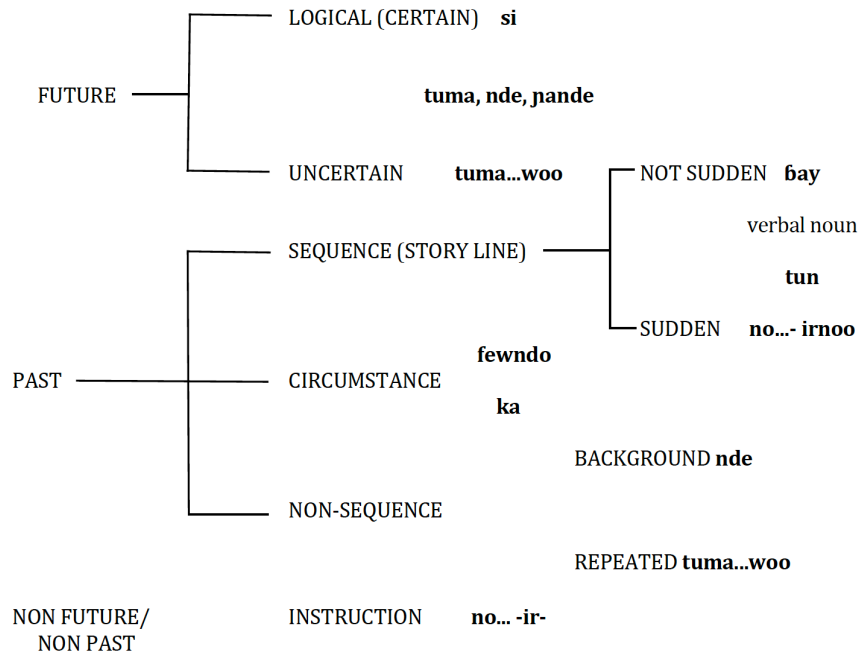


Figure 6.8: Temporal connectives meaning ‘when’ in Pular according to Evans (2017: 61)

bay, *tun*, and *no...-irnoo*, respectively, where the difference in ‘suddenness’ or ‘urgency’ is reflected, to some extent, in the English translations.

- (1) a. **Bay** o hewtii ka suudu, bumbe ben badii mo. Iisaa wi’i be: “Hidon sikka mido waawi wadude dun?” Be jaabii mo: “Hiyyii, yaa an Koohoojo.”
 ‘When he entered the house, the blind men came to him, and Jesus said to them, “Do you believe that I am able to do this?” They said to him, “Yes, Lord.”’ (Matthew 9:28)
- b. **Tun** o hewti, o badii Iisaa, e house himo wi’a: “Karamoko’en!” Onsay o hirbii mo.
 ‘And when he came, he went up to him at once and said, “Rabbi!” And he kissed him’ (Mark 14:45)
- c. **No** Elisabaatu nanirnoo salminaango Mariyama ngon, biddo makko on memminii ka nder reedu makko, kanko Elisabaatu o heewi Ruuhu Seniido on
 ‘And when Elizabeth heard the greeting of Mary, the baby leaped in her womb. And Elizabeth was filled with the Holy Spirit’ (Luke 1:41)

Tun and *no...-irnoo*, along with several other connectives from the overview in Figure 6.8, are not visible from the Kriging map but are represented, albeit to a much lesser extent, in the dataset of *when*-parallels. We can abstract from the individual forms

and use their meanings as provided by Evans (2017) and run Kriging on the functional labels instead of the forms themselves. We can then look at where in the semantic map the different functions of WHEN, as encoded by Pular connectives, cluster. To do so, unambiguous connectives with their functional labels are first annotated automatically. These include the following:

- *si*, which is used ‘to refer to a proposed or potential action in the future by relating it to another action which should precede it’ (Evans 2017: 61). However, a closer look at all the occurrences of *si* suggested that what is meant by FUTURE:CERTAIN in the overview are, in fact, non-past clauses with a universal (as opposed to existential) interpretation, as in (2).

(2) *Kono si fewta e Joomiraado on, taway tiggaare nden ittaama ka berde mabbe.*
But when one turns to the Lord, the veil is removed. (2 Corinthians 3:16)

The ‘logical’ or ‘certainty’ relation referred to be Evans (2017), therefore, may have to do with the observation that what is described by the main eventuality occurs *whenever* the eventuality described by the WHEN-clause occurs. Recall that this is precisely the use of universal WHEN commented on in the previous chapter. All these occurrences were therefore tagged as NONPAST:UNIVERSAL.

- *nande* and *tuma* (on its own, i.e. without *woo* following): Evans (2017: 62) explains that ‘uncertainty has basically to do with the time the event will happen, not whether the event will happen (although inevitably there is also some implication that way as well)’. This time, a closer look at the relevant examples indicates that these are instead non-past existential WHEN-clauses, as in (3).

(3) *Tuma o aroyi, o weebitanay aduna on fi junuubu e peewal e naawoore*
And when he comes, he will convict the world concerning sin and
righteousness and judgment (John 16:8)

These examples are therefore tagged as NONPAST:EXISTENTIAL.

- *bay*, verbal nouns (*-ugol*) or *tun*, and *no...irnoo*: as explained above, these all introduce sequential events in the past with different degrees of ‘suddenness’ or ‘urgency’. Among these connectives, *bay* is the only frequent one, the others occurring relatively rarely in our dataset. All of them are therefore tagged as PAST:SEQUENCE.
- *fewndo*-clauses were all tagged as PAST:CIRCUMSTANCE. A closer look at all the occurrences indicates that this connective is used predominantly in existential, but stative/durative past clauses, as in (4).

- (4) **Fewndo** o joododi e mabbe fii paamugol, o yetti bireedi on, o du'ii, o tayiti, o jonni be.

When he was at table with them, he took the bread and blessed and broke it and gave it to them. (Luke 24:30)

Connectives with more than one meaning were then manually disambiguated depending on the context. These are *tuma...woo* and *nde*, which can be used, according to Evans (2017), in clauses referring to the future or to the past. However, *tuma...woo* can convey repetition (i.e. a universal reading), not only in past but also in nonpast sentences, as (5) clearly shows. In fact, no example of this construction in past sentences was found in the dataset, so all its occurrences were tagged as NONPAST:UNIVERSAL.

- (5) **Tuma** men hoynaa **woo**, men du'anoo. **Tuma** men cukkaa **woo**, men wakkiloo. **Tuma** men jo'aa **woo**, men jaabora no newori. Men woni tuundi oo aduna, men hawkaa e di fow, haa weetaango hande ngoo.

When reviled, we bless; when persecuted, we endure; when slandered, we entreat. We have become, and are still, like the scum of the world, the refuse of all things. (1 Corinthians 4:11-13)

Nde, which is widely represented in the dataset, was tagged as PAST:BACKGROUND when in past clauses, as indicated by Evans (2017); however, it was also often found in pre-matrix position in past sentences, unlike what is stated by the author, as in (6c). Regardless of position, *nde* in past sentences was assumed to introduce *Past:Background*. A close look at the use of *nde* in non-past sentences showed that it is used both existentially, as in (6a) and universally (6b). These were tagged as NONPAST:EXISTENTIAL and NONPAST:UNIVERSAL, respectively.

- (6) a. **dun**, **nde** mi aroyi, mi wattiday e be subidon ben dokke mon den, wondude e bataake, be naba Yerusalaam
 And when I arrive, I will send those whom you accredit by letter to carry your gift to Jerusalem. (1 Corinthians 16:3)
- b. **Nde** neddo dolnudo, jom aalaaji ayni galle mun woo, jawle mun den daday.
 When a strong man, fully armed, guards his own palace, his goods are safe (Luke 11:21)
- c. **Nde** Iisaa hewtunoo e ley kin yibbehi, o tiggitii, o wi'i
 And when Jesus came to the place, he looked up and said to him (Luke 19:5)

Ka, expressing PAST:CIRCUMSTANCE was left out because of a high degree of homophony with other instances of *ka* and/or multifunctionality, which made it difficult to annotate it with confidence.

Additionally, *bawto* was annotated as AFTER (see Bah 1980), while *ontuma* and *onsay*, both meaning ‘at that time’ (Evans 2017: 64), were tagged as MAINCLAUSE if no other connective was found, else as *other*.³

The result of running Kriging on the new labels is shown in Figure 6.9.

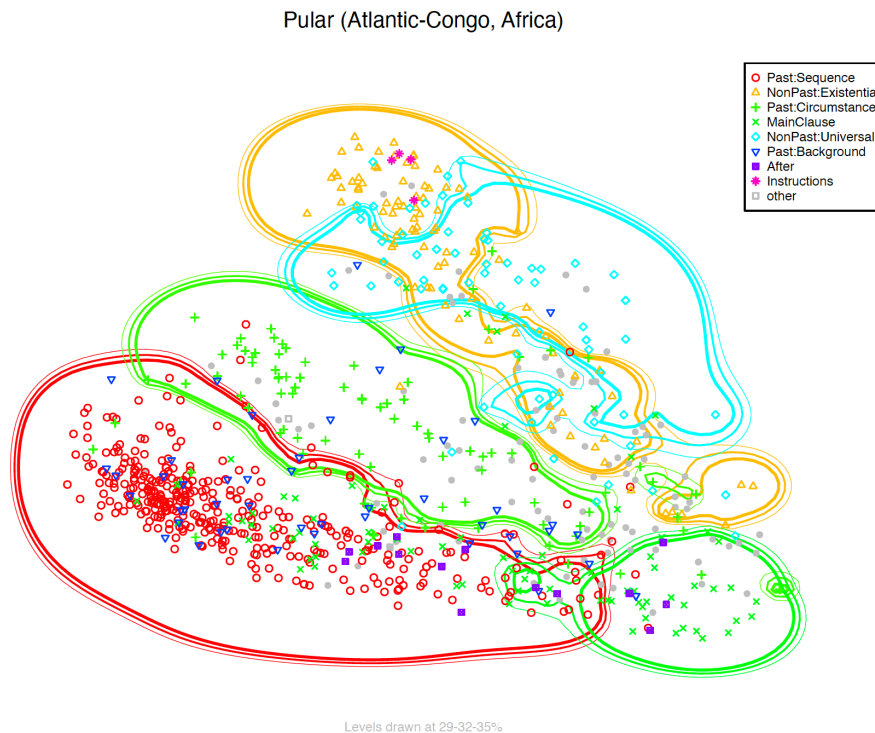


Figure 6.9: Pular WHEN-map with Kriging based on the functional annotation of Pular connectives

The new Kriging map shows better-defined areas for the functional scopes of the various Pular connectives. Besides confirming that both universal and existential non-past clauses cluster at the top of the map, it shows a more clear-cut separation between the top area and the bottom area by means of connectives conveying what Evans (2017) calls PAST:CIRCUMSTANCE. The intermediate status of these constructions between sequential and non-sequential uses of WHEN-clauses referred to by Evans (2017) thus finds clear confirmation in the intermediate position of this functional cluster between the WHEN-clauses at the top of the map and those at the bottom in the semantic map. It also provides further confirmation that one of the main differences between the top and bottom area, the latter corresponding to the vast majority of absolute constructions and conjunct participles in the Old Church Slavonic map, is the stress on sequentiality conveyed by the situations corresponding to the data points at the bottom. We also observe that main clauses dominate the very bottom right corner, as expected, where

³The annotated dataset is available in the project repository at <https://doi.org/10.6084/m9.figshare.24166254>.

Old Church Slavonic and Ancient Greek main clauses corresponding to English *when* also cluster (cf. Figure 6.12).

It is important, however, to stress that cases such as Pular simply confirm that the bottom area of the semantic map, within which most participle constructions fall in Early Slavic, corresponds to a relatively well-defined function from the perspective of temporal relations, namely that between existential-eventive clauses, which translates into the sequentiality between events being stressed. What it does *not* help clarify is the competing motivations governing the choice between *when*-counterparts that also occur in the same area and participle clauses. In all likelihood, this is a matter of information structure and discourse organization, which the semantic map does not capture well.

In Chapters 1-4, the analysis of the information-structural properties of subjects in *jegda*-clauses and participle clauses suggested that *jegda*-clause may be a better choice when there is a topic continuation from the immediately preceding discourse, as opposed to a topic shift, more likely to be signalled by some sentence-initial framing participle. This observation was prompted, among other things, by the fact that several *jegda*-clauses have a null subject despite more frequently having a different subject from that of their matrix clause. As we will see in the next section, the topic-shift function of framing participles finds some cross-linguistic parallels in other types of null constructions, particularly those used with a clause-linkage function. However, other factors related to discourse organization and Time-Space-Participant-connectedness (Rothstein 2003, Fabricius-Hansen 2006) are at play, which different phenomena that have fallen so far under the generic notion of ‘null constructions’ in the world’s languages can help clarify.

6.3 Null constructions in a cross-linguistics perspective

As discussed in previous chapters, the discourse functions of conjunct participles in Ancient Greek and Early Slavic can partly be inferred compositionally from the relative order and tense-aspect of participle and matrix clause (Haug 2012), allowing us to single out their usage as INDEPENDENT RHEMES, which are discourse-coordinated and can be found stacked up as sequences of foregrounded clauses controlled by an argument of the matrix clause, as in the Old Church Slavonic and Ancient Greek example in (7a) and (7b).

- (7) a. *i abie tekъ edinъ otъ*
 and immediately run.PTCP.PFV.M.NOM.SG one.M.NOM.SG from
nichъ. i priemъ gobъ.
 he.GEN.PL and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG
isplъnъ ocъta. i vъzneъzъ
 fill.PTCP.PFV.M.NOM.SG vinegar.GEN.SG and put.PTCP.PFV.M.NOM.SG
na trъstъ. napaěše i
 on reed.ACC.SG give.to.drink.IMPF.3.SG he.ACC.SG

- b. *kai eutheōs dramōn heis ex*
 and immediately run.PTCP.PFV.M.NOM.SG one.NOM.SG from
autōn kai labōn spongon
 he.GEN.PL and take.PTCP.PFV.M.NOM.SG sponge.ACC.SG
plēsas te oxous kai
 fill.PTCP.PFV.M.NOM.SG with vinegar.GEN.SG and
peritheis kalamō epotizen auton
 put.PTCP.PFV.M.NOM.SG reed.DAT.SG give.to.drink.IMPF.3.SG he.ACC.SG
 ‘Immediately one of them ran and took a sponge, filled it with sour wine and
 put it on a reed, and offered it to him to drink’ (Matthew 27:48)

Pre-matrix perfective conjunct participles in New Testament Greek and Old Church Slavonic were singled out as most typically INDEPENDENT RHEMES, although an (often ambiguous) FRAME reading can be available for sentence-initial conjunct participles. We have seen in Chapters 1 and 2 that absolutes regularly occur sentence-initially, often introducing a series of foregrounded clauses, and can instead be considered typical background material (i.e. FRAMES) regardless of tense-aspect (cf. also Pedrazzini 2022).

We can test this claim by identifying typical INDEPENDENT RHEMES and FRAMES among null alignments using the linguistic annotation in PROIEL, which may allow us to look at differences in where they cluster in the semantic map. Using Ancient Greek, because of its greater coverage of the New Testament, all pre-matrix aorist conjunct participles with an aorist main verb were tagged as INDEPENDENT RHEMES and all sentence-initial absolute constructions as FRAMES. This is reminiscent of what we did in Figure 6.4b; this time, however, *while*-situations are not included, and a specific configuration of conjunct participles (pre-matrix, aorist) is isolated as a construction of its own. After running Kriging on the newly labelled data points, we obtain the map in Figure 6.10a (remaining null alignments are labelled as ‘other_NULL’). For ease of comparison, Figure 6.10b highlights all and only the null observations belonging to the Kriging areas for the identified typical FRAMES and INDEPENDENT RHEMES.

As the figures show, typical INDEPENDENT RHEMES and FRAMES (or, in other terms, foreground and background matter, respectively) are predominant in a dedicated Kriging area at the bottom half of the map, stretching out from the area corresponding to GMM cluster 4 (i.e. the BL area) towards the right side of the map, where other, non-further-defined NULL constructions are found. INDEPENDENT RHEMES and FRAMES each receive a contiguous, but relatively well-defined Kriging area.

It is interesting to notice that typical FRAMES are found above INDEPENDENT RHEMES in the map, that is, closer to lexified WHEN-clauses.

As discussed in Chapter 5, the connective *when* (and similarly *lorsque* and *quand* in the literature on French, as well as English *while*) have been widely recognized as ‘triggers’ or ‘clues’ for backgrounding rhetorical relations in formal frameworks of discourse representation (Reese et al. 2007; Asher 2004; Prévot 2004; Asher et al. 2007), namely

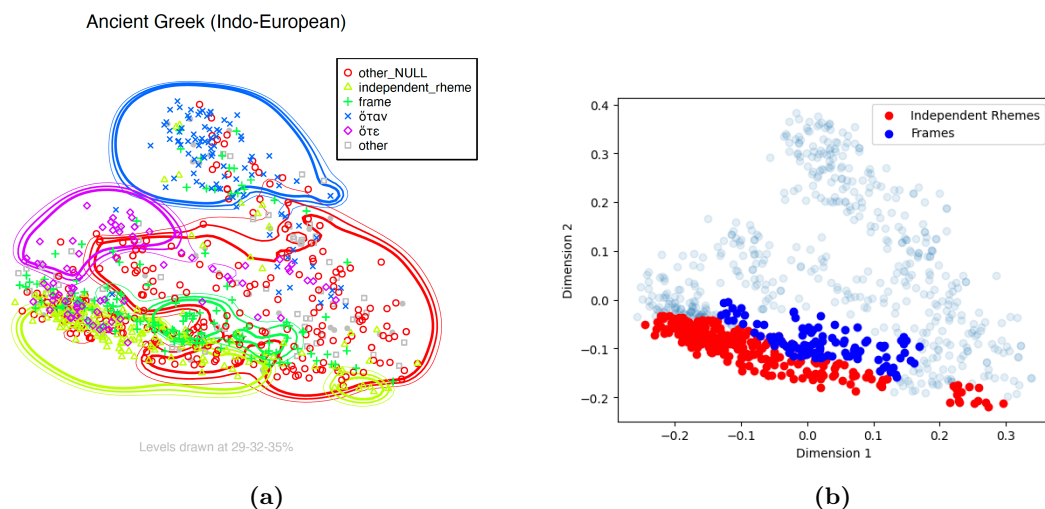


Figure 6.10: (a) Ancient Greek WHEN-map with Kriging based on functional labels, namely typical INDEPENDENT RHEMES (i.e. pre-matrix, perfective, with an aorist main verb) and typical FRAMES (i.e. sentence-initial absolute constructions); (b) highlights of any point found inside the Kriging areas for FRAMES (blue) and INDEPENDENT RHEMES (red).

as introducers of a background frame for a foregrounded event(uality). If an equivalence be made in discourse-structural terms between *hóte/hótan* and *when* as *Background-triggers*, then the relative greater closeness of FRAMES to *hóte* and *hótan* in the map in Figure 6.10a adds a further layer of distinctions within the WHEN-map—that between backgrounds and foregrounds (or FRAMES and INDEPENDENT RHEMES). If this were the case, then the semantic map for Ancient Greek is close to the idea of a continuum between the various constructions, including null ones—from typical foregrounded clauses at the bottom to typical backgrounded clauses at the top of the map.

As Figures 6.11a and 6.11b show, even by simply highlighting finite *when*-counterparts, absolute constructions and conjunct participles on the MDS map, we clearly see that absolute constructions occupy an intermediate position along the *y*-axis between the bulk of conjunct participles (at the very bottom of the map) and *when*-counterparts (*jegda*, *hóte/hótan*, at the top of the map).

In a way, this continuum also reflects the intuition that the closer we get to the conjunct participles at the bottom of the map, the less subordinate-like the constructions are from the pragmatic perspective—from *when*-counterparts at the top (most subordinate-like) to the conjunct participles at the bottom (least subordinate-like). At the same time, the constructions should be less and less subordinate-like moving rightwards along the *x*-axis, since the bottom right corner is where null constructions are found in *all* languages in the dataset, albeit to very different extents. This is, in fact, reflected in the location of Ancient Greek and Old Church Slavonic independent clauses corresponding to *when*-clauses in the semantic map. Figure 6.12 shows the average location of different

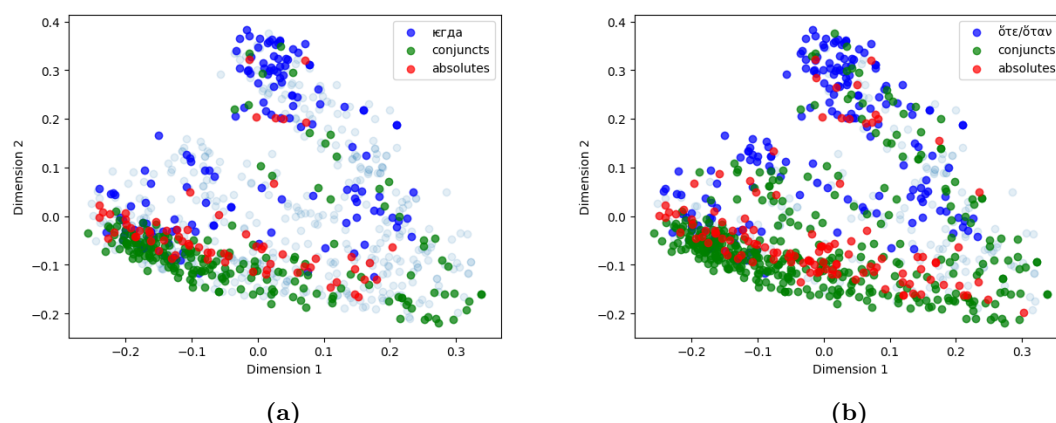


Figure 6.11: MDS maps of *WHEN*, highlighting all finite subordinates (*jegda* and *hóte/hótan*), absolute constructions and conjunct participles in (a) Old Church Slavonic and (b) Ancient Greek

constructions corresponding to *when* in Ancient Greek and Old Church Slavonic. To obtain the map, the Ancient Greek and Old Church Slavonic counterparts to *when* were first divided into groups, corresponding to different constructions in the two languages. Each construction was then represented on the map by averaging all the observations for that particular group. In this way, we can then check where in the map the average observation for a particular construction is found.

As expected, independent clauses are found, on average, towards the bottom right of the semantic map, similarly to prepositional phrases (e.g. *vъ syrě drěvě*, corresponding to English ‘when the tree is green’, in Old Church Slavonic literally ‘in the moist [as opposed to dry] tree’), among others.

The blurred boundaries between coordination and subordination when it comes to participle clauses had already been stressed by Bary & Haug (2011) in their account of Ancient Greek conjunct participles. As we could already gather from the cross-linguistic pattern of concentration of null observations to the bottom half of the map (cf. the heatmap in Figure 5.9 in Chapter 5), this is by no means unique to Old Church Slavonic and Ancient Greek. The dependence of clauses on context for their interpretation as coordinate or subordinate structures is, in fact, a major feature of languages that make extensive use of converbal forms and/or that allow clause chaining and has thus been an important topic in the relevant literature (see, among others, Finer 1984, 1985; Roberts 1988a; Bickel 1991; Bisang 1995b; König 1995; Broadwell 1997; Longacre 2007; Dooley 2010; Nonato 2014). The rest of this section touches on a number of (often interconnected) phenomena, observed in languages that appear to be typologically similar to Old Church Slavonic and Ancient Greek (i.e. with a similar split between lexified *when*-counterparts and null constructions), or whose Kriging map only contains null constructions. Section 6.3.1 looks at clause chaining, which has already been mentioned in passing in previous

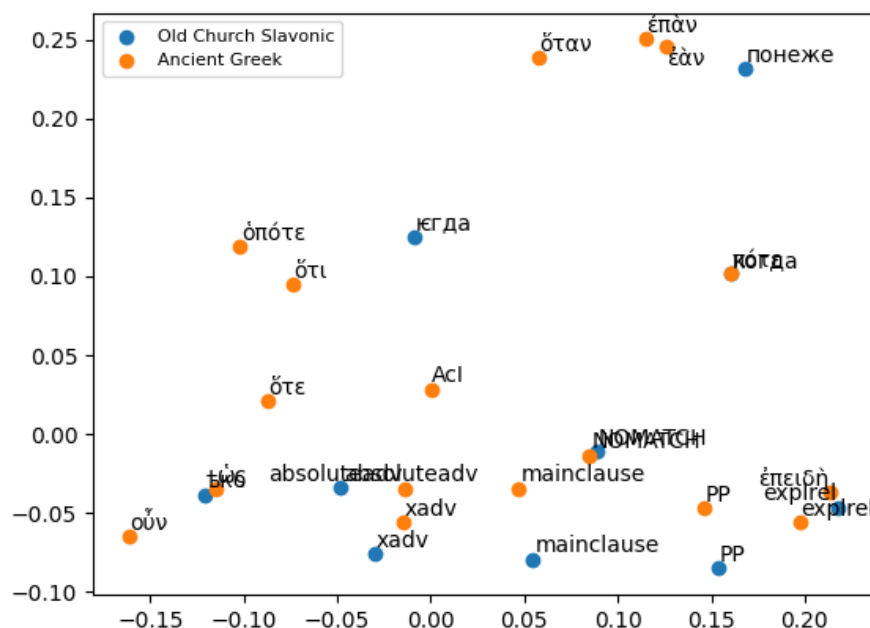


Figure 6.12: Location of the average observation for each different construction in Old Church Slavonic and Ancient Greek. *AcI* = accusative with infinitive, *absoluteadv* = absolute construction, *xadv* = conjunct participle, *PP* = prepositional phrase, *explrel* = explicit relative.

sections, before looking at examples of bridging (or clause-linkage) (Section 6.3.2), switch-reference (Section 6.3.3), and insubordination (Section 6.3.4). As we will see, the clear parallels to these phenomena, attested in genetically and areally unrelated languages, provide some independent evidence in support of the observations made so far on the functional scope of null constructions in the WHEN-map, but will also provide some clarifications regarding the division of labour between absolute constructions and conjunct participles.

6.3.1 Clause chaining

Clause chaining is the ‘possibility of long sequences of foreground clauses with operator dependence, typically within the sentence’ (Dooley 2010: 3). *Operator dependence* refers to the fact that the clauses in clause chains are typically *deranked* verb forms (Stassen 1985; Croft 1990; Cristofaro 1998, 2019), that is, lacking marking of one or more tense, aspect, or mood distinctions (‘operators’) compared to independent clauses in the same language. They can be additionally marked by specific means signalling their status as ‘medial’ clauses or ‘converbs’, as the Avar and Korean examples in the previous chapter, for instance, showed. However, as is clear from Dooley’s (2010) definition, the presence of converbal morphology in a language (e.g. Italian or Ukrainian, as we saw in the previous

chapter) does not necessarily entail the possibility of clause chaining, which specifically refers to the possibility of having sequences of *foregrounded* clauses, namely clauses that function as independent predications from the discourse perspective. The Avar sentence in (8) is a prototypical example of clause chaining.⁴

(8) Avar (North Caucasian)

Eadal ččun vaqun hebsaġat qvaṭive tuhun vačana
 water dip.PFV.CVB rise.PFV.CVB immediately out go.PFV.CVB behold
Ġisa. Hebmexal zobgi kibilizabun, mikkidul suratalda boržun
 Jesus then sky split.PFV.CVB dove image fly.PFV.CVB
bačun, Allahasul Ruħ Ġisaqe rešštuneb bixana.
 approach.PFV.CVB God Spirit Jesus rest see.AOR

‘And when Jesus was baptized, immediately he went up from the water, and behold, the heavens were opened to him, and he saw the Spirit of God descending like a dove and coming to rest on him’ (Matthew 3:16).

As Dooley (2010) observes, a clause chaining sentence may, and often does, *also* contain backgrounded clauses, which can have the same type of operator dependence as the foregrounded clauses, that is, they can be morphologically identical to foregrounded clauses, but their content is presuppositional and serves to limit the scope of the main predication, similarly to how we have been defining FRAMES in the previous chapters. The Old East Slavic and Mbyá Guaraní sentences in (9) and (10), respectively, from our corpus exemplify this.

- (9) *i prišdъ kъ prpdbъnumu antoniju jeho*
 and come.PTCP.PFV.M.NOM.SG to saint.DAT.SG Antonij.DAT 3.SG.M.GEN
že viděvъ i padъ pokloni
 PTC see.PTCP.PFV.M.NOM.SG and fall.PTCP.PFV.M.NOM.SG bow.AOR.3.SG
sja jemu sъ slzami molja sja
 REFL 3.SG.M.DAT with tear.INST.PL beg.PTCP.IPFV.M.NOM.SG REFL
jemu da by u nego
 3.SG.M.DAT so that be. AOR.3.SG at 3.SG.M.GEN
bylъ
 be.PTCP.RESULT.ACT.M.NOM.STRONG

‘When he came to St. Antonij and saw him, he prostrated himself and bowed, begging him with tears to let him stay with him’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 31b)

(10) Mbyá Guaraní (Tupian)

Ha'e nunga ha'e kuery oendu vy ko'ẽmba'i ramove templo py oike
 such things they COLL hear SS daybreak right after temple in enter
vy nhombo'e
 SS teach

⁴All glossing in this chapter is my own unless otherwise stated.

‘And when they heard that, they entered the temple early in the morning and taught’ (Acts 5:21)

In (9), the conjunct participles *prišvḍ* (lit.) ‘having come’ and *viděṿ* (lit.) ‘having seen’ can be considered as background material (i.e. FRAMES), unlike the morphologically identical participle *paḍ* (lit.) ‘having fallen’, which belongs instead to the foreground, as also reflected in the English translation. Similarly, the first verb in the Mbyá Guaraní chain sentence in (10) belongs to the background, while the following is part of the foreground.

As Dooley (2010: 12) observes, a clause in a chaining structure can be ambiguous between foreground and background, similarly to the ambiguity we have seen between INDEPENDENT RHEME and FRAME participles sentence-initially, as the example in (11) shows.

- (11) a. *i povelěṿ narodoṃ vbzlešti na*
 and command.PTCP.PFV.M.NOM.SG crowd.DAT.PL lay down.INF on
trěvě i prieṃ pěṭ chlēḅ i
 grass.LOC and take.SPTCP.PFV.M.NOM.SG five.ACC bread.GEN.PL and
dvě rybě ṿbzrěṿ na nbo
 two.ACC fish.ACC look up.PTCP.PFV.M.NOM.SG to heaven.ACC
blgsvi. i přeloṃ chlēby
 bless.AOR.3.SG and break.PTCP.PFV.M.NOM.SG bread.ACC
dasṭ oučēnikoṃ. oučēnici že narodoṃ.
 give.AOR.3.SG disciple.DAT.PL disciple.NOM.PL PTC crowd.DAT.PL

- b. *kai keleusas tous okhlous*
 and command.PTCP.PFV.M.NOM.SG the.ACC.PL crowd.ACC.PL
anaklithenai epi tou khortou labōn
 lay down.INF.PAS.AOR on the.GEN grass.GEN take.PTCP.PFV.M.NOM.SG
tous pente artous kai tous duo ikthuas
 the.ACC.PL five bread.ACC.PL and the.ACC.PL two fish.ACC.PL
anablepsas eis ton ouranon eulogesen
 look up.PTCP.PFV.M.NOM.SG to the.ACC heaven.ACC bless.AOR.3.SG
kai klasas edōken tois
 and break.PTCP.PFV.M.NOM.SG give.AOR.3.SG the.DAT.PL
mathetais tous artous hoi de
 disciple.DAT.PL the.ACC.PL bread.ACC.PL the.NOM.PL PTC
mathetai tois okhlois
 disciple.NOM.PL the.DAT.PL crowd.DAT.PL

‘Then he ordered the crowds to sit down on the grass, and taking the five loaves and the two fish, he looked up to heaven and said a blessing. Then he broke the loaves and gave them to the disciples, and the disciples gave them to the crowds’ (Matthew 14:19)

In (11), the series of conjunct participles could be seen as ambiguous between background and foreground material. This is also reflected in the different English translations

—the English Standard Version (given under the example), for instance, interprets the clause corresponding to the first conjunct participle as an independent clause (*he ordered the crowds*), the second as a subordinate (*taking the five loaves*), and all the others as independent clauses (*he looked up... said a blessing ... broke the loaves ... gave them to the disciples*). Other versions show different interpretations. For example, the King James Version (and similarly NKJV) translates the first two clauses as main clauses (*he commanded... and took the five loaves*), the third as subordinate (*looking up*), then all the remaining ones as main clauses (*he blessed, and brake, and gave*).

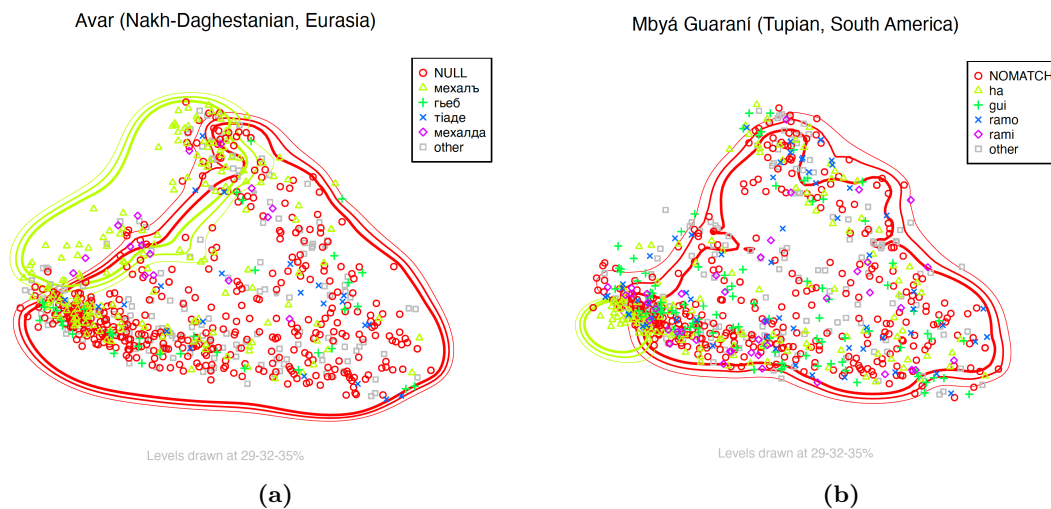


Figure 6.13: Kriging maps for Avar (Nakh-Daghestanian, Eurasia) and Mbyá Guaraní (Tupian, South America)

As pointed out by Dooley (2010: 13), not every language with a chaining construction is necessarily a clause-chaining language. Whichever the case for Early Slavic, there are clear similarities in the use of chaining constructions in clause-chaining languages such as Avar and Mbyá Guaraní and series of conjunct participles in Old Church Slavonic. Even among clause-chaining languages, there are major differences in the functions of chaining constructions, as is also reflected in the different maps for Mbyá Guaraní (Figure 6.13b) and Avar (Figure 6.13a), despite both being clause-chaining languages.

Several authors who have proposed explanations for the use of clause chaining as a narrative strategy have put forward criteria that could be subsumed under a general principle of *event grouping*, as is referred to by Dooley (2010: 14). In a nutshell, the assumption is that the *syntactical* integration of clauses in narrative chaining reflects some level of *conceptual* conflation as well (Dooley 2010: 14; cf. also Chafe 1994: 145; Givón 2001; Talmy 1991). A chaining construction may thus constitute a complete discourse or ‘developmental’ unit (Levinsohn 2009). In Thurman’s (1975: 344) words:

Primarily final verbs [i.e. the finite verb at the end of a chaining construction, A/N] are used to alter the perspective on the topic at hand by terminating one chain so that another may begin. This may mean a change in participant

orientation so that another character in the narrative becomes prominent; or it may involve reorientation only in time or space.

Similarly, Fabricius-Hansen (2007) and Sæbø (2011b) argued that absolute constructions, and similarly co-predicative participles, provide a guarantee of tense-space-participant connectedness and ‘serve to build groups of events or states, expressing that the host and supplement eventualities form interesting sums of eventualities’ whereby ‘the core event and co-eventualities all add up to one super-group eventuality’ (Sæbø 2011b: 1438).

6.3.2 Bridging

Background clauses in clause chains, just like FRAMES, are presuppositional (in the sense of Lambrecht 1994: 36-73), that is, they contain reference to given or accessible information and, because they can link back to the previous discourse, they often work as bridging devices. It is not uncommon for languages with clause chaining to use bridging systematically at the beginning of most chaining constructions by repeating the content of the previous foregrounded clause and re-purposing it, as it were, as the background for a new clause chain. This is a phenomenon variously referred to, among others, as *bridging constructions* (e.g. Guérin & Aiton 2019), *tail-head linkage* (e.g. de Vries 2005), *recapitulation clauses* (e.g. Stirling 1993), *summary-head linkage* (Thompson, Longacre & Hwang 2007), or simply *clause linkage* (AnderBois & Altshuler 2022). While bridging is not used as systematically in Early Slavic as in some of these languages, we have seen how we frequently find FRAMES using either previously foregrounded eventualities or event participants as part of the background for a new portion of foregrounded discourse to come (cf. Chapter 2, Section 2.3.2). Bridging as a systematic discourse-organization device follows precisely this overall principle, as is clear from its definition in Guérin & Aiton (2019: 2–3):

A bridging construction is a linkage of three clauses. The first clause of the construction (i.e., the reference clause) is the final clause in a unit of discourse. The second clause (i.e., the bridging clause) recapitulates the reference clause. It usually immediately follows the reference clause but it acts as the initial (albeit non-main) clause of a new discourse unit. The primary discourse function of a bridging construction is to add structure and cohesion: recapitulation backgrounds the proposition of the reference clause and foregrounds the clause following the bridging clause. This third clause is discourse-new and typically sequentially ordered.

Or, in AnderBois & Altshuler’s (2022: 802) schematization:

[REFERENCE]_R. [[BRIDGE]_B CONTINUATION]_C

where **R** is the reference clause as per the definition by Guérin & Aiton (2019) above, **B** is the bridging clause which, together with the discourse unit following the bridging

clause (the third, foregrounded clause of Guérin & Aiton’s definition), constitute the continuation clause (C).

AnderBois & Altshuler (2022) studied the phenomenon in A’ingae (Cofán; see the map in Figure 6.14, where only *null* constructions have a Kriging area) and gave an SDRT account of the rhetorical relations systematically triggered by bridging constructions.

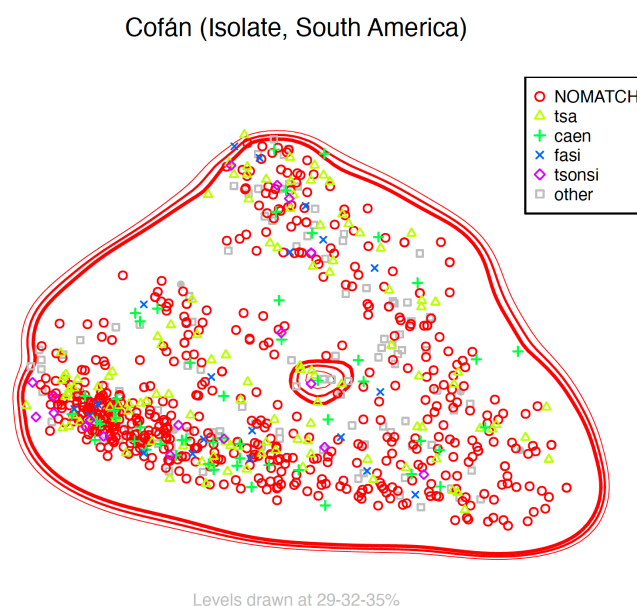


Figure 6.14: Kriging map for Cofán (Isolate, South America)

According to them, bridging requires the relations $Narration(R,[B,C])$ and $Background_{Forward}(B,C)$. In other words, the reference clause attaches to the complex discourse unit made of the bridging clause and continuation clause via $Narration$, whereas the bridging clause attaches to the continuation clause via $Background_{Forward}$. This is the same principle that we observed in the case study on strategically annotated treebanks in Chapter 2, specifically when we observed that dative absolutes appeared to systematically repeat *some* portion of a previously foregrounded unit, either the immediately preceding clause or some clause not far behind in the discourse. This is what we see in (12), repeated from Chapter 2, and (13)-(14).

- (12) [*prizovi* *mi* *ju*]_R *da* *se* *poveselju* *dn̄s*
 summon.IMP 1.SG.DAT 3.SG.F.ACC so that REFL enjoy.PRS.1.SG today
s *njeju.* *z̄elo* *bo* *jako* *slyšalb* *jesm*
 with 3.SG.F.INST much for that hear.PTCP.RESULT.M.SG be.PRS.1.SG
o *njei* *godě* *mi* *jestb.*
 about 3.SG.F.LOC pleasant.ADV 1.SG.DAT be.PRS.3.SG
 [[*prizvaně* *že* *byvši* *ei*]_B.
 summon.PTCP.PFV.PASS.F.DAT.SG PTC be.PTCP.PFV.F.DAT.SG 3.SG.DAT
pride *k* *njemu*]_C
 come.AOR.3.SG to 3.SG.DAT

‘‘Summon her, so that I can enjoy myself with her today. From what I’ve heard of her, I am much attracted by her’’. After being summoned, she came to him.’ (*Life of Mary, Abraham’s Niece*, Bdinski Sbornik f. 7v)

- (13) *nǔ sice sǔtvorivě da [azǔ ti lǔjagu na*
 but thus do.IMP.1.DU that 1.SG.M.NOM PTC lie down.PRS.1.SG on
vozě]R ty že mogyi na koni
 chariot.LOC.SG 2.SG.NOM PTC be able.PTCP.IPFV.M.NOM on horse.LOC.SG
ěchati. to že blaženyi sǔ vǔsjakyimǔ sǔměrenijemǔ
 ride.INF so PTC blessed.M.NOM.SG with all.N.INST.SG humility.N.INST.SG
vǔstavǔ sěde na koni. [[a
 stand up.PTCP.PFV.M.NOM.SG sit.AOR.3.SG on horse.LOC.SG and
onomu že legǔšju na vozě.]B i
 that.DAT.SG PTC lie down.PTCP.PFV.M.DAT.SG on chariot.LOC.SG and
idjaše putǔmǔ raduja se i
 go.IMP.3.SG way.INST.SG rejoicing.PTCP.IPFV.M.NOM.SG REFL and
slavja ba.]C
 glorify.PTCP.IPFV.M.NOM.SG god.GEN

‘[Iziaslav:] ‘‘But let’s do this: I’ll lie down in the chariot, and you can ride the horse’’. The Blessed Theodosius humbly got up and mounted his horse. And as he [Iziaslav] lay down in the chariot, he [Theodosius] continued on his way, rejoicing and glorifying God.’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 43b)

- (14) *i došedǔ velika dne.*
 and reach.PTCP.PFV.M.NOM.SG great.M.GEN.SG day.M.GEN.SG
vskrsnǔja po obyčajju. prazdnova
 resurrection.N.GEN.SG according to custom.DAT.SG celebrate.AOR.3.SG
světlo. [vpade v bolěznǔ.]R [[razbolěvšju bo se
 joyfully fall.AOR.3.SG in illness.ACC fall ill.PTCP.PFV.M.NOM.SG PTC REFL
jemu.]B i bolěvšju. dnii .e. po
 3.SG.M.DAT and be ill.PTCP.PFV.M.DAT.SG day.GEN.PL five after
semǔ byvšju večeru. povelě
 this.LOC.SG be.PTCP.PFV.M.DAT.SG evening.DAT.SG command.AOR.3.SG
iznesti se na dvorǔ.]C
 carry.INF REFL in yard.ACC

‘But when the great day of Easter arrived [lit. when he reached the great day of Easter], according to custom, he celebrated it joyfully. He fell ill. Having fallen ill and having been ill for five days, once in the evening he ordered to be carried out to the yard’ (PVL, 62d)

In (12), the dative absolute *prizvaně byvši* ‘being summoned’ repeats the main verb of the previous sentence (*prizovi* ‘summon’), repurposing it as the background for the new portion of the discourse (*pride* ‘she came’), sequentially located *after* the summoning event. This could be analysed following the structure described by Guérin & Aiton (2019) and AnderBois & Altshuler (2022), where *prizovi* is part of the reference clause, *prizvaně byvši* is the bridging clause and *pride* is part of the continuation clause. Very

similarly, in (13), *ljagu* ‘I will lie down’ can be seen as the reference clause, *onomu že legošju* ‘having lain down’ as the bridging clause, and *idjaše* ‘he went’ as the continuation clause. Likewise, in (14), *vpade v bolěznъ* ‘he got sick’ would be the reference clause, the immediately following *razbolěvšju bo se jemu* ‘having gotten sick’ the bridging clause, and *povelě* ‘he commanded’ as the continuation clause.

As the simple example in (15) shows, bridging clauses in A’ingae are introduced by subordinators with switch-reference morphology, indicating whether the subject of the bridging clause is the same or different from that of the continuation clause (cf. Fischer & van Lier 2011; AnderBois, Altshuler & Silva 2023).⁵

- (15) *Tse’faei’ccuyi* [*aje chattian.*]_R [[*Aje chattian-si*]_B *tsa pushesuja*
 immediately fever alleviate fever alleviate-DS ANA woman
poi’yi’ccongā cui’ñā]_C.
 all serve
 ‘and the fever left her, and she began to serve them’ (Matthew 8:15)

Marking bridging clauses for switch reference is very common among languages where such marking is available (cf. Stirling 1993; de Vries 2005; Guérin & Aiton 2019). In fact, we find many similarities between the use of participle constructions in Early Slavic and switch reference systems of several language families, most notably many Papuan (Roberts 1997) and South American (van Gijn, Haude & Muysken 2011; van Gijn 2012, 2016b; Overall 2014, 2016) families, beyond bridging constructions. We now therefore turn to looking at some of these parallels in more detail.

6.3.3 Switch reference

In a nutshell, *switch reference* is a morphological system for tracking referents in an ongoing discourse (Roberts 2017: 538). In a ‘canonical’ switch-reference system (cf. Haiman & Munro 1983: ix), a clause is marked to signal whether its subject is co-referential or not with the subject of another, usually adjacent, clause. With subject co-reference, a same-subject marker is used (SS), else a different-subject marker is employed (DS). Switch reference is overwhelmingly present in languages that allow and use clause chaining extensively, in which case switch reference marking would typically occur on each medial clause leading up to the final verb. (16) is an example of ‘canonical’ switch reference in Huichol-Wixárika from our corpus.⁶ In (16a), *me-’u’-axüa-cu* has the DS marker *-cu*, indicating that its subject referent differs from that of its matrix clause. Conversely, *me-’u’-axüa-ca* (16b) has the SS marker *-ca*, indicating co-referentiality with the matrix subject.

⁵As we will see in more detail in Section 6.3.3, ‘switch-reference’ does not specifically only refer to different-subject marking, as the term has sometimes been used in the literature on Early Slavic dative absolutes (e.g. Collins 2011), but it is the general term used to describe the phenomenon whereby clauses are marked according to whether their subject is co-referential or not with that of the matrix (i.e. it includes both different-subject and same-subject markers).

⁶In the Huichol example, I kept the spelling of the Bible translation in Mayer & Cysouw’s (2014) corpus. Note, however, that this is not the most common orthography found in most studies on Huichol today.

(16) Huichol-Wixárika (Uto-Aztecan)

a. *Hesüana me-'u'-axüa-cu müpaü*
 to.him 3.PL.SBJ-VIS-arrive.PL-DS thus
ti-ni-va-ru-ta-hüave
 DISTR-NARR-3.PL.NSBJ-PL-SG-say
 'When they came to him he said to them' (Acts 20:18)

b. *Hesüana me-'u'-axüa-ca müme, müpaü*
 to.him 3.PL.SBJ-VIS-arrive.PL-SS men thus
me-te-ni-ta-hüave
 3.PL.SBJ-DISTR-NARR-SG-say
 'When the men had come to him they said' (Luke 7:20)

However, definitions of switch reference based on canonical switch reference and on purely morphosyntactic criteria (most notably Finer 1984, 1985) have now long been challenged and revisited to account for the fact that several languages use switch reference markers regardless of subject co-reference, to indicate some other type of (dis)continuity. This discussion is extraordinarily reminiscent of previous literature on the functions and origin of Early Slavic participle constructions,⁷ and the parallels between the use of participles in Early Slavic and switch-reference systems in the world's languages are, in fact, several.

In her seminal account of switch reference and discourse representation, Stirling (1993) defines switch reference more broadly as a 'clause level function which does not deal with the reference of NPs as such but with degrees and types of cohesion between eventualities' (Stirling 1993: 136; similarly Mithun 1993 and more recently McKenzie 2012, 2015a,b and Keine 2013) so that while referential (dis-)continuity is the basic meaning marked by all switch-reference systems, these may be seen as 'types of clause linkage rather than as mere devices of referential tracking' (Stirling 1993: 151). She identifies at least six main factors, or 'pivots', playing a role in the use of switch-reference markers among the world's languages, which provide a useful basis for comparison with the different contexts in which Early Slavic dative absolutes as opposed to conjunct participles are used:

- (1) Referential (non-)identity between one or more NPs. This largely corresponds to Haiman & Munro's (1983) definition of canonical switch reference, although Stirling's approach to referential continuity also encompasses cases in which what is tracked are syntactic roles other than subjects (e.g. objects) or semantic roles (e.g. agents).

⁷Corin (1995: 268), for example, argued that 'exceptions to the subject non-identity condition imply neutralization of the sole syntactic feature which differentiates the function of absolute and non-absolute participial constructions', and similarly Večerka (1996: 190); Bauer (2000: 280) considers the lack of co-reference with the matrix subject in Old Church Slavonic as evidence that 'absolute constructions were fully understood' and therefore not a Greek imitation. Collins (2004, 2011) instead challenges views based on allegedly 'canonical' constraints on subject non-coreference, providing instead possible explanations based on discourse-structuring criteria.

- (2) The agentivity of important event participants: ‘the SS/DS markers signal not simply continuity of reference, but more complexly, continuity of reference and of agentivity value: SS always implies both same referent and same agentivity or control value (whether high or low), while DS marks a discontinuity in either reference or agentivity value (high to low or low to high)’ (Stirling 1993: 150; see also Foley & Van Valin 1984).
- (3) The time of the event: temporal shifts between situations or ‘from one situation or event complex to another’ (Stirling 1993: 220) can trigger DS marking (similarly also Roberts 1988a: 98–100). Conversely, temporal continuity may license SS marking even without referential identity (see Mithun 1993: 126; van Gijn 2016a: 35).
- (4) The location of the event: a shift in location can trigger DS marking. This seems to occur most frequently with verbs of motion (cf. also Roberts 1988a: 60).
- (5) Mood of the clause, whereby DS marking may be used to indicate a shift between realis and irrealis mood. Related to this are shifts in what is referred to as ‘actuality’ (cf. Kaplan 1989: 19), which a change between current and reported speaker can indicate.
- (6) ‘Continuance or shift out of a cohesive sequence of events’. This includes DS marking used for signalling an unexpected, ‘surprise’ change in the course of events (see also Roberts 1988a: 98–100), as well as a change in *grounding* (Grimes 1975; Hopper 1979 [2020]) or *transitivity* (in the sense of Hopper & Thompson 1980).

It is interesting to see that all of these factors, to different extents, also explain several of the occurrences of both ‘canonical’ and ‘non-canonical’ participle constructions, particularly dative absolutes whose subject is co-referential with the one of the matrix clause.

Point (2), namely a change of agentivity as a possible trigger for DS marking, has also been pinpointed as one of the most common explanations for co-referential dative absolutes in (Old) Church Slavonic. Collins (2011: 110–112), for example, argues that a change in semantic role or shift in the degree of agentivity can warrant the use of an absolute construction, as (17) and (18) exemplify. Similarly, Kurešević (2006: 105) observes that some of the several occurrences of co-referential dative absolutes in medieval Serbian/Serbian Church Slavonic are motivated by the different semantic roles of the subjects in the absolute and in the matrix clause. In languages with switch reference, it is not uncommon to find comparable instances. In Eastern Pomo, for example, changes in agentivity regularly license the use of DS markers, despite subject co-reference, as in (19).⁸

⁸The opposite also occurs, i.e. continuity in agentivity licenses the use of SS even if the subjects are not coreferential).

- (17) *i bystъ idoštemъ imъ. ištistišę*
 and happen.AOR.3.SG go.PTCP.IPFV.DAT.PL 3.PL.DAT cleanse.AOR.3.PL
sę
 REFL
 ‘And it came to pass that, as they went, they were cleansed’ (Luke 17.14)
- (18) *i prišedšju vyšegorodu razbolě sę*
 and arrive.PTCP.PFV.M.DAT.SG Vyšegorod.DAT get sick.AOR.3.SG REFL
velmi.
 very
 ‘And when he arrived in Vyšegorod, he got very sick’ (*Primary Chronicle*, Codex Laurentianus f. 54d)
- (19) Eastern Pomo (Pomoan, North America; from McLendon 1978, cited in Roberts 2017: 549)
- a. *Há: káluhu-y si:má:mérqaki:hi*
 1.SG.A go.home-SS went to bed
 ‘I went home and then went to bed’
- b. *Há: káluhu-qan mí:p mérqaki:hi*
 1.SG.A go.home-DS 3.SG.A went to bed
 ‘I went home and he went to bed’
- c. *Há: xá: qákki-qan wi qa:lál tá:la*
 1.SG.A water bathe-DS 1.SG.U sick become
 ‘I took a bath and got sick’

The Eastern Pomo examples show some of the contexts in which DS and SS markers are used. (19a) is a typical, ‘canonical’ example of SS marking, indicating referential identity between the subjects of the two clauses. (19b), conversely, is a typical example of DS marking, indicating referential non-identity between the subjects of the two clauses. (19c) is an example of ‘non-canonical’ switch-reference marking: DS is used even if the subjects of the two clauses have the same referent, which can be explained with a shift in agentivity or semantic role of the subject referent, from agent/actor to experiencer/undertaker (Roberts 2017: 549).

A similar explanation can be given to both the Old Church Slavonic example (17) and the Old East Slavic example (18), where the subjects of the absolute construction and of the matrix clause are co-referential, but a shift in the agentivity of the subject can be observed. Note how, incidentally, the matrix clause in the Old East Slavic example in (18) and that in the Eastern Pomo example (19c) have verbs with the same meaning in the two languages (i.e. ‘get sick’, a typically non-agentive predicate).

Point (3) and (4), namely temporal (dis-)continuity between situations and shifts in the location of an event as factors in the use of a switch-reference marker, also find some

parallels in the use of absolute constructions as opposed to conjunct participles in Early Slavic. Among the world's languages, this is the case, for instance, in Amele, which may use DS marking even with subject co-reference if a new temporal setting or location is introduced, as the examples in (20) exemplify.

(20) Amele (Nuclear Trans New Guinea; from Stirling 1993: 114)⁹

- a. *Eu 1977 jagel November na odo-co-b cul-ig-Ø-en*
 that 1977 month November in do-DS-3.SG leave-1.PL-3.SG-REMP
 'That was in November 1977 that he did that and then he left it for us'
- b. *Age ceta gul-do-co-bil l-i bahim na tac-ein.*
 3.PL yam carry-3.SG-DS-3.PL go-PRED.SS floor on fill-3.PL.REMP
 'They carried the yams on their shoulders and went and filled up the yam store'

According to Roberts (1987, 1988a,b), the DS marker on *odo-co-b* 'did' in (20a) can be explained with the fact that there is a temporal shift between the first and the second clause, as also signalled by the time adverbial 'in November 1977'. Similarly, the DS marking in *gul-do-co-bil* 'carried' in (20b), despite its subject being coreferential with the one of the matrix clause, can be motivated by the change in place from the DS-marked clause to the following. In the words of Roberts (1988a: 60):

The explanation given by native speakers for such instances is that "something has changed" or this is "a new situation", and often it is obvious that the change being indicated is deictic rather than syntactic and that these deictic changes are in the area of world, time, or place reference points. For example, a change of time marked by the SR system is often backed up by a temporal expression; a change of place marked by the SR system occurs most frequently with verbs of motion, and a change of location can also be indicated by a locative expression; a change of world marked by the SR system is normally a switch from an intended or proposed action to the real action itself or vice versa

Many co-referential dative absolutes in the corpus, particularly in Old East Slavic, fit Roberts's (1988a) above characterization, whereby a change in place (often accompanied by a change in time setting) can motivate the use of DS marking. (21)-(23) are all co-referential dative absolutes from Old East Slavic texts where movement of some kind from one setting to another is involved.

- (21) *i všedšim v gorodž utěšista*
 and enter.PTCP.PFV.DAT.PL in city.ACC.SG console.AOR.3.DU
volodimercě
 people of Vladimir.ACC.PL
 'And when they entered the city, they consoled the people of Vladimir' (*Suzdal Chronicle*, Codex Laurentianus f. 126r)

⁹These examples are originally from Roberts (1988a: 61), but I am here using the more explicit glossing by Stirling (1993), which, for example, also adds glosses for zero-marking on third persons singular.

- (22) *i v̄ jedin̄ ot dnii š̄d̄š̄ju k̄*
 and in one.M.ACC from day.M.GEN.PL go.PTCP.PFV.M.DAT.SG to
tomu blagomu i bonos̄nomu oc̄ju
 that.M.DAT.SG good.M.DAT.SG and God-bearing.M.DAT.SG father.M.DAT.SG
našemu feodosiju. i jako v̄nide v̄ chram̄
 our.M.DAT.SG Feodosij.DAT and as enter.AOR.3.SG in palace.ACC.SG
ideže b̄ knęz̄ s̄de. i se
 where be.IMPF.3.SG prince.NOM.SG sit.PTCP.IPFV.M.NOM.SG and behold
vid̄ mnogyya igrajušta před̄ nim̄
 see.AOR.3.SG much.M.ACC.PL play.PTCP.IPFV.ACC.PL before 3.SG.M.INST
 ‘Once our good and God-bearing father Theodosius came to the prince and as he entered the palace where the prince was, he saw many musicians playing in front of him’ (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 59g)

- (23) *polovci že gjurgevi ni po strěl̄*
 Polovtsy.NOM.PL PTC Yuri’s.NOM.PL NEG with arrow.DAT.SG
pustiv̄še togda poběgoša. a potom olgoviči.
 release.PTCP.PFV.NOM.PL then run.AOR.3.PL but after Olgoviči.NOM.PL
a potom gjurgi z dětmi. i běžašt̄m
 but after Yuri.NOM with child.INST.PL and run.PTCP.IPFV.DAT.PL
im̄ čres̄ rut̄. mnogo družiny
 3.PL.M.DAT through Rut’.ACC much.N.NOM.SG retinue.GEN.SG
potope v rutu. b̄ bo grezok̄. i
 drown.AOR.3.SG in Rut’.LOC be.IMPF.3.SG because swampy.NOM.SG and
běžašt̄m im̄ ověch̄ izbiša. a
 run.PTCP.IPFV.DAT.PL 3.PL.M.DAT this.M.GEN.PL kill.AOR.3.PL but
drugyya iz̄vimaša. i poludne priběgoša ko
 other.M.ACC.PL capture.AOR.3.PL and noon.LOC come running.AOR.3.PL to
dněpru.
 Dnieper.DAT

‘Yuri’s Polovtsy, without shooting arrows at that time, ran, and then the Olgoviči, and then Yuri with his children. And when they fled through the [river] Rut’, several of the retinues drowned in Rut’ because it was swampy. And as they were fleeing, they killed some and captured others. And by noon, they had run to the Dnieper’ (*Suzdal Chronicle*, Codex Laurentianus f. 111v)

(21)-(23) are only some of the several examples of coreferential dative absolutes introducing a shift in setting. The absolute in (21) introduces the arrival of Yaropolk Rostislavich in Vladimir, where he becomes Grand Prince, after travelling from Moscow and, before that, Chernihiv. In (22), Feodosij Pečerskij (Theodosius of Kyiv) arrives to visit his brother, the Grand Prince Iziaslav Yaroslavich of Kyiv. The example is from a series of passages in which the narrative goes back and forth from the prince’s palace to Feodosij’s monastery. In (23), where the broader context is also provided, the rapid shifts in location are clear from the repeated verb (*po*)*bežati* ‘run, flee’, following the movement of Yuri and Yuri’s Polovtsy and of the Olgoviči through the river Rut’ till the Dnieper.

Point (5) from Stirling's (1993) pivots, namely shifts in the mood or actuality from the marked clause to the next, also finds a potential parallel in Early Slavic. This is specifically in regard to Stirling's observation that a shift in actuality can be seen in correspondence to changes between current and reported speakers, a change which can be marked by DS. In Early Slavic it is relatively common to see dative absolutes introducing reported speech, even if the subject of its syntactic matrix is coreferential with the subject of the absolute. (24)-(27) are some of such examples.

- (24) *onomu že rekšu něstō*
 that.M.DAT.SG PTC say.PTCP.PFV.M.DAT.SG not be.PRS.3.SG
došla zapadnyichō stranō arianōskaja
 reach.PTCP.RESULT.F.NOM.SG Western.GEN.PL land.GEN.PL Arian.NOM.SG
chula
 controversy.NOM.SG

'He said that the Arian controversy had not reached the Western lands' (*Vita of Isaakios of Dalmatos*, Codex Suprasliensis f. 99r)

- (25) *prišedšju že igumenu ko mně i*
 arrive.PTCP.PFV.M.DAT.SG PTC abbot.M.DAT.SG to 1.SG.M.DAT and
rekšju mi poidevě v pečeru k
 say.PTCP.PFV.M.DAT.SG 1.SG.M.DAT go.IMP.1.DU in cave.ACC to
feodosjevi [...] onomu glštju ko
 Feodosij.SG.M.DAT [...] that.M.DAT.SG say.PTCP.IPFV.M.DAT.SG to
mně oudariša v bilo. mně že
 1.SG.M.DAT hit.AOR.3.PL in semantron.ACC 1.SG.M.DAT PTC
rekuštju prokopachō ouže
 say.PTCP.IPFV.M.DAT.SG dig.AOR.1.SG already

'So, the abbot came to me and said: "Let's go to the cave to Feodosij" [...] And when he told me: "They hit the semantron," I said: "I finished digging"' (*Primary Chronicle*, Codex Laurentianus f. 70b)

- (26) *onēmō mnjaštemō jako bratii polunoštbnōje*
 that.DAT.PL think.PTCP.IPFV.DAT.PL that brethren.DAT midnight.ACC.SG
pěnije sōvrvšajuštemō. i tako paky otidoša.
 prayer.ACC.SG finish.PTCP.IPFV.DAT.PL and so again leave.AOR.3.PL
čajušte donōdeže sii sōkončajutō pěnije.
 wait.PTCP.IPFV.NOM.PL until this.M.NOM.PL finish.PRS.3.PL singing.ACC
i togda vōšbdōše vō crkōvō pojemljutō
 and then enter.PTCP.PFV.NOM.PL in church.ACC take.PRS.3.PL
vse sūštaja vō i
 all.N.ACC.PL be.PTCP.IPFV.N.ACC.PL in 3.SG.F.ACC

'They thought that the brethren was [still] finishing singing the midnight prayers, so they [the robbers] left again, waiting until they finished singing, when they could then enter the church and take everything that was in it' (*Life of Feodosij Pečerskij*, Uspenskij Sbornik f. 46g)

- (27) *otcъ* *že* *našъ* *sergii* *poklonъ*
 father.NOM.SG PTC our.M.NOM.SG Sergij.NOM bow.PTCP.PFV.M.NOM.SG
se *i* *reče* *jako* *gvi* *godě* *tako* *i* *budi*
 REFL and say.AOR.3.SG that lord.DAT pleasant so then be.IMP.3.SG
blsvenъ *gъ* *vo* *věki.* *i* *vsěmъ*
 blessed.M.NOM.SG lord.M.NOM.SG for ever and all.DAT.PL
rekšimъ *aminъ*
 say.PTCP.PFV.DAT.PL Amen
 ‘Our father Sergij bowed and answered: “As the Lord pleases, so be it; blessed be the Lord forever!”, and they all said “Amen!”’ (*Life of Sergij of Radonezh* f. 102v)

In (24) the dative absolute *rekšu* ‘(he) said’ has no apparent matrix clause and, in light of similar non-canonical usages of switch-reference markers in the world’s languages, its use could then be explained by the shift to reported speech that the absolute introduces.

The absolutes in (25) introduce the reported speech of two different protagonists (one of them in the first person) and are part of a longer dialogue chain.

In (26), the absolute introduces the thoughts of the subject. In this case, the reported thought is introduced as an indirect speech, as *jako* signals, and this time the absolute clause is overtly coordinated with the matrix clause, as is also stressed by the discourse-coordinating connective *tako* ‘so, thus’.

Finally, in (27), the post-matrix, overtly coordinated absolute *rekšimъ* ‘(they) said’ does not seem to have any other apparent function but introducing a reported speech. The speaker introduced by the absolute *is* different from the previously reported speaker, which was, in fact, introduced by the finite clause *reče* ‘(he) said’.

Turning now to the final point from Stirling’s (1993) pivots, the use of switch reference to signal a change in *grounding*, namely from *foregrounding* to *backgrounding* or vice-versa, is, as we have repeatedly seen, the main constant function of dative absolutes and conjunct participles emerged from the previous chapters. In Chapter 2, we have also seen that even in post-matrix position the temporal relation of dative absolutes to their matrix clause is most typically consistent with the temporal relation they have with the matrix when they are fronted, i.e. as typical FRAMES. Given the attested, albeit much less frequent, use of post-matrix absolute constructions as ELABORATIONS and the fact that these are all non-coreferential with the subject of their matrix clause, we can consider referential (non-)identity as the basic criterion, as it were, whereby an absolute is chosen over a conjunct participle and vice-versa. That is, given a discourse function, if both conjunct participles and absolute constructions are equally viable means for fulfilling that function, then the former is used when the subject referents are the same, the latter if they differ. This is in many ways similar to what has been claimed about switch reference systems cross-linguistically, namely that ‘all other things being equal, the same referent marker [...] appears in situations of referential identity between the subject participants of two clauses, and the different referent marker [...] when there is no such referential identity.

However, other factors, to do with discourse cohesion may overrule this basic set-up'¹⁰ (van Gijn 2016a: 3). Regarding referential (non-)identity as the baseline of all switch-reference systems, Stirling (1993: 151), after listing the six 'pivots' for switch-reference marking reported above, claims that all switch-reference systems mark at least point (1), namely referential (non-)identity. However, the common cross-linguistic coverage of all six pivots indicates that switch-reference may be regarded as a type of clause linkage and not simply a referential-tracking device.

At an event-ontological level, it has been observed that there are clear ties between disjoint reference and backgrounding and between coreference and foregrounding. Building on Hopper & Thompson's (1980) notion of 'transitivity' as a value assignable to clauses on the basis of properties such as the number of participants (i.e. the basic notion of transitivity), the *Aktionsart* (e.g. 'kinesis', i.e. states versus non-states; 'punctuality', i.e. punctual versus durative, or 'nonpunctual') or the individuation of the event participants, Stirling (1993: 147–150) highlights the apparent correlation between backgrounding and low transitivity, and between foregrounding and high transitivity, adding that

chains of foregrounded clauses (within one episode) typically continue to talk about the same participants rather than introducing new ones, thereby maintaining same subject. Furthermore, chains of foregrounded clauses tend to present sequential events, while simultaneous events (along with states and on-going or repeated events) appear in background clauses.

This is an observation that can be easily applied to INDEPENDENT RHEMES (foregrounded clauses), typically encoded by conjunct participles and typically involving the same participants as their matrix (and any other conjunct participle dependent on it) and presenting sequential events, as opposed to FRAMES (backgrounded clauses), more typically encoded by absolute constructions and normally involving overlap/simultaneity/inclusion. The similarity between absolute constructions and switch-reference markers had, in fact, been observed before by Bickel (1999), who argued that many formal synchronic switch-reference systems have originated from absolute constructions 'in pragmatic competition with conjunct participles (*participia coniuncta*) that show case agreement with a coreferential argument of the matrix, occupy roughly the same adsentential position as absolutes, and fulfill a similar discourse function' (Bickel 1999: 46), which, Bickel argues, is the grammaticalization paths followed by the switch-reference systems of languages in the Muskogean, Yuman and Uto-Aztecan families.

There are thus clearly stronger connections between DS markers in switch-reference systems and absolute constructions, and between SS markers and conjunct participles, than the mere marking of referential (non-)identity. The languages mentioned in this

¹⁰I have masked the reference to the specific markers in Tsafiki, namely SS *-to* and DS *-sa*, from the source since it is clear from van Gijn's (2016a) overview that this statement is generally applicable to switch-reference systems cross-linguistically, where participant (non-)identity is the basic factor even when several other discourse variables can supersede those factors.

section that are also in our dataset, namely Amele and Huichol, had already been assigned to the same pattern (B) as Old Church Slavonic in the classification in the previous Chapter. Their Kriging maps, particularly the one for Huichol are, in fact, quite similar to the Old Church Slavonic one (Figure 6.1a above), as Figures 6.15a and 6.15b show.

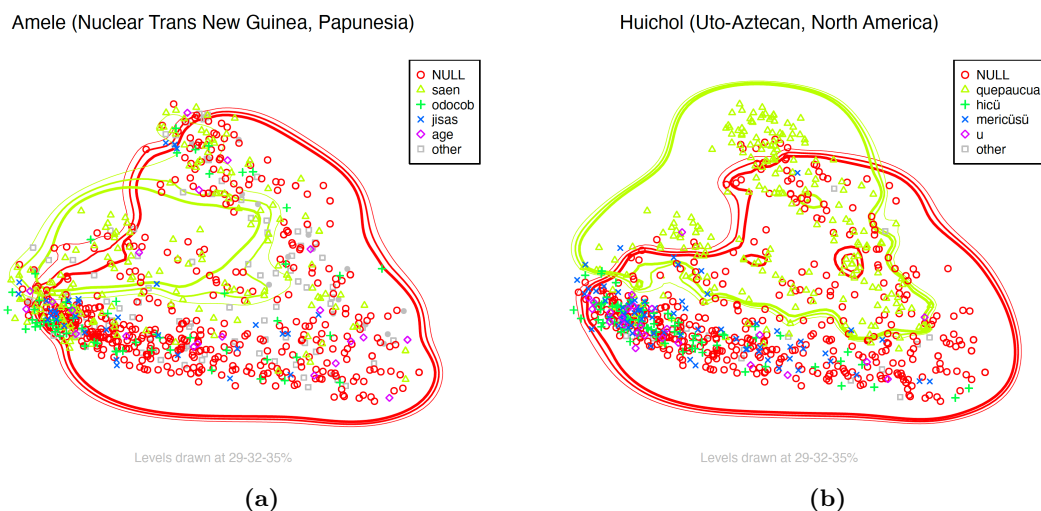


Figure 6.15: Kriging maps for Amele (Nuclear Trans New Guinea, Papunesia) and Huichol (Uto-Aztecan, North America)

Since both Amele and Huichol, as well as Cofán (A’ingae) from the previous section, also use converbal/medial forms marked for switch reference as temporal clauses, in an ideal scenario, a complete semantic map for WHEN would also include different switch-reference markers. This would allow us to capture the competition not only between potential subordinating conjunctions (if present at all) and null constructions, but also between different null constructions. Doing this at scale would require detailed annotation, at least, at the morphosyntactic level, as the one we currently have for Old Church Slavonic and Ancient Greek in the PROIEL treebanks, which did, in fact, allow us to add a further layer of distinctions in their semantic map. Gathering such information for all the languages in the dataset would be hardly feasible, but a preliminary, automatic test on Amele, Huichol, and Cofán indicates that such direction could lead to very informative maps and may thus be worth pursuing for more languages. All three languages have relatively isolable switch-reference morphemes. Amele has *-co/-ce-* + personal endings for DS and *-me-* + personal endings for SS.¹¹ Huichol has *-ku* and *-ka* (spelled as *-cu* and *ca* in our dataset) for DS and SS respectively (Comrie 1983). Cofán has *-si* for DS and *-pa/-mba* for SS (AnderBois et al. 2023). Similarly to what did for Old Church Slavonic and Ancient Greek participles in Figures 6.11a and 6.11b, the placeholders ‘DS’ and ‘SS’ were

¹¹Following Roberts (1988a: 48) summary of endings, this gives the following possible markers for DS: *-comin*, *-com*, *-cob*, *-cohul*, *-cobil*, *-comun*, *-cemín*, *-cem*, *-ceb*, *-cehul*, *-cebil*, *-cemun*; and these for SS: *-meig*, *-meg*, *-meu*, *-meb*, *-mesi*, *-mei*. Note that these are only the ‘sequential’ switch-reference endings in Amele, namely those that are used to express the sequentiality of events. Those expressing simultaneity are not isolable as they involve stem reduplication.

inserted before any word in the corpus ending with the respective markers in each of these languages, thus allowing the model to capture the placeholders as dummy subordinators. After realigning the processed texts to the English version and extracting *when* and its parallels in the three languages, we can superimpose the occurrences of *when* aligned to the dummy switch-reference subordinators on the raw MDS map. Figures 6.16a-6.16c show the results for the three languages. Based on the previous Kriging analysis, Amele and Huichol also have lexified *when*-counterparts, whereas Cofán does not. These are, therefore, also superimposed on the map for the former two to allow for easier comparison with the maps for Old Church Slavonic and Greek in Figures 6.11a and 6.11b.

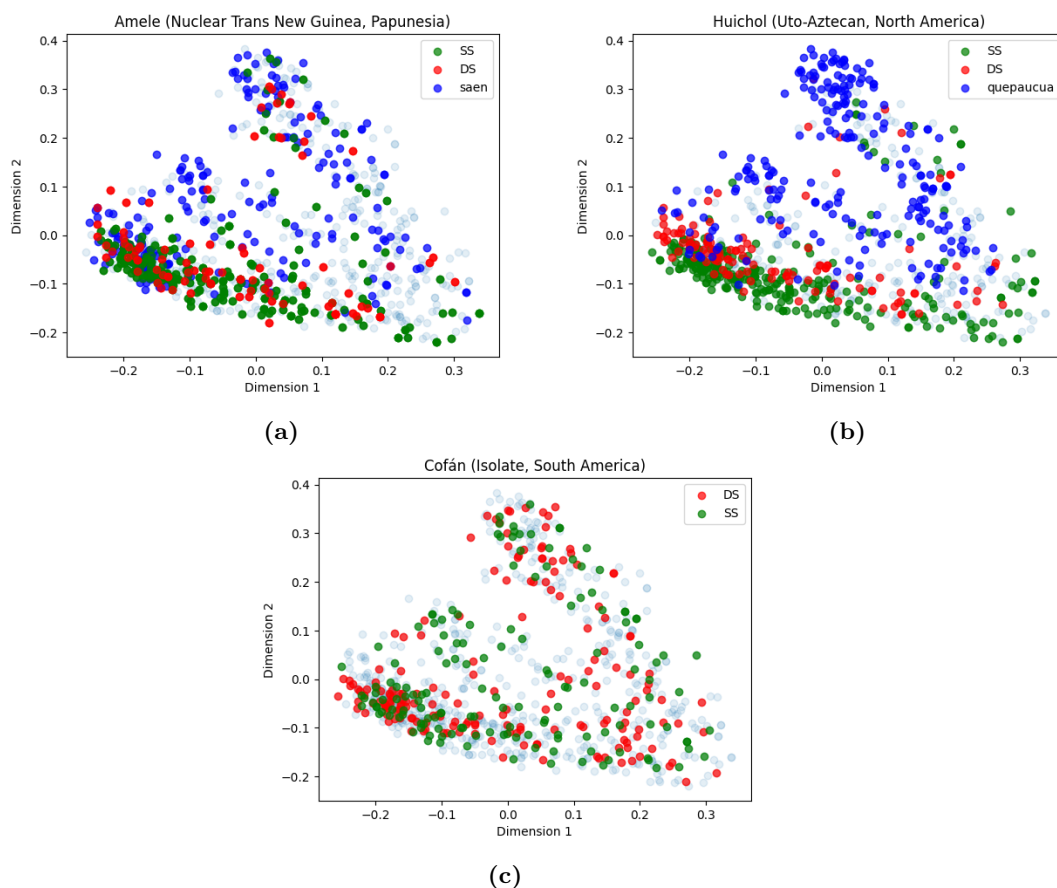


Figure 6.16: MDS map of *WHEN* highlighting overt subordinators and switch reference markers in (a) Amele (Nuclear Trans New Guinea, Papunesia), (b) Huichol (Uto-Aztecan, North America) and (c) Cofán (Isolate, South America).

The result for Cofán indicates that switch-reference markers in this language have a much wider scope and are likely to be used with different functions than Old Church Slavonic absolutes and conjunct participles. Judging from the Kriging map shown in the previous section, Cofán does not appear to have a lexified counterpart to *when* but instead uses switch-reference markers for the whole semantic space in the *WHEN*-map.

On the other hand, the maps for Amele and Huichol, but particularly the latter, bear great similarities with the ones we have seen for Old Church Slavonic and Ancient

Greek in Figures 6.11a and 6.11b. At least in Huichol, we observe that same continuum from *when*-counterparts (*quepaucua*) at the top of the map to SS markers at the very bottom (corresponding to what the highlighted conjunct participles in the Old Church Slavonic and Ancient Greek maps), with DS markers also clustering in the bottom area, but in a relatively defined band above SS markers, similarly to the position occupied by absolute constructions in Figures 6.11a and 6.11b. In the Amele map, this continuum is less neat but does overall reflect a very similar distribution between switch reference markers and *when*-counterpart (*saen*). Besides, the Amele map only includes switch-reference markers indicating sequentiality, not those conveying simultaneity, for reasons of isolability of the morphemes, as explained earlier. Further annotation may thus improve the map considerably.

6.3.4 Insubordination

As we saw in Chapter 2, seemingly independent dative absolutes often become clearly backgrounding devices for some portion of discourse to come when units larger than clauses and sentences are taken into consideration (cf. Collins 2011), and this may well be the case for all the examples of seemingly independent absolutes seen above, as in (24)-(27).¹² However, what is interesting to notice there is that, even at the inter-clausal level, such ‘non-canonical’ instances of dative absolutes could be motivated as devices to indicate a shift in actuality, as Stirling (1993) puts it, and that these instances have cross-linguistic counterparts in the way switch reference systems may use DS markers for goals other than subject co-reference. Yet, both functions, namely the use of seemingly independent absolutes as background for discourse segments larger than sentences and as markers of shifts in actuality, are by no means mutually exclusive. Similar functions have been observed cross-linguistically on so-called ‘in subordinate’ constructions (Evans 2007; Evans & Watanabe 2016).

Insubordination is widely attested among the world’s languages independently of switch-reference systems and can be defined, from the synchronic perspective, as ‘the independent use of constructions exhibiting prima facie characteristics of subordinate clauses’ (Evans & Watanabe 2016: 2), representing a ‘redeployment of linkages from intra-clausal to general discourse links’ (Evans 2007: 370).

Within switch-reference systems, insubordination has been observed by Wilkins (1988: 154–156), for example, in Mparntwe Arrernte (Pama–Nyungan, Australia), where SS- or DS-marked clauses may occur on their own in a sentence without any apparent matrix clause (a phenomenon referred to as *trans-sentential switch-reference* by Wilkins), which, in most occurrences, is explained with the fact that ‘a main clause in a text or discourse can, later

¹²As with coreferential dative absolutes, syntactically independent dative absolutes have also been specifically discussed in previous literature as ‘non-canonical’ occurrences of the construction. On the topic of independent/coordinated absolutes, see, for example, Corin (1995: 262-276), Večerka (1961: 50–51), Alekseev (1987), Gebert (1987: 568), Remneva (1989: 35-39), Collins (2011: 113-126).

in the text or discourse, be referred to by an anaphor or can be ellipsed, as long as it is contextually recoverable' (Wilkins 1988: 154).

In the switch reference system of Yuman languages (Cochimí-Yuman), the SS marker *-k* is typically used, at the inter-clausal level, to represent closely related events as belonging to a higher level event (i.e. the typical stacking of clauses in chains to form a sentence), while the DS marker *-m* is used to link clauses 'representing less closely related event or states' (Mithun 2008: 100; see also Winter 1976; Slater 1977; Langdon & Munro 1979; Hardy 1982; Gordon 1983). However, some languages within the family (in the Hualapai branch) have extended the use of the markers to 'marking relations of full sentences to each other in larger discourse contexts, providing higher-level chunking of experience into events' (Mithun 2008: 100), which is very similar to what has been claimed about non-canonical dative absolutes in studies such as Collins (2011).

Insubordination with functions similar to those observed on non-canonical absolutes is well-attested also beyond switch-reference markers. The adverbializer *=go* in Navajo (Athabaskan-Eyan-Tlingit, North America), for example, which marks the subordinate relations of adverbial clauses within sentences often also 'specifies relations of sentences to larger stretches of discourse' (Mithun 2008: 72).

In Navajo narrative discourse, we can find a series of clauses, seemingly independent, some marked with *=go*, some unmarked (i.e. formally independent). According to Mithun, the latter advance the storyline, whereas the former 'provide background, incidental information, explanation, and emotional evaluation' and, as observed by Mithun, can thus be reconducted within models of narrative structure (e.g. Hopper 1979 [2020]; Chafe 1994) to foreground and background, respectively.

Similarly, in Central Alaskan Yup'ik (Eskimo-Aleut), participles can be used in subordination to 'provide background information, parenthetical comments off the event line, elaboration, and evaluation' and are 'often used when speakers open narratives or episodes by setting the scene' (Mithun 2008: 90). As pointed out by Heine, Kaltenböck & Kuteva (2016: 51), the functions of in subordinate constructions observed by Mithun are described in the literature on parentheticals as 'meta-discursive' (e.g. Hansen 1998: 236), 'meta-textual' (e.g. Traugott & Dasher 2002b: 155), and 'meta-communicative' (e.g. Grenoble 2004: 1953). This is precisely how Collins (2011) describes one of the main functions of 'independent' participle constructions. (28)-(30) are examples of independent participle clauses which, as Collins (2011: 120-122) observes, function as 'meta-discursive', 'meta-comment', or 'meta-textual' devices.

- (28) *troje jestv nemoštvo mi rozuměti a četvora ne*
 three be.PRS.3.SG impossible 1.SG.DAT understand.INF and fourth NEG
rozumějo slěda orlou lěteštou
 understand.PRS.1.SG track.GEN.SG eagle.DAT.SG fly.PTCP.IPFV.M.DAT.SG
xristosovo vřšstije i potii zmiině po
 Christ's.NOM.SG ascension.NOM.SG and path.GEN.PL serpent's.GEN.SG over

kameni *dijavolъ* *ne* *obrěte* *bo* *slěda*
 rock.LOC.SG devil.NOM NEG find.AOR.3.SG because trace.GEN.SG
grěxovъnaago *na* *tělesi* *xsvě* *i* *pouti*
 sin's.GEN.SG on body.LOC.SG Christ's.LOC.SG and path.GEN.SG
lodiĵa *po* *vodě* *plovoušti* **crъkvi**
 boat.F.GEN.SG through water.LOC float.PTCP.IPFV.F.GEN.SG Church.DAT
aky *vъ* *poučině* *žitъja* *sego* *naděždeĵo*
 like in gulf.LOC.SG life.GEN.SG this.GEN.SG hope.F.INST.SG
jaže *vъ* *xsa* *krъsmъ*
 which.F.NOM.SG in Christ.GEN cross.INST.SG
pravimě *i* *poutii* *mouže* *vъ*
 govern.PTCP.PASS.IPFV.F.DAT.SG and path.GEN.PL man.M.GEN.SG in
ĵonosti *roždenaago* *otъ* *světaago* *dxa* *i* *otъ* *dvca*
 youth.LOC born.M.GEN.SG from holy.GEN spirit.GEN and from virgin.GEN
 'Three things it is impossible for me to understand, and a fourth I do not
 understand: the track of an eagle flying (Christ's Ascension); and a serpent's
 paths over a rock (the Devil; for he did not find a trace of sin on Christ's body);
 and the path of a boat sailing through the water (like the Church being governed
 in the gulf of this life by hope in Christ, by the Cross); and the paths of a man in
 youth (the One born of the Holy Spirit and the Virgin)' (*Izbornik of 1073*, f.
 156c1–20)

- (29) a. *mol* ***ležjošte*** *na* *loži*
 prayer.NOM.SG lie.PTCP.IPFV.NOM.PL in bed.LOC
 'A prayer [when people are] lying in bed' (*Euchologium Sinaiticum* f. 37a)

- b. *mol* *egda* ***chotešte*** *vinogradъ*
 prayer.NOM.SG when want.PTCP.IPFV.NOM.PL vineyard.ACC.SG
saditi
 plant.INF
 'A prayer when [people are] wanting to plant a vineyard' (*Euchologium
 Sinaiticum* f. 13b)

- (30) *ot* *mě:stę:tb :* *ko* *ga:vo:šb :* *i :* *ko :* *so:di:lb :*
 from Mestjata.GEN.SG to Gavša.DAT.SG and to Sdila.DAT.SG
po:py:ta:i:=ta *mi :* *konę :* *a :* ***mě:stę:ta :*** ***sę :***
 find.IMP=PTC 1.DAT.SG horse.GEN.SG and Mestjata.NOM.SG REFL
va:ma ***poklanę :*** *a:že :* *va :* *cv:to :* *na:do:bb :*
 2.DAT.DU bow.PTCP.IPFV.M.NOM.SG if 2.DAT.DU something needed
a *solī=ta :* *ko monъ :* *a* *gramotuo :* *vodai=ta* *a*
 and send.IMP=PTC to 1.DAT.SG and letter.ACC.SG give.IMP=PTC and
uo *pavla :* *skota* *poprosi:=ta* *a* *mъ:stę*
 from Pavel.GEN.SG money.ACC.SG ask.IMP=PTC and Mestja[ta]
 'From Mestjata to Gavša and to Sdila. Get a horse for me. And Mestjata bows to
 you (both). If you need anything, send to me, and give (the messenger) a letter.
 And ask for money from Pavel, and Mestja[ta] (bows to you).' (*Birchbark letter
 N422*, translation by Schaeken 2018b: 146)

(28), from the *Izbornik of 1073* (also known as *Izbornik of Svjatoslav*)¹³, presents a passage from the Book of Proverbs (30:18–19) interpolated with exegetical commentaries (rendered in Collins’s (2011) translation with parentheses). These comments, including the third (the dative absolute), are clearly ‘syntactic parentheticals functioning as metadiscourse rather than as part of the basic text’ (Collins 2011: 121).

(29a-b) are standard examples of the rubrics with which *euchologia* introduce prayers to be recited aloud by the clerical readers in the specified occasion. In the *Euchologium Sinaiticum*, we also find several examples of the meta-comment *imę rekъ* ‘having said the name’ in correspondence of names to be said out loud throughout the prayer (ten occurrences, according to Nahtigal 1942: xxxix, as reported in Collins 2011: 121). As Collins (2011: 121) points out, ‘the indirect command conveyed by *imę rekŭ* can be understood as a continuation of this same implicit speech act [the one conveyed by the rubrics]: Here say the beneficiary’s name, and then continue the prayer’.

(30) contains the nominative absolute *a męstęta sę vama poklanę* ‘and Mestjata bows to you (both)’, where *you* refers to addressees of the letter (Gavša and Sdila) and the clause is an instruction to the messenger to bow to the addressees on Mestjata’s behalf, which reflects the common practice of reading the letters aloud to the addressees (Zaliznyak 2004: 297; Schaeken 2018b: 11-12). The metatextual function of the participle construction in this example is backed up by several other occurrences of instructions with similar functions (cf., for example, birchbark letters N354 and N358).

All these examples indicate that the ‘non-canonical’, independent use of participle constructions corresponds, in fact, to a rather well-attested and well-understood pattern.

6.4 Summary

This chapter addressed the issue of understanding the functions of null constructions in the semantic map of WHEN, started in the previous chapter, not only at the level of their competition with *when*-counterparts, but also with respect to the competition between different null constructions, which the semantic map of WHEN, without further annotation, could not address.

Section 6.1 looked at the distribution of absolute constructions and conjunct participles in the map for Old Church Slavonic by adding targeted annotation, allowing to distinguish the two constructions among otherwise unspecified null alignments to *when*. The analysis showed that, as anticipated in the previous chapter, both absolutes and conjunct participles cluster at the bottom of the map and occupy rather contiguous than fully overlapping areas, with absolutes relatively closer to the main *jegda*-cluster at the top of the map. Adding *while* to the Kriging maps indicated that absolutes in Old

¹³The *Izbornik of 1073* is not part of the datasets used in this thesis. The passage and translation are from Collins (2011: 120), but transcription and glossing are mine and follow the photographic edition in Morozov (1880).

Church Slavonic cover much of the same ground as English *while*, unlike Greek, whose Kriging for area for absolute constructions, both in the *when*-only and in the *when-while*-map, corresponds largely to uses of English *when*. This difference corresponds to the observation made in Chapter 1 on Old Church Slavonic-Greek mismatches, which showed that there is a small but consistent group of constructions in Greek (*en tōi* + infinitive) that are translated into absolutes in Old Church Slavonic and convey temporal inclusion of the main eventuality in the eventuality introduced by the adverbial clause, similarly to the function of English *while*.

Section 6.2 looked into an example of language (Pular) with relatively well-defined lexified counterparts to *when* not only for the top and middle cluster as in several other languages (including Old Church Slavonic), but also for the bottom left cluster. Following grammatical descriptions of this language, functional labels were substituted for the particular connectives, with the aim of pinpointing functional differences between the top half and bottom half of the semantic map of WHEN, as well as specifically clarifying differences in the meaning of the middle-left cluster (corresponding to the Kriging area for Greek *hóte* and part of the Kriging area for Old Church Slavonic *jegda*) and the bottom-left cluster. The analysis confirmed, as was suspected in the previous chapter, that the bottom area of the semantic map, within which most participle construction fall in Early Slavic, corresponds to a relatively well-defined function from the perspective of temporal relations, namely that between existential-eventive clauses, which translates into the sequentiality between events being stressed.

Finally, Section 6.3.4 looked at various well-studied phenomena from languages that displayed similar patterns to Old Church Slavonic in the previous Chapter or that seemingly *only* use null constructions across the semantic map of WHEN, specifically clause chaining, bridging, switch reference and insubordination. The goal was to clarify the competing motivations governing the choice between different types of null constructions, as well as between null constructions and WHEN-counterparts that appear to be in intense competition with null constructions in the bottom region of the semantic map.

First, we substituted functional labels (INDEPENDENT RHEMES and FRAMES) in the Ancient Greek map (because of its greater coverage of the New Testament) for specific configurations of participle constructions known from previous chapters to be correlated with those functions. The maps generated with these functional labels indicated more clearly that typical framing participles (sentence-initial absolutes) and INDEPENDENT RHEMES (pre-matrix, aorist participles with an aorist matrix) dominate in contiguous, but not fully overlapping areas in the map, with FRAMES being relatively closer to finite competitors at the top of the map than INDEPENDENT RHEMES. The same pattern was observed by simply overlaying the observation for the three constructions to the raw MDS map.

The overview on clause chaining suggested that similar criteria of discourse organization govern the use of chaining constructions in clause-chaining languages and participle

constructions in Old Church Slavonic, including the possibility of *backgrounded* clauses alternating with *foregrounded* clauses, as well as overarching principles of tense-space-participant connectedness and event conflation/integration.

We then looked at how the ‘topic-shift’ function of dative absolutes mentioned in previous chapters, whereby elements of a foregrounded eventuality in the previous discourse get repurposed as background for a new portion of foregrounded discourse to come, has, in fact, clear parallels in bridging constructions as a systematic, formalizable phenomenon. In particular, we saw that formal accounts of bridging based on the SDRT rhetorical relations *Background* and *Narration*, which we already partially implemented in Chapter 5, are clearly applicable to several similar examples of absolute constructions in Old Church Slavonic.

The clearest parallels to the use of conjunct participles as opposed to absolute constructions, as well as the common thread between all the phenomena touched upon in this section, were observed in synchronic switch-reference systems across the world’s languages. Especially considering the areal and genetic unrelatedness of the languages mentioned in this overview to Old Church Slavonic, the similarities between the former and the latter were found to be quite striking. In particular, the automatic addition of switch reference markers to the semantic map of WHEN for Amele, Huichol and Cofán suggested that languages with a switch reference system *and* a lexicalized counterpart to *when* (i.e. in our case Amele and Huichol, not Cofán) show an extremely similar distribution of *when*-counterparts, DS-markers and SS-markers, approximately placed in a continuum comparable to *jegda*-clauses, absolutes, and conjunct participles in Old Church Slavonic. Overall, they provided some independent evidence that the use of absolute constructions as opposed to conjunct participles is regulated by very similar principles to the ‘pivots’ that play a role in the use of switch-reference markers cross-linguistically, as described in detail, among others, by Stirling (1993).

Finally, we saw how insubordination, namely the ‘independent use of constructions exhibiting *prima facie* characteristics of subordinate clause’ (Evans & Watanabe 2016: 2), and the independent uses of dative absolutes analysed by Collins (2011) also bear great similarities, particularly in light of the fact that insubordination and switch-reference are often related in the same languages.

Crucially, all these principles are also able to motivate several of the uses of ‘non-canonical’ participle constructions, particularly dative absolutes, which had otherwise often been written off as ‘aberrations’ or as signs of ‘decay’ by the earlier scholarship on Early Slavic (e.g., among many others, Večerka 1961: 49; Večerka 1996: 190; Corin 1995: 268).

Chapter 7

Conclusions

In the Introduction, the goal of this thesis was set up to be the study of ‘the semantic and pragmatic properties of Early Slavic conjunct participles and absolute constructions to understand what principles motivate their selection over one another and over *jegda*-clauses at the synchronic level’. The investigation, I believe, led to some genuine progress on several fronts and raised some new questions on others.

The stated goal was addressed by adopting two broadly different approaches: a quantitative, corpus-based one, exploiting data from Early Slavic treebanks, and a typological one, leveraging massively parallel corpora consisting of translations of the New Testament in over 1400 linguistic varieties, including Old Church Slavonic.

Since very early on in the investigation, it became clear that the competition between conjunct participles, absolute constructions, and *jegda*-clauses occurs at the level of discourse organization, where the major common denominator is that all of them may function as event-framing devices. The concept of ‘framing device’ was meant as a discourse segment that is used ‘to limit the applicability of the main predication to a certain restricted domain’, which is Chafe’s (1976: 50) definition of *frame setters* and was consistently associated with the *background* content of a (complex or elementary) discourse unit. The main contrast throughout the thesis, then, was between *background* clauses and discourse segments corresponding to the *foreground* content of a predication. The background-foreground contrast informed much of the discussion, encompassing all theoretical frameworks followed in this thesis.

From the investigation, the following findings can be highlighted.

7.1 Overview of the main findings

(a.) **Conjunct participles are typical foreground material**

Early Slavic conjunct participles can be both background and foreground clauses, but in their most typical functions they are part of the foreground content of a discourse unit. Like their Ancient Greek counterparts, following Bary & Haug (2011) and Haug (2012), they are most typically either INDEPENDENT RHEMES or ELABORATIONS, which

are both foregrounded material in the sense that they encode new information and are part of the main line of events, but with very different semantic properties. As INDEPENDENT RHEMES, they are discourse-coordinated and are typically interpreted as independent clauses, as the second and third conjunct participle in (1). As ELABORATIONS, they add granularity to the matrix event and are most often interpreted as manner or means adverbials, as in (2).

- (1) *borisa že [oubivše]_{frame} okaněni*
 Boris.SG.M.GEN PTC kill.PTCP.PFV.M.NOM.PL cursed.M.NOM.PL
[ouvertěvše]_{indrheme} v šaterě [vzloživše]_{indrheme} na
 wrap.PTCP.PFV.M.NOM.PL in tent.SG.M.ACC load.PTCP.PFV.M.NOM.PL on
kola povezoša i
 cart.SG.N.ACC take.AOR.3.PL 3.SG.M.ACC
 ‘After killing Boris, the raiders wrapped him in a canvas, loaded him on a cart, and took him off’ (*Primary Chronicle*, Codex Laurentianus 46b)

- (2) *stoěchu plačjušte se i*
 stand.IMPV.3.PL cry.PTCP.IPFV.M.NOM.PL REFL and
divešte se
 marvel.PTCP.IPFV.M.NOM.PL REFL
 ‘They stood crying and marvelling’ (*Life of Mary, Abraham’s Niece*, Bdinski Sbornik f. 16r)

Their temporal effects are also very different. INDEPENDENT RHEMES can (and normally do) introduce new temporal referents, and when perfective (which is most of the time in narrative discourse), they always induce narrative progression. ELABORATIONS depend on the matrix event for their temporal reference and always temporally coincide with the matrix event.

When occurring in chaining constructions, the leftmost conjunct participles can be ambiguous between INDEPENDENT RHEMES and, in Bary & Haug’s (2011) terminology, FRAMES, which set the stage for and provide temporal anchoring for the matrix event, as in (3). In specific configurations, they are instead most clearly FRAMES, most notably in sentence-initial position with the subject following the participle, as in the first conjunct participle in (1) above, or as part of two or more pre-matrix coordinated conjunct participles, as in (4), where the coordination suggests that they are found in the specifier of the matrix I’ (as with FRAMES) rather than adjoined to I’ (as with INDEPENDENT RHEME) (Haug 2012).

- (3) *i [rastrzavši]_{indrheme/frame} ryzi svoe biěše se*
 and tear.PART.PRF.M.NOM.SG robe.ACC.PL her beat.IMPV.3.SG REFL
po licju i chotěše se sama udaviti ot
 on face.DAT and want.IMPV.3.SG REFL self suffocate.INF from

pečaly

sorrow.GEN.SG

‘And after tearing her own robes/she tore her own and she started striking her own face and wanted to suffocate herself from sorrow’ (*Life of Mary, Abraham’s Niece*, Bdinski Sbornik f. 3v)

- (4) [*Sia* *že* *vəsi* *slyšavše* *i*
this.N.ACC.PL PTC this.N.ACC.PL hear.PTCP.PFV.M.NOM.SG and
viděvše] *množaišo* *věro* *i* *usrōdie* *kō*
see.PTCP.PFV.M.NOM.SG more.F.ACC faith.F.ACC and zeal.N.ACC to
stmu *pokazovaachō*
saint.M.DAT.SG show.IMPF.3.PL

‘After hearing and seeing all these things, they started showing greater faith and zeal towards the Saint’ *Life of Ivan of Rila*, Zografski Sbornik f. 102r)

(b.) **Dative absolutes are typical background material**

Early Slavic dative absolutes are instead typically FRAMES, i.e. background clauses. As such, they depend on the context for their own temporal reference, which is given by previously mentioned or easily inferable events. When occurring after the matrix clause, where they are much less frequent, they are most often interpreted as clause-final restrictive adjuncts (Fabricius-Hansen & Haug 2012b). These are very similar to FRAMES from the temporal and event-structuring perspective, since they restrict the domain of the matrix clause by introducing a *state* within which the main eventuality holds, as in (5). Like clause-final restrictive adjuncts, post-matrix absolutes may also receive a contrastive reading, as in (6).

- (5) *onō* *že* *šedō* *sēde* *na*
that.M.NOM.SG PTC go.PTCP.PFV.M.NOM.SG sit.AOR.3.SG on
stolē *černigově* *jaroslavu*
throne.SG.M.LOC Chernigov.SG.M.LOC Yaroslav.M.DAT.SG
suštju *nověgorodě* *togda*
be.PTCP.IPFV.M.DAT.SG Novgorod.SG.M.LOC then

‘He thus departed thence and established himself upon the throne of Chernigov, while Yaroslav was in Novgorod.’ (*Primary Chronicle*, Codex Laurentianus f. 50b)

- (6) *i* *povelē* *prinesti* *svěšte*
and command.AOR.3.SG bring.INF candle.PL.F.ACC
gorěšte. *i* *prižagati* *lice*
burn.PTCP.IPFV.F.ACC.PL and burn.INF face.SG.N.ACC
paule. *propovědniku* *vъpījōštu*
Paul’s.SG.N.ACC.STRONG preacher.M.DAT.SG cry out.PTCP.IPFV.M.DAT.SG
i *glagolōštu*
and say.PTCP.IPFV.M.DAT.SG

‘And he commanded to bring burning candles and burn Paul’s face, while the

preacher was crying out, saying' (*Vita of Paul and Juliana*, Codex Suprasliensis f. 7r)

A small number of post-matrix occurrences can function as ELABORATIONS, similarly to post-matrix conjunct participles, as in (7).

- (7) *i to rekō paki po pervoe*
 and that.N.ACC.SG say.PTCP.PFV.M.NOM.SG more on righteous.N.ACC.SG
drōžaše se dobroe ustroenie. bogu
 stay.IMPF.3.SG REFL good.N.ACC.SG behaviour.SG.N.ACC god.SG.M.DAT
pomagajuštu emu na blagoē
 help.PTCP.IPFV.M.DAT.SG 3.SG.M.DAT on good.N.ACC.SG
proizvolenie
 intention.N.ACC.SG

'After saying this, he stayed on the righteous path more than before, with God helping him in his good intentions' (*Life of Sergij of Radonezh* f. 46v)

However, several elaborating absolutes occur in rather predictable, perhaps formulaic, verb-subject combinations (as in (7) above), suggesting that the role of absolutes as ELABORATIONS is quite peripheral.

(c.) **Perfective *jegda*-events can include perfective matrix events**

Jegda-clauses can be analysed as background clauses by default because of the explicit subordinator *jegda* 'when' (following SDRT's account of *when*-clauses, see Asher et al. 2007), but their precise interpretation depends on several distinctions at the level of event structure (e.g. events versus states, durativity versus punctuality, telicity versus atelicity). Despite clear overlaps with participle FRAMES in their distribution and discourse functions, at least one clear semantic difference emerged between *jegda*-clauses and participle constructions. Namely, participle clauses, when perfective, always induce narrative progression, so that a perfective matrix clause is always interpreted as occurring (*just*) *after* the participle event. This led to the observation that perfective participle FRAMES always introduce a consequent state within which the matrix clause occurs. While this is also possible with perfective *jegda*-clauses, perfective *jegda*-clauses can introduce a durative but *bounded* referent *within* which the matrix event occurs, which is a temporal reading that does not seem to be available to perfective participle constructions. (8) is one such example.

- (8) *Bystō že egda kr̄stīšę se vsi*
 happen.AOR.3.SG PTC when baptize.AOR.3.PL REFL all.PL.M.NOM
ljud̄be. isou kr̄st̄b̄šju se i
 people.PL.M.NOM Jesus.DAT baptize.PTCP.PFV.M.SG.DAT REFL and
moleštju se. otvr̄ze se nebo
 pray.PTCP.IPFV.M.SG.DAT REFL open.AOR.3.SG REFL heaven.SG.N.NOM

'It came to pass that, when all the people were baptized, as Jesus was baptized and as he was praying, the heaven was opened' (Luke 3:21)

(d.) **INDEPENDENT RHEMES trigger Strong Narration, FRAMES trigger Background**

A formalization was proposed under an SDRT framework whereby conjunct participles functioning as INDEPENDENT RHEMES trigger a Strong Narration rhetorical relation (Bras et al. 2001) to the next event in a sequence. Strong Narration involves a strict discourse topic requirement and a temporal relation of *abutement* (strict sequentiality) to the following event in a series (i.e. either another INDEPENDENT RHEME or its matrix clause). Pre-matrix FRAMES are instead formalized as triggers of a Background_{forward} relation (Asher et al. 2007), which does not attach directly to the foreground (e.g. a series of INDEPENDENT RHEMES), but introduces a more complex structure consisting of a Framing Topic, linked to the background clause(s) (e.g. dative absolutes or framing conjunct participles) via Background and to the foreground clause(s) via Elaboration. The discourse topic requirement of INDEPENDENT RHEMES therefore nicely fits the possible interpretation of sentence-initial conjunct participles as FRAMES, as well as the predominance of sentence-initial dative absolutes over those in any other positions in the sentence. In other words, while INDEPENDENT RHEMES involve a strict discourse topic requirement, FRAMES introduce one explicitly.

(e.) **Sequentiality and event conflation as defining principles?**

The competition between participle adjuncts and *jegda*-clauses cannot be captured in terms of discrete, categorical variables but must be modelled as a continuum, allowing a large degree of overlap. Yet, there are very clear cross-linguistic patterns in the way this continuum is divided between overtly subordinated (finite) constructions and juxtaposed (non-finite or ‘unbalanced’, cf. Cristofaro 2013a,b, 2019) constructions. A semantic map of WHEN built from around 1400 linguistic varieties, including Old Church Slavonic, showed that ‘null’ constructions (e.g. participles, converbs, simple main clause juxtaposition, that is, non-overtly subordinated constructions) cluster in particular regions, indicating that they are not equally viable as alternatives to any use of WHEN, but carry particular meanings that make them less suitable for some of its functions. Comparisons with genealogically and areally unrelated but typologically similar languages showed that the majority of participle constructions in the semantic map correspond to cross-linguistic means used to stress the sequentiality of events, whereas *jegda*-clauses show virtually no competition when it comes to *universal* uses of WHEN in any tense, as well as *existential* uses *in non-past tenses*.

The stronger suitability of participle constructions to express sequentiality may be connected to the different temporal semantic properties mentioned above. These are the Strong Narration relation seemingly triggered by most pre-matrix conjunct participles (since these are most typically INDEPENDENT RHEMES) and the possibility for perfective *jegda*-clauses to be interpreted as *including* the following event in line

(matrix or adjunct), which, as we saw, is instead never a possible reading for perfective participle constructions (which always induce narrative progression). It is possible that participle FRAMES contribute to stressing the sequentiality of events by triggering a Background relation, which entails the explicit introduction of a Framing Topic. The Framing Topic, then, ensures discourse cohesion between the foreground events and satisfies the strict discourse topic requirement of INDEPENDENT RHEMES.

Similar criteria of discourse organization govern the use of chaining constructions in clause-chaining languages and participle constructions in Old Church Slavonic, both of which are regulated, behind the specific principles and constraints defined in each language, by the discourse distinction between foreground and background (Dooley 2010). Crucially, both clause chaining in clause-chaining languages and participle constructions (whether as FRAMES, INDEPENDENT RHEMES or ELABORATIONS) in Early Slavic may be seen as a broader strategy for event grouping (or ‘conflation’) and as a guarantee of tense-space-participant connectedness (Dooley 2010; Saebo 2011).

(e.) **‘Non-canonical’ participle constructions are canonical after all**

Virtually all ‘non-canonical’ usages of participle constructions mentioned or discussed in the previous literature on Early Slavic participles (i.e. co-referential dative absolutes, syntactically independent absolutes and conjunct participles, and participle constructions with no apparent matrix clause; cf. Collins 2011) are very well-attested in cross-linguistically phenomena that are functionally very similar to Early Slavic participles.

In particular, a shift in agentivity of a subject, a shift in time, location or grounding (from background to foreground) can all warrant the use of absolute constructions even when their subject is co-referential with the one of their matrix clause (i.e. the most common ‘non-canonical’ example of absolute constructions), similarly to how these same factors can explain the use of different-subject markers in several switch-reference systems across the world’s languages.

A shift in mood or actuality (Stirling 1993), for example when introducing direct speech, as well as parenthetical or metadiscursive functions, may instead warrant insubordination (i.e. the use as independent clauses) of conjunct participles and dative absolutes.

Given how both subject co-referentiality in dative absolutes and insubordination of participle constructions are found particularly frequently in original (non-translated) sources and outside of sacred and liturgical Old Church Slavonic texts, these clear cross-linguistic parallels incidentally also support the ‘nativeness’ of the constructions and their independent use from their Greek counterparts.

7.2 Where to from here?

The cross-linguistic perspective on Early Slavic participle constructions has shown, I believe, that token-based typological approaches can be very fruitful when adopted to study constructions in historical languages, for which elicitation of new data is not an option, for obvious reasons. The final chapter, in particular, has highlighted the great similarities between the functions of phenomena attested cross-linguistically in languages areally and genealogically unrelated to Old Church Slavonic, particularly switch-reference systems, and provided Early Slavic examples that clearly support such similarities. These examples were found upon close reading the occurrences of the constructions and checking their surrounding discourse. However, some of these phenomena lend themselves well to larger-scale, quantitative analyses.

In particular, clause bridging, presented in Section 6.3.2 of Chapter 6, has a regular, schematizable structure (cf. Guérin & Aiton 2019 and AnderBois & Altshuler 2022), which is worthwhile attempting to capture quantitatively in Early Slavic, by leveraging the recurrence of lemmas in adjacent portions of discourse. We have preliminarily shown (particularly in Section 2.3.2 of Chapter 2, besides Section 6.3.2 of Chapter 6) that clause bridging can be considered one of the functions of dative absolutes, but the close reading of examples from the corpus indicated that it likely deserves a separate, dedicated study investigating how pervasive it is, whereas in this thesis it only received relatively sporadic mentions.

Similar observations hold for the correspondences to switch-reference systems in the world's languages. As pointed out in Section 6.3.3 of Chapter 6, a large-scale computational comparison would require annotation at the morphosyntactic level for more languages than current computational tools can achieve, especially considering that several switch-reference languages are extremely low-resourced. However, we have seen that it is possible to obtain some clear patterns by automatically adding annotation for switch-reference markers in languages where these are relatively isolable. A promising next step, therefore, is to annotate several more languages for switch reference marking, following the plethora of existing descriptions of such systems. A larger parallel dataset with switch-reference annotation may then allow for semantic maps capturing finer-grained semantic and discourse distinctions in the usage of switch-reference markers and make better-informed comparisons with the distributions of participle constructions in Early Slavic.

An aspect that was not taken into account in the semantic map of WHEN, nor in this thesis as a whole, is the discourse functions of much more polyfunctional subordinators, most notably Old Church Slavonic *jako*, whose meanings span 'when', 'as', 'if', 'because', 'how', 'since', among others, and similarly Ancient Greek *hōs*, which is very similar in meaning to Old Church Slavonic *jako*. 'When' used as a question word (*kogda* in Old Church Slavonic) has also emerged from the semantic maps of Old Church Slavonic and is

likely to deserve separate investigation. The semantic map of WHEN pointed to potential cross-linguistic patterns of coexpression, whereby the question word *when*, when not separately lexicalized, is colexified with universal WHEN. However, this issue was not addressed directly, and our study hardly allows for proper generalizations on this matter.

From a theoretical perspective, I only touched upon the idea that clause chaining in Early Slavic and clause-chaining languages may convey a higher degree of event integration than what is conveyed by *jegda*-clauses and finite temporal subordinates in general. However, event conflation and the structure of macro-events are theoretically well-established, so a comprehensive account of finite and non-finite competitors in Early Slavic may benefit from fully modelling the competition within an event conflation theory as conceived and progressively refined mainly by Leonard Talmy since the 1970s (Talmy 1972, 1975, 1985, 1991, 2000).

Finally, the formulaic tendencies observed on dative absolutes across the corpus certainly deserve a dedicated study. We have seen (particularly throughout Chapter 2, but also as generally transpired from the low lexical variation) that dative absolutes tend to occur in very predictable participle lemma-subject lemma combinations and, not infrequently, in what seems like fixed or semi-fixed expressions. We have also noticed how some of these expressions are attested across different texts and genres, for example in generic temporal FRAMES such as *pozdě byvōšu* ‘when it got late’ and variations of the basic phrase *ne věduštju nikomuže* ‘without anyone knowing’, as well as seemingly semi-fixed expressions of the type ‘God <HELP> <PERSON>’, where HELP is generally one of a set of verbs semantically similar to *help* (e.g. *sōbljudati* ‘guard, watch over’, *pomagati* ‘help’, or *ukrēpljati* ‘give strength’). Other expressions, however, seem predictable only contextually to a specific text or genre, for instance in ‘reigning <RULER> in <PLACE>’ (e.g. *crstvujūštju olbkse vō csrigradě* ‘when Oleksa was reigning in Constantinople’), which are predominantly found in the chronicles. The ‘predictability’ of dative absolutes, then, seems to be well-suited for a study of the occurrences of the construction as potential instantiations of a limited set of *constructions* in the Construction Grammar sense of the term (i.e. ‘learned pairings of form with semantic or discourse functions’; Goldberg 2005: 5). The question may be fruitfully addressed under a distributional semantic framework and using language models trained on Early Slavic to capture the productivity of absolute constructions at different levels of abstraction (i.e. a lower level of abstraction may consider position and tense-aspect combinations as a construction, e.g. <absolute lemma-PFV + matrix lemma-AOR>; a higher level of abstraction may only consider aspect; the highest level may consider any configuration, and so on), in the spirit of what studies such as Perek (2016), Bozzone (2014), Rodda, Probert & McGillivray (2019) and Rodda (2021) have done on different phenomena in English and Ancient Greek. Despite the high diachronic and diatopic variation of historical languages, historical language modelling has been advancing at a rapid pace, which makes

distributional semantic studies of variation in participle constructions much more likely to be achievable than had been previously conceived.¹

¹Ancient Greek word embeddings I have developed with colleagues, for example, are described in Stopponi, Pedrazzini, Peels, McGillivray & Nissim (2023a) and McGillivray, Nissim, Pedrazzini, Peels-Matthey & Stopponi (Forthcoming) and available in Stopponi, Pedrazzini, Peels-Matthey, McGillivray & Nissim (2023b). Language modelling of Early Slavic, unsurprisingly, has so far received much less (and virtually no) attention, except for preliminary attempts by Pedrazzini (2023).

Appendices

Appendix A

Massively parallel dataset: corpus representativeness

A.1 Language families

Table A.1 lists all the world’s language families according to the Glottolog database (<https://glottolog.org>; Nordhoff & Hammarström 2011, Hammarström et al. 2023), how many languages belong to each of them, always following Glottolog, compared to how many languages from each family are represented in the massively parallel Bible corpus used in Chapters 5 and 6. In the table, *bible_* refers to the massively parallel corpus and *world_* to the languages family from Glottolog. *raw* are the raw number of languages belonging to the relevant family, *rel* is the relative frequency of these in relation to the total number of languages in the respective dataset (the parallel Bible dataset for *bible_*, the whole Glottolog language database for *world_*).

family	world_raw	world_rel	bible_raw	bible_rel
Atlantic-Congo	1380	18.1%	249.0	17.2%
Austronesian	1289	16.9%	246.0	17.0%
Indo-European	595	7.8%	110.0	7.6%
Sino-Tibetan	441	5.8%	90.0	6.2%
Afro-Asiatic	371	4.9%	47.0	3.3%
Nuclear Trans New Guinea	313	4.1%	94.0	6.5%
Pama-Nyungan	268	3.5%	13.0	0.9%
Otomanguean	180	2.4%	79.0	5.5%
Isolate	167	2.2%	16.0	1.1%
Austroasiatic	160	2.1%	10.0	0.7%
Sign Language	156	2.1%	Absent	Absent
Tai-Kadai	91	1.2%	3.0	0.2%
Dravidian	84	1.1%	6.0	0.4%
Mande	77	1.0%	24.0	1.7%
Tupian	73	1.0%	17.0	1.2%
Central Sudanic	65	0.9%	15.0	1.0%
Uto-Aztecan	64	0.8%	26.0	1.8%

Nilotic	57	0.7%	21.0	1.5%
Arawakan	56	0.7%	23.0	1.6%
Nuclear Torricelli	55	0.7%	9.0	0.6%
Algic	48	0.6%	8.0	0.6%
Athabaskan-Eyak-Tlingit	46	0.6%	7.0	0.5%
Quechuan	45	0.6%	27.0	1.9%
Turkic	45	0.6%	21.0	1.5%
Uralic	41	0.5%	8.0	0.6%
Kru	38	0.5%	9.0	0.6%
Hmong-Mien	37	0.5%	3.0	0.2%
Mayan	35	0.5%	25.0	1.7%
Sepik	34	0.4%	8.0	0.6%
Pano-Tacanan	33	0.4%	11.0	0.8%
Nakh-Daghestanian	32	0.4%	6.0	0.4%
Lower Sepik-Ramu	30	0.4%	4.0	0.3%
Cariban	29	0.4%	8.0	0.6%
Salishan	27	0.4%	Absent	Absent
Artificial	26	0.3%	2.0	0.1%
Nuclear-Macro-Je	25	0.3%	7.0	0.5%
Tucanoan	23	0.3%	16.0	1.1%
Timor-Alor-Pantar	23	0.3%	Absent	Absent
Chibchan	21	0.3%	9.0	0.6%
Lakes Plain	20	0.3%	Absent	Absent
Ta-Ne-Omotc	19	0.2%	7.0	0.5%
Dogon	19	0.2%	1.0	0.1%
Siouan	18	0.2%	1.0	0.1%
Mixe-Zoque	17	0.2%	8.0	0.6%
Yam	17	0.2%	Absent	Absent
Anim	16	0.2%	Absent	Absent
Mongolic-Khitani	16	0.2%	3.0	0.2%
Pidgin	16	0.2%	1.0	0.1%
Border	15	0.2%	2.0	0.1%
North Halmahera	15	0.2%	3.0	0.2%
Nubian	14	0.2%	Absent	Absent
Angan	13	0.2%	4.0	0.3%
Ndu	13	0.2%	3.0	0.2%
Japonic	13	0.2%	1.0	0.1%
Tor-Orya	13	0.2%	Absent	Absent
Khoe-Kwadi	13	0.2%	2.0	0.1%
Iroquoian	13	0.2%	2.0	0.1%
Tungusic	12	0.2%	1.0	0.1%
Totonacan	12	0.2%	8.0	0.6%
Eskimo-Aleut	12	0.2%	4.0	0.3%
Worroran	12	0.2%	Absent	Absent
Western Daly	11	0.1%	Absent	Absent
Great Andamanese	11	0.1%	Absent	Absent

Geelvink Bay	10	0.1%	1.0	0.1%
Miwok-Costanoan	10	0.1%	Absent	Absent
Sko	10	0.1%	Absent	Absent
Songhay	10	0.1%	2.0	0.1%
Heibanic	10	0.1%	1.0	0.1%
Nyulnyulan	10	0.1%	Absent	Absent
Ijoid	10	0.1%	Absent	Absent
Gunwinyguan	10	0.1%	2.0	0.1%
Saharan	10	0.1%	Absent	Absent
Surmic	10	0.1%	1.0	0.1%
Dagan	9	0.1%	4.0	0.3%
Chocoan	9	0.1%	3.0	0.2%
South Bougainville	9	0.1%	1.0	0.1%
Maban	9	0.1%	Absent	Absent
Cochimi-Yuman	9	0.1%	Absent	Absent
Koiarian	8	0.1%	5.0	0.3%
Greater Kwerba	8	0.1%	Absent	Absent
Narrow Talodi	8	0.1%	Absent	Absent
Dajuic	7	0.1%	Absent	Absent
Muskogean	7	0.1%	Absent	Absent
Matacoan	7	0.1%	6.0	0.4%
Pomoan	7	0.1%	Absent	Absent
Bosavi	7	0.1%	3.0	0.2%
Tuu	7	0.1%	Absent	Absent
Kiwaian	6	0.1%	Absent	Absent
East Strickland	6	0.1%	2.0	0.1%
Baining	6	0.1%	1.0	0.1%
Kartvelian	6	0.1%	1.0	0.1%
Nambiquaran	6	0.1%	1.0	0.1%
Left May	6	0.1%	1.0	0.1%
Lengua-Mascoy	6	0.1%	2.0	0.1%
Chumashan	6	0.1%	Absent	Absent
Mailuan	6	0.1%	Absent	Absent
Kiowa-Tanoan	6	0.1%	1.0	0.1%
Kadugli-Krongo	6	0.1%	Absent	Absent
Zaparoan	6	0.1%	1.0	0.1%
Wakashan	6	0.1%	Absent	Absent
South Bird's Head Family	6	0.1%	Absent	Absent
Arawan	6	0.1%	2.0	0.1%
Yareban	5	0.1%	2.0	0.1%
Caddoan	5	0.1%	Absent	Absent
Abkhaz-Adyge	5	0.1%	1.0	0.1%
South Omotic	5	0.1%	1.0	0.1%
Yanomamic	5	0.1%	2.0	0.1%
Pauwasi	5	0.1%	1.0	0.1%
Kxa	5	0.1%	Absent	Absent

Guaicuruan	5	0.1%	4.0	0.3%
Yeniseian	5	0.1%	1.0	0.1%
Chukotko-Kamchatkan	5	0.1%	1.0	0.1%
Keram	5	0.1%	Absent	Absent
Barbacoan	5	0.1%	4.0	0.3%
Eleman	5	0.1%	Absent	Absent
Mirndi	5	0.1%	Absent	Absent
Guahiboan	5	0.1%	3.0	0.2%
Misumalpan	5	0.1%	2.0	0.1%
Sahaptian	5	0.1%	Absent	Absent
Yuat	5	0.1%	1.0	0.1%
Nimboranic	5	0.1%	Absent	Absent
Huitotoan	5	0.1%	2.0	0.1%
West Bird's Head	5	0.1%	Absent	Absent
Naduhup	4	0.1%	1.0	0.1%
Huavean	4	0.1%	1.0	0.1%
Koman	4	0.1%	1.0	0.1%
Tangkic	4	0.1%	Absent	Absent
North Bougainville	4	0.1%	1.0	0.1%
Suki-Gogodala	4	0.1%	Absent	Absent
Sentanic	4	0.1%	Absent	Absent
Chicham	4	0.1%	4.0	0.3%
Iwaidjan Proper	4	0.1%	1.0	0.1%
Walioic	4	0.1%	Absent	Absent
Chapacuran	4	0.1%	Absent	Absent
Yukaghir	4	0.1%	1.0	0.1%
Eastern Jebel	4	0.1%	Absent	Absent
Blue Nile Mao	4	0.1%	Absent	Absent
Mangarrayi-Maran	4	0.1%	Absent	Absent
Maiduan	4	0.1%	Absent	Absent
Wintuan	4	0.1%	Absent	Absent
Eastern Trans-Fly	4	0.1%	2.0	0.1%
Turama-Kikori	4	0.1%	Absent	Absent
Maningrida	4	0.1%	1.0	0.1%
Chinookan	3	0.0%	Absent	Absent
Lepki-Murkim-Kembra	3	0.0%	Absent	Absent
Kwalean	3	0.0%	1.0	0.1%
Giimbiyu	3	0.0%	Absent	Absent
Kayagaric	3	0.0%	Absent	Absent
Yangmanic	3	0.0%	Absent	Absent
Tamaic	3	0.0%	Absent	Absent
Dizoid	3	0.0%	Absent	Absent
Kunimaipan	3	0.0%	3.0	0.2%
ArafundiArafundi	3	0.0%	Absent	Absent
East Bird's Head	3	0.0%	3.0	0.2%
West Bomberai	3	0.0%	Absent	Absent

Kalapuyan	3	0.0%	Absent	Absent
Kolopom	3	0.0%	Absent	Absent
Saliban	3	0.0%	Absent	Absent
Tsimshian	3	0.0%	Absent	Absent
Jarrakan	3	0.0%	Absent	Absent
Aymaran	3	0.0%	1.0	0.1%
Kamula-Elevala	3	0.0%	2.0	0.1%
Mixed Language	3	0.0%	Absent	Absent
Mairasic	3	0.0%	Absent	Absent
Kuliak	3	0.0%	Absent	Absent
Bororoan	3	0.0%	Absent	Absent
Amtu-Musan	2	0.0%	Absent	Absent
Kresh-Aja	2	0.0%	Absent	Absent
Gumuz	2	0.0%	1.0	0.1%
Teberan	2	0.0%	2.0	0.1%
Senagi	2	0.0%	1.0	0.1%
Harakmbut	2	0.0%	1.0	0.1%
Manubaran	2	0.0%	1.0	0.1%
Doso-Turumsa	2	0.0%	Absent	Absent
Inanwatan	2	0.0%	Absent	Absent
Furan	2	0.0%	Absent	Absent
Nyimang	2	0.0%	Absent	Absent
GarrwanGarawa	2	0.0%	1.0	0.1%
Limilngan-Wulna	2	0.0%	Absent	Absent
Peba-Yagua	2	0.0%	1.0	0.1%
Cahuapanan	2	0.0%	1.0	0.1%
Eastern Daly	2	0.0%	Absent	Absent
Kakua-Nukak	2	0.0%	1.0	0.1%
Bulaka River	2	0.0%	Absent	Absent
Keresan	2	0.0%	Absent	Absent
Tarascan	2	0.0%	2.0	0.1%
Haida	2	0.0%	Absent	Absent
Zamucoan	2	0.0%	2.0	0.1%
East Kutubu	2	0.0%	Absent	Absent
Chonan	2	0.0%	Absent	Absent
Jarawa-Onge	2	0.0%	Absent	Absent
Palaihnihan	2	0.0%	Absent	Absent
Konda-Yahadian	2	0.0%	Absent	Absent
Yuki-Wappo	2	0.0%	Absent	Absent
Mombum-Koneraw	2	0.0%	Absent	Absent
Yawa-Saweru	2	0.0%	1.0	0.1%
Baibai-Fas	2	0.0%	Absent	Absent
Tequistlatecan	2	0.0%	1.0	0.1%
Rashad	2	0.0%	Absent	Absent
Temeinic	2	0.0%	Absent	Absent
Hibito-Cholon	2	0.0%	Absent	Absent

Piawi	2	0.0%	1.0	0.1%
Koreanic	2	0.0%	1.0	0.1%
Bunaban	2	0.0%	Absent	Absent
Katukinan	2	0.0%	Absent	Absent
Northern Daly	2	0.0%	Absent	Absent
Kwomtari-Nai	2	0.0%	Absent	Absent
Southern Daly	2	0.0%	1.0	0.1%
Bogia	2	0.0%	Absent	Absent
ChimakuanChimakum	2	0.0%	Absent	Absent
Somahai	2	0.0%	Absent	Absent
Boran	2	0.0%	2.0	0.1%
Coosan	2	0.0%	Absent	Absent
Marrku-Wurrugu	2	0.0%	Absent	Absent
Pahoturi	2	0.0%	Absent	Absent
Katla-Tima	2	0.0%	Absent	Absent
Bayono-Awbono	2	0.0%	Absent	Absent
Namla-Tofanma	2	0.0%	Absent	Absent
Kaure-Kosare	2	0.0%	Absent	Absent
Hurro-Urartian	2	0.0%	Absent	Absent
Uru-Chipaya	2	0.0%	1.0	0.1%
Araucanian	2	0.0%	1.0	0.1%
Puri-Coroado	1	0.0%	Absent	Absent
Kawesqar	1	0.0%	Absent	Absent
Ainu	1	0.0%	Absent	Absent
Kamakanan	1	0.0%	Absent	Absent
Hatam-Mansim	1	0.0%	Absent	Absent
Shastan	1	0.0%	Absent	Absent
Yokutsan	1	0.0%	Absent	Absent
Nivkh	1	0.0%	1.0	0.1%
Chiquitano	1	0.0%	1.0	0.1%
Jicaquean	1	0.0%	1.0	0.1%
Ticuna-Yuri	1	0.0%	1.0	0.1%
Taulil-Butam	1	0.0%	Absent	Absent

Table A.1: Language families represented in the massively parallel Bible dataset compared with the world’s language families according to Glottolog.

Appendix B

Language classification by pattern of co-expression

Table B.3 contains all the languages in the massively parallel dataset used in Chapters 5 and 6 to generate the semantic map of WHEN. In the table, languages are ordered by their ISO 639-3 code and classified by pattern and subpattern, which was assigned automatically as outlined below. Each language has an indication of its name, language family, and main macroarea in which it is attested. This metadata, as explained in Section 5.2 of Chapter 5, is obtained from the Glottolog database (<https://glottolog.org>; Nordhoff & Hammarström 2011, Hammarström et al. 2023).

B.1 Pattern and subpattern assignment

The three basic patterns are defined on the basis of the top-left (TL), mid-left (ML) and bottom-left (BL) areas of the semantic map of when and assigned as shown in Table B.1.

pattern	areas coexpressed
A	TL = ML = BL
B	(TL = ML) ≠ BL
C	TL ≠ (ML = BL)
D	TL ≠ ML ≠ BL
E	(TL = BL) ≠ ML

Table B.1: Basic coexpression patterns

Subpatterns are assigned to languages that do not fit one of the five basic patterns above, due to one or more of the clusters being under the scope of more than one linguistic item. Depending on how many items there are in total for the three areas, there is a different number of total logical combinations, all of which are taken into consideration when assigning a subpattern. To take the possible combinations into account, we use a construction with the following template:

$$[]_{TL} []_{ML} []_{BL}$$

where each slot is filled by arbitrary letters standing for each different linguistic item having scope over the relevant cluster. To capture that two or more clusters are colexified, the same letter is repeated in all those clusters and, conversely, different letters indicate different linguistic items. In this way, we can schematically represent more complex patterns than those in Table B.1. For example, below, (a.) corresponds to pattern D, since the three clusters are dominated by three different linguistic items. Instead, (b.) represents a language in which the TL cluster is dominated by both a lexical item unique to that cluster (X in the template), but the same item that has scope on the ML cluster has also scope over the TL cluster (Y in the template). (c.) instead corresponds to a language in which the ML and BL clusters are colexified, while the TL cluster is under the scope of not one, but two lexical items.

- (a.) [X]_{TL} [Y]_{ML} [Z]_{BL}
- (b.) [XY]_{TL} [Y]_{ML} [Z]_{BL}
- (c.) [XY]_{TL} [Z]_{ML} [Z]_{BL}

In the overview in Table B.2, for ease of presentation, the three slots in the template are only separated by a space, but the order remains the same as in the template shown above. For example, (c.) above is represented as **XX Z Z**. ? indicates ‘no lexical item’, representing a language in which the relevant cluster has no Kriging area.

configuration	subpattern label
XY Y X	BxE
X XY Y	BxC
X Y XY	CxE
XW Y Z	D2a
X YW Z	D2b
X Y ZW	D2c
XZ Y Y	C3
X YZ X	E3
X X YZ	B3
XY X X	CxA
X XY X	ExA
X X XY	BxA
X YX Z	DxX
X Y ZX	DxX
XY Y Z	DxX
X Y ZY	DxX
XZ Y Z	DxX
X XY XY	AxC
XY X XY	AxE
XY XY X	AxB
XYZ X X	AxC3
X XYZ X	AxE3
X X XYZ	AxB3
XYZ X Y	D-Other
X XYZ Y	D-Other
X Y XYZ	D-Other
XY XY XY	A2
? X X	A?C
? X Y	B?D?E
X ? X	A?E
X ? Y	B?C?D
X X ?	A?B
X Y ?	C?D?E

Table B.2: Subpatterns of coexpression

iso	name	family	macroarea	pattern	subpattern
aai	Arifama-Miniafia	Austronesian	Papunesia	E	E
aak	Ankave	Angan	Papunesia	A	A
aau	Abau	Sepik	Papunesia	D	DxX
aaz	Amarasi	Austronesian	Papunesia	C	C
abt	Ambulas	Ndu	Papunesia	A	A
abx	Inabaknon	Austronesian	Papunesia	D	D
aby	Aneme Wake	Yareban	Papunesia	A	A
aca	Achagua	Arawakan	South America	C	C
acd	Gikyode	Atlantic-Congo	Africa	D	D

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
ace	Acehnese	Austronesian	Papunesia	D	DxX
acf	Saint Lucian Creole French	Indo-European	Eurasia	A	A
ach	Acoli	Nilotic	Africa	E	E
acn	Longchuan Achang	Sino-Tibetan	Eurasia	E	E
acr	Achi	Mayan	North America	C	CxA
acu	Achuar-Shiwiar	Chicham	South America	A	A
ade	Adele	Atlantic-Congo	Africa	D	D
adh	Adhola	Nilotic	Africa	C	C
adi	Bori-Karko	Sino-Tibetan	Eurasia	C	C
adj	Adioukrou	Atlantic-Congo	Africa	A	A
adl	Galo	Sino-Tibetan	Eurasia	A	A
aeb	Tunisian Arabic	Afro-Asiatic	Eurasia	C	AxC
aer	Eastern Arrernte	Pama-Nyungan	Australia	A?E	A?E
aeu	Akeu	Sino-Tibetan	Eurasia	A	A
aez	Amele	Nuclear Trans New Guinea	Papunesia	B	B
afr	Afrikaans	Indo-European	Eurasia	C	C
agd	Agarabi	Nuclear Trans New Guinea	Papunesia	B	B
agg	Angor	Senagi	Papunesia	A	A
agm	Angaataha	Angan	Papunesia	A	A
agn	Agutaynen	Austronesian	Papunesia	C	C
agr	Aguaruna	Chicham	South America	A	A
agt	Central Cagayan Agta	Austronesian	Papunesia	B?C?D	B?C?D
agu	Aguacateco	Mayan	North America	C	C
agw	Kahua	Austronesian	Papunesia	A	A
agx	Aghul	Nakh-Daghestanian	Eurasia	A	A
ahk	Akha	Sino-Tibetan	Eurasia	A	A
aia	Arosi	Austronesian	Papunesia	A	A
aii	Assyrian Neo-Aramaic	Afro-Asiatic	Eurasia	A	A
aim	Aimol	Sino-Tibetan	Eurasia	D	D
aji	Ajië	Austronesian	Papunesia	D	D2c
ajz	Amri Karbi	Sino-Tibetan	Eurasia	A	A
akb	Batak Angkola	Austronesian	Papunesia	B?C?D	B?C?D
ake	Akawaio-Ingariko	Cariban	South America	A	A
akh	Angal Heneng	Nuclear Trans New Guinea	Papunesia	A	A
akp	Siwu	Atlantic-Congo	Africa	C	C
ald	Alladian	Atlantic-Congo	Africa	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
alj	Alangan	Austronesian	Papunesia	D	D
aln	Gheg Albanian	Indo-European	Eurasia	A	A
alp	Alune	Austronesian	Papunesia	C	CxA
alq	Algonquin	Algic	North America	A	A
alt	Southern Altai	Turkic	Eurasia	A	A
aly	Alyawarr	Pama-Nyungan	Australia	A	A
alz	Alur	Nilotic	Africa	C	CxA
ame	Yanesha'	Arawakan	South America	D	D2a
amf	Hamer-Banna	South Omotic	Africa	B	B
amh	Amharic	Afro-Asiatic	Eurasia	A	A
amk	Ambai	Austronesian	Papunesia	E	E
amm	Ama	Left May	Papunesia	B	B
amn	Amanab	Border	Papunesia	A	A
amp	Alamblak	Sepik	Papunesia	A	A
amr	Amarakaeri	Harakmbut	South America	A	A
amu	Guerrero Amuzgo	Otomanguean	North America	E	AxE
amx	Anmatyerre	Pama-Nyungan	Australia	C	C
anm	Anal	Sino-Tibetan	Eurasia	A	A
ann	Obolo	Atlantic-Congo	Africa	C	CxA
anv	Denya	Atlantic-Congo	Africa	B	B
any	Anyin	Atlantic-Congo	Africa	C	C
aoi	Anindilyakwa	Gunwinyguan	Australia	A	A
aoj	Mufian	Nuclear Torricelli	Papunesia	C	C
aom	Ömie	Koiarian	Papunesia	A?E	A?E
aon	Bumbita Arapesh	Nuclear Torricelli	Papunesia	B	B
aoz	Uab Meto	Austronesian	Papunesia	A	A
apb	Sa'a	Austronesian	Papunesia	A	A
ape	Bukiyip	Nuclear Torricelli	Papunesia	B	B
apn	Apinayé	Nuclear-Macro-Je	South America	B	B
apr	Arop-Lokep	Austronesian	Papunesia	A	A
apt	Apatani	Sino-Tibetan	Eurasia	D	D
apu	Apurinã	Arawakan	South America	A	A
apw	Western Apache	Athabaskan-Eyak-Tlingit	North America	A	A
apy	Apalaí	Cariban	South America	C	C
apz	Safeyoka	Angan	Papunesia	E	E
arb	Standard Arabic	Afro-Asiatic	Eurasia	C	CxA
are	Western Arrarnta	Pama-Nyungan	Australia	E	E
arl	Arabela	Zaparoan	South America	A	A
arn	Mapudungun	Araucanian	South America	B	B
arp	Arapaho	Algic	North America	C	C
ary	Moroccan Arabic	Afro-Asiatic	Eurasia	A	A
arz	Egyptian Arabic	Afro-Asiatic	Eurasia	C	CxA

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
asg	Western-Kambari-Cishingini	Atlantic-Congo	Africa	D	DxX
aso	Dano	Nuclear Trans New Guinea	Papunesia	A	A
ata	Pele-Ata	Isolate	Eurasia	A	A
atb	Zaiwa	Sino-Tibetan	Eurasia	B	B
atd	Ata Manobo	Austronesian	Papunesia	B	B
atg	Ivbie North-Okpela-Arhe	Atlantic-Congo	Africa	C	C
att	Pamplona Atta	Austronesian	Papunesia	C	C
auc	Waorani	Isolate	Eurasia	A	A
aui	Anuki	Austronesian	Papunesia	A	A
auy	Awiyaana	Nuclear Trans New Guinea	Papunesia	A	A
ava	Avar	Nakh-Daghestanian	Eurasia	B	B
avt	Au	Nuclear Torricelli	Papunesia	D	D
avu	Avokaya	Central Sudanic	Africa	B?C?D	B?C?D
awa	Awadhi	Indo-European	Eurasia	A	A
awb	Awa	Nuclear Trans New Guinea	Papunesia	A	A
awi	Aekyom	Kamula-Elevala	Papunesia	A	A
ayo	Ayoreo	Zamucoan	South America	C	C
ayr	Central Aymara	Aymaran	South America	B	B
aze	South Azerbaijani	Turkic	Eurasia	A	A
azg	San Pedro Amuzgos Amuzgo	Otomanguean	North America	B	B
azz	Highland Puebla Nahuatl	Uto-Aztecan	North America	A	A
bak	Bashkir	Turkic	Eurasia	A	A
bam	Bambara	Mande	Africa	A	A
ban	Balinese	Austronesian	Papunesia	C	C3
bao	Waimaha	Tucanoan	South America	A	A
bar	Bavarian	Indo-European	Eurasia	C	C
bav	Vengo	Atlantic-Congo	Africa	BxC	BxC
bba	Baatonum	Atlantic-Congo	Africa	D	DxX
bbb	Barai	Koiarian	Papunesia	A	A
bbc	Batak Toba	Austronesian	Papunesia	A?E	A?E
bbj	Ghomálá'	Atlantic-Congo	Africa	B	B
bbr	Girawa	Nuclear Trans New Guinea	Papunesia	A	A
bcc	Southern Balochi	Indo-European	Eurasia	E	ExA
bch	Bariai	Austronesian	Papunesia	D	D

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
bci	Baoulé	Atlantic-Congo	Africa	C	CxA
bcl	Coastal-Naga Bikol	Austronesian	Papunesia	C	C
bco	Kaluli	Bosavi	Papunesia	A	A
bcw	Bana	Afro-Asiatic	Eurasia	A	A
bdd	Bunama	Austronesian	Papunesia	D	D
bdh	Baka	Central Sudanic	Africa	C	CxA
bea	Beaver	Athabaskan- Eyak-Tlingit	North America	C	C
bef	Benabena	Nuclear Trans New Guinea	Papunesia	A	A
bel	Belarusian	Indo-European	Eurasia	A	A
bem	Bemba	Atlantic-Congo	Africa	A	A
ben	Bengali	Indo-European	Eurasia	B	B
beq	Beembe	Atlantic-Congo	Africa	A	A
bfd	Bafut	Atlantic-Congo	Africa	E	ExA
bfo	Malba Birifor	Atlantic-Congo	Africa	D	D2c
bgr	Bawm Chin	Sino-Tibetan	Eurasia	B?C?D	B?C?D
bgs	Tagabawa	Austronesian	Papunesia	C	C
bgz	Banggai	Austronesian	Papunesia	D	DxX
bhg	Binandere	Nuclear Trans New Guinea	Papunesia	A	A
bhl	Bimin	Nuclear Trans New Guinea	Papunesia	B	B
bhp	Bima	Austronesian	Papunesia	D	DxX
bib	Bissa	Mande	Africa	A	A2
big	Biangai	Kunimaipan	Papunesia	C	C
bim	Bimoba	Atlantic-Congo	Africa	B	B
bis	Bislama	Indo-European	Eurasia	C	CxA
biu	Biete	Sino-Tibetan	Eurasia	C	C
biv	Southern Birifor	Atlantic-Congo	Africa	E	E
bjp	Fanamaket	Austronesian	Papunesia	A	A
bjr	Binumarien	Nuclear Trans New Guinea	Papunesia	A	A
bjv	Nangnda	Central Sudanic	Africa	E	ExA
bkd	Talaandig- Binukid	Austronesian	Papunesia	C	C
bkq	Bakairí	Cariban	South America	A	A
bku	Buhid	Austronesian	Papunesia	B	B
bkv	Bekwarra	Atlantic-Congo	Africa	A	A
blh	Kuwaa	Kru	Africa	C	C
blw	Balangao	Austronesian	Papunesia	B	B
blz	Balantak	Austronesian	Papunesia	D	D
bmb	Bembe	Atlantic-Congo	Africa	D	D

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
bmh	Kein	Nuclear Trans New Guinea	Papunesia	A	A
bmk	Ghayavi	Austronesian	Papunesia	A	A
bmq	Bomu	Atlantic-Congo	Africa	C	CxA
bmr	Muinane	Boran	South America	A	A
bmu	Burum-Mindik	Nuclear Trans New Guinea	Papunesia	B	B
bnj	Eastern Tawbuid	Austronesian	Papunesia	E	E
bnp	Bola	Austronesian	Papunesia	A	A
boa	Bora	Boran	South America	A	A
boj	Anjam	Nuclear Trans New Guinea	Papunesia	A	A
bom	Berom	Atlantic-Congo	Africa	A	A
bon	Bine	Eastern Trans- Fly	Papunesia	A	A
bov	Tuwuli	Atlantic-Congo	Africa	D	D
box	Buamu	Atlantic-Congo	Africa	C	C
bpr	Koronadal Blaan	Austronesian	Papunesia	D	D
bps	Sarangani Blaan	Austronesian	Papunesia	D	D
bqc	Boko	Mande	Africa	D	DxX
bqj	Bandial	Atlantic-Congo	Africa	A	A
bqp	Busa	Mande	Africa	C	C
bre	Breton	Indo-European	Eurasia	A	A
bru	Eastern Bru	Austroasiatic	Eurasia	D	DxX
bsc	Bassari-Tanda	Atlantic-Congo	Africa	B	BxA
bsn	Barasana-Eduria	Tucanoan	South America	A	A
bsp	Baga Sitemu	Atlantic-Congo	Africa	C	C
bss	Akoose	Atlantic-Congo	Africa	C	C
btd	Batak Dairi	Austronesian	Papunesia	D	D
bth	Biatah Bidayuh	Austronesian	Papunesia	E	ExA
bto	Iriga Bicolano	Austronesian	Papunesia	A?E	A?E
bts	Batak Simalun- gun	Austronesian	Papunesia	D	D
btt	Bete-Bendi	Atlantic-Congo	Africa	Other	Other
btx	Batak Karo	Austronesian	Papunesia	D	D
bud	Ntcham	Atlantic-Congo	Africa	D	D
bug	Buginese	Austronesian	Papunesia	C	C
buk	Bugawac	Austronesian	Papunesia	B	B
bul	Bulgarian	Indo-European	Eurasia	B	BxA
bum	Bulu	Atlantic-Congo	Africa	D	D
bus	Bokobaru	Mande	Africa	C	C
bvr	Burarra	Maningrida	Australia	A	A
bvz	Bauzi	Geelvink Bay	Papunesia	B?C?D	B?C?D

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
bwq	Southern Bobo Madaré	Mande	Africa	B	BxA
bwu	Buli	Atlantic-Congo	Africa	C	C
bxh	Buhutu	Austronesian	Papunesia	E	E
bxr	Russia Buriat	Mongolic-Khitani	Eurasia	A	A
byr	Baruya	Angan	Papunesia	A	A
byx	Qaqet	Baining	Papunesia	B?C?D	B?C?D
bzd	Bribri	Chibchan	North America	A	A
bzh	Mapos Buang	Austronesian	Papunesia	C	C
bzi	Bisu	Sino-Tibetan	Eurasia	B	B
bzj	Belize Kriol English	Indo-European	Eurasia	A	A
caa	Chortí	Mayan	North America	A	A
cab	Garifuna	Arawakan	South America	C	C
cac	Chuj	Mayan	North America	B	BxA
caf	Southern Carrier	Athabaskan-Eyak-Tlingit	North America	B	B
cag	Nivaclé	Matacoan	South America	D	D
cak	Kaqchikel	Mayan	North America	A	A
cao	Chácobo	Pano-Tacanan	South America	A	A
cap	Chipaya	Uru-Chipaya	South America	B	B
caq	Car Nicobarese	Austroasiatic	Eurasia	C	C
car	Galibi Carib	Cariban	South America	BxE	BxE
cas	Mosetén-Chimané	Isolate	Eurasia	A	A
cat	Catalan	Indo-European	Eurasia	B	BxA
cav	Cavineña	Pano-Tacanan	South America	A	A
cax	Lomeriano-Ignaciano Chiquitano	Chiquitano	South America	C	C
cbc	Carapana	Tucanoan	South America	A	A
cbi	Cha'palaa	Barbacoan	South America	A	A
cbk	Chavacano	Indo-European	Eurasia	C	C3
cbr	Cashibo-Cacataibo	Pano-Tacanan	South America	A	A
cbs	Cashinahua	Pano-Tacanan	South America	A	A
cbt	Shawi	Cahuapanan	South America	A	A
cbu	Candoshi-Shapra	Isolate	Eurasia	A	A
cbv	Kakua	Kakua-Nukak	South America	A	A
cce	Chopi	Atlantic-Congo	Africa	BxE	BxE
cco	Comaltepec Chinantec	Otomanguean	North America	A	A
ceb	Cebuano	Austronesian	Papunesia	A	A
ceg	Chamacoco	Zamucoan	South America	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
cek	Eastern Khumi Chin	Sino-Tibetan	Eurasia	B	BxA
ces	Czech	Indo-European	Eurasia	C	CxA
cfm	Falam Chin	Sino-Tibetan	Eurasia	A	A
cgc	Kagayanen	Austronesian	Papunesia	D	D
cha	Chamorro	Austronesian	Papunesia	C	C
chd	Highland Oaxaca Chontal	Tequistlatecan	North America	E	E
che	Chechen	Nakh- Daghestanian	Eurasia	E	E
chf	Tabasco Chontal	Mayan	North America	A	A
chk	Chuukese	Austronesian	Papunesia	A	A
chq	Quiotepec Chinantec	Otomanguean	North America	C	C
chr	Cherokee	Iroquoian	North America	A	A
chu	Church Slavic	Indo-European	Eurasia	B	B
chv	Chuvash	Turkic	Eurasia	E	E
chz	Ozumacín Chinantec	Otomanguean	North America	C	C
cjo	Ashéninka Pajonal	Arawakan	South America	C	C
cjp	Cabécar	Chibchan	North America	C	C
cjs	Shor	Turkic	Eurasia	A	A
cjv	Chuave	Nuclear Trans New Guinea	Papunesia	A	A
ckb	Central Kurdish	Indo-European	Eurasia	A	A
cko	Anufo	Atlantic-Congo	Africa	C	CxA
ckt	Chukchi	Chukotko- Kamchatkan	Eurasia	A	A
cle	Lealao Chinantec	Otomanguean	North America	A	A
clu	Caluyanun	Austronesian	Papunesia	D	D
cly	Eastern Highland Chatino	Otomanguean	North America	E	AxE
cme	Cerma	Atlantic-Congo	Africa	C	C
cmo	Central Mnong	Austroasiatic	Eurasia	A	A
cnh	Haka Chin	Sino-Tibetan	Eurasia	A	A
cni	Asháninka	Arawakan	South America	D	D
cnl	Lalana Chinantec	Otomanguean	North America	C	C
cnt	Tepetotutla Chi- nantec	Otomanguean	North America	D	D
cnw	Ngawn Chin	Sino-Tibetan	Eurasia	A	A
coe	Koreguaje	Tucanoan	South America	A	A
cof	Tsafiki	Barbacoan	South America	C	C

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
cok	Santa Teresa Cora	Uto-Aztecan	North America	B	BxA
con	Cofán	Isolate	Eurasia	A	A
cop	Coptic	Afro-Asiatic	Eurasia	C	C
cor	Cornish	Indo-European	Eurasia	A	A
cot	Caquinte	Arawakan	South America	A	A
cpa	Palantla Chinantec	Otomanguean	North America	D	D
cpb	Ucayali-Yurúa Ashéninka	Arawakan	South America	D	D
cpc	Ajyíninka Apurucayali	Arawakan	South America	D	D
cpu	Pichis Ashéninka	Arawakan	South America	D	D
cpy	South Ucayali Ashéninka	Arawakan	South America	D	D
crh	Crimean Tatar	Turkic	Eurasia	C	C
crm	Moose Cree	Algic	North America	A	A
crn	El Nayar Cora	Uto-Aztecan	North America	C	CxA
crq	Iyo'wujwa Chorote	Matacoan	South America	B	B
crs	Seselwa Creole French	Indo-European	Eurasia	A	A
crt	Iyojwa'ja Chorote	Matacoan	South America	B	B
crx	Central Carrier	Athabaskan-Eyak-Tlingit	North America	A	A
csk	Jola-Esulalu	Atlantic-Congo	Africa	A	A
cso	Sochiapam Chinantec	Otomanguean	North America	C	CxA
csy	Sizang Chin	Sino-Tibetan	Eurasia	E	E
cta	Tataltepec Chatino	Otomanguean	North America	A	A
ctd	Tedim Chin	Sino-Tibetan	Eurasia	A	A
ctp	Western Highland Chatino	Otomanguean	North America	A	A
ctu	Chol	Mayan	North America	C	C
cub	Cubeo	Tucanoan	South America	A	A
cuc	Usila Chinantec	Otomanguean	North America	D	D2a
cui	Cuiba	Guahiboan	South America	A	A
cuk	San Blas Kuna	Chibchan	North America	A	A
cul	Culina	Arawakan	South America	A	A
cut	Teutila Cuicatec	Otomanguean	North America	A	A
cux	Tepeuxila Cuicatec	Otomanguean	North America	B	AxB
cwe	Kwere	Atlantic-Congo	Africa	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
cwt	Kuwaataay	Atlantic-Congo	Africa	D	DxX
cya	Nopala Chatino	Otomanguean	North America	A	A
cym	Welsh	Indo-European	Eurasia	A	A
czt	Zotung Chin	Sino-Tibetan	Eurasia	A	A
daa	Dangaleat	Afro-Asiatic	Eurasia	D	DxX
dad	Marik	Austronesian	Papunesia	B	B
dah	Gwahatike	Nuclear Trans New Guinea	Papunesia	B	B
dak	Dakota	Siouan	North America	C	C
dan	Danish	Indo-European	Eurasia	C	C
dar	North-Central Dargwa	Nakh- Daghestanian	Eurasia	A	A
ddo	Tsez	Nakh- Daghestanian	Eurasia	A	A
ded	Dedua	Nuclear Trans New Guinea	Papunesia	A	A
des	Desano	Tucanoan	South America	A	A
deu	German	Indo-European	Eurasia	C	C
dgc	Casiguran- Nagtipunan Agta	Austronesian	Papunesia	D	D
dgi	Northern Dagara	Atlantic-Congo	Africa	D	DxX
dgr	Dogrib	Athabaskan- Eyak-Tlingit	North America	C	C
dgz	Daga	Dagan	Papunesia	A	A
dhg	Dhangu	Pama-Nyungan	Australia	E	E
dhm	Zemba	Atlantic-Congo	Africa	A?B	A?B
dig	Digo	Atlantic-Congo	Africa	A	A
dik	Southwestern Dinka	Nilotic	Africa	B?C?D	B?C?D
dip	Northeastern Dinka	Nilotic	Africa	D	D2c
dis	Dimasa	Sino-Tibetan	Eurasia	A	A
dje	Zarma	Songhay	Africa	C	C
djk	Aukan	Indo-European	Eurasia	C	C
djr	Djambarrpuyngu	Pama-Nyungan	Australia	Other	Other
dks	Southeastern Dinka	Nilotic	Africa	E	E
dnj	Nuclear Dan	Mande	Africa	C	C
dob	Dobu	Austronesian	Papunesia	B	B
dop	Lukpa	Atlantic-Congo	Africa	BxE	BxE
dow	Doyayo	Atlantic-Congo	Africa	C	C
dru	Rukai	Austronesian	Papunesia	C	C
dtp	Kadazan Dusun	Austronesian	Papunesia	D	DxX

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
dts	Toro So Dogon	Dogon	Africa	A	A
due	Umiray Dumaget Agta	Austronesian	Papunesia	C	C
dug	Duruma	Atlantic-Congo	Africa	A	A
duo	Dupaninan Agta	Austronesian	Papunesia	C	C
dur	Dii	Atlantic-Congo	Africa	A	A
dwr	Dawro	Ta-Ne-Omotic	Africa	A	A
dww	Dawawa	Austronesian	Papunesia	A	A
dyi	Djimini Senoufo	Atlantic-Congo	Africa	D	DxX
dyo	Jola-Fonyi	Atlantic-Congo	Africa	C	C
dyu	Dyula	Mande	Africa	D	D
ebk	Eastern Bontok	Austronesian	Papunesia	B	B
efi	Efik	Atlantic-Congo	Africa	A	A
eka	Ekajuk	Atlantic-Congo	Africa	D	DxX
eko	Koti	Atlantic-Congo	Africa	D	D
ell	Modern Greek	Indo-European	Eurasia	D	D
emi	Mussau-Emira	Austronesian	Papunesia	A	A
emp	Northern Emberá	Chocoan	South America	B	B
enb	Markweeta	Nilotic	Africa	C	AxC
eng	English	Indo-European	Eurasia	A	A
enm	Middle English	Indo-European	Eurasia	A	A
enx	Enxet Sur	Lengua-Mascoy	South America	D	D
epo	Esperanto	Artificial Language	Papunesia	A	A
eri	Ogea	Nuclear Trans New Guinea	Papunesia	B	B
ese	Ese Ejja	Pano-Tacanan	South America	A	A
esi	North Alaskan Inupiatun	Eskimo-Aleut	Eurasia	B	B
esk	Seward Alaska Inupiatun	Eskimo-Aleut	Eurasia	A	A
est	Estonian	Uralic	Eurasia	A	A
esu	Central Alaskan Yupik	Eskimo-Aleut	Eurasia	A	A
etr	Edolo	Bosavi	Papunesia	A	A
etu	Ejagham	Atlantic-Congo	Africa	B	B
eus	Basque	Isolate	Eurasia	A	A
eve	Even	Tungusic	Eurasia	A	A
ewe	Ewe	Atlantic-Congo	Africa	C	C
ewo	Ewondo	Atlantic-Congo	Africa	A	A
eza	Ezaa	Atlantic-Congo	Africa	B	B
faa	Fasu	Isolate	Eurasia	A	A
fad	Wagi	Nuclear Trans New Guinea	Papunesia	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
fai	Faiwol	Nuclear Trans New Guinea	Papunesia	B	B
fal	South Fali	Atlantic-Congo	Africa	C	AxC3
fao	Faroese	Indo-European	Eurasia	A	A
ffm	Maasina Fulfulde	Atlantic-Congo	Africa	C	C
fij	Fijian	Austronesian	Papunesia	A	A
fil	Filipino	Austronesian	Papunesia	C	C3
fin	Finnish	Uralic	Eurasia	A	A
fon	Fon	Atlantic-Congo	Africa	B	B
for	Fore	Nuclear Trans New Guinea	Papunesia	A	A
fra3	French	Indo-European	Eurasia	A	A
fry	Western Frisian	Indo-European	Eurasia	C	C
fub	Adamawa Fulfulde	Atlantic-Congo	Africa	C	C
fue	Borgu Fulfulde	Atlantic-Congo	Africa	C	C
fuf	Pular	Atlantic-Congo	Africa	D	D
fuh	Western Niger Fulfulde	Atlantic-Congo	Africa	C	CxA
fuq	Central-Eastern Niger Fulfulde	Atlantic-Congo	Africa	C	C
fuv	Hausa States Ful- fulde	Atlantic-Congo	Africa	D	D
gaa	Ga	Atlantic-Congo	Africa	C	C
gag	Gagauz	Turkic	Eurasia	A	A
gah	Alekano	Nuclear Trans New Guinea	Papunesia	A	A
gai	Borei	Lower Sepik- Ramu	Papunesia	E	E
gam	Kandawo	Nuclear Trans New Guinea	Papunesia	E	E
gaw	Nobonob	Nuclear Trans New Guinea	Papunesia	E	E
gbi	Galela	North Halmahera	Papunesia	A	A
gbo	Northern Grebo	Kru	Africa	D	DxX
gbr	Gbagyi	Atlantic-Congo	Africa	B	B
gde	Gude	Afro-Asiatic	Eurasia	D	D
gdg	Ga'dang	Austronesian	Papunesia	D	D2a
gdn	Umanakaina	Dagan	Papunesia	E	E
gdr	Wipi	Eastern Trans- Fly	Papunesia	C	C
geb	Kire	Lower Sepik- Ramu	Papunesia	C	C
gej	Gen	Atlantic-Congo	Africa	D	D2c

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
gfk	Patpatar	Austronesian	Papunesia	E	AxE
ghe	Southern Ghale	Sino-Tibetan	Eurasia	B	BxA
ghs	Guhu-Samane	Nuclear Trans New Guinea	Papunesia	A	A
gid	Gidar	Afro-Asiatic	Eurasia	B?C?D	B?C?D
gil	Kiribati	Austronesian	Papunesia	C	C
giz	South Giziga	Afro-Asiatic	Eurasia	C	C
gjn	Gonja	Atlantic-Congo	Africa	D	DxX
gkn	Gokana	Atlantic-Congo	Africa	A	A
gkp	Guinea Kpelle	Mande	Africa	A	A
gla	Scottish Gaelic	Indo-European	Eurasia	A	A
gle	Irish	Indo-European	Eurasia	A	A
glk	Gilaki	Indo-European	Eurasia	A	A
glv	Manx	Indo-European	Eurasia	A	A
gmv	Gamo	Ta-Ne-Omoti	Africa	B	B
gnb	Gangte	Sino-Tibetan	Eurasia	B?C?D	B?C?D
gnd	Zulgo-Gemzek	Afro-Asiatic	Eurasia	B?C?D	B?C?D
gng	Ngangam	Atlantic-Congo	Africa	D	D
gnn	Gumatj	Pama-Nyungan	Australia	A	A
gnw	Western Bolivian Guaraní	Tupian	South America	A	A
gof	Gofa	Ta-Ne-Omoti	Africa	A	A
gog	Gogo	Atlantic-Congo	Africa	D	D2c
gom	Goan Konkani	Indo-European	Eurasia	B	B
gor	Gorontalo	Austronesian	Papunesia	C	C
got	Gothic	Indo-European	Eurasia	B	B
gqr	Gor	Central Sudanic	Africa	B	AxB
grc	Ancient Greek	Indo-European	Eurasia	D	D
grt	Garo	Sino-Tibetan	Eurasia	B	B
gso	Southwest Gbaya	Atlantic-Congo	Africa	A	A
gub	Guaajára	Tupian	South America	B	B
guc	Wayuu	Arawakan	South America	A	A
gud	Yocoboué Dida	Kru	Africa	C	CxA
gug	Paraguayan Guaraní	Tupian	South America	B?C?D	B?C?D
guh	Guahibo	Guahiboan	South America	E	E
gui	Eastern Bolivian Guaraní	Tupian	South America	A	A
guj	Gujarati	Indo-European	Eurasia	A	A
guk	Northern Gumuz	Gumuz	Africa	A	A
gul	Sea Island Creole English	Indo-European	Eurasia	A	A
gum	Guambiano	Barbacoan	South America	A	A
gun	Mbyá Guaraní	Tupian	South America	B	B

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
guo	Guayabero	Guahiboan	South America	A	A
guq	Aché	Tupian	South America	A	A
gur	Farefare	Atlantic-Congo	Africa	B?C?D	B?C?D
guw	Gun	Atlantic-Congo	Africa	C	CxA
gux	Gourmanchéma	Atlantic-Congo	Africa	D	D
guz	Gusii	Atlantic-Congo	Africa	E	E
gvc	Kotiria	Tucanoan	South America	A	A
gvf	Golin	Nuclear Trans New Guinea	Papunesia	A	A
gvl	Gulay	Central Sudanic	Africa	A	A
gvn	Kuku-Yalanji	Pama-Nyungan	Australia	C	C
gwi	Gwich'in	Athabaskan- Eyak-Tlingit	North America	E	ExA
gya	Northwest Gbaya	Atlantic-Congo	Africa	D	DxX
gym	Ngäbere	Chibchan	North America	B	B
gyr	Guarayu	Tupian	South America	C	C
hae	Eastern Oromo	Afro-Asiatic	Eurasia	E	E
hag	Hanga	Atlantic-Congo	Africa	C	C
hak	Hakka Chinese	Sino-Tibetan	Eurasia	A	A
hat	Haitian	Indo-European	Eurasia	A	A
hau	Hausa	Afro-Asiatic	Eurasia	D	DxX
haw	Hawaiian	Austronesian	Papunesia	Other	Other
hay	Haya	Atlantic-Congo	Africa	D	DxX
hch	Huichol	Uto-Aztecan	North America	B	B
heb	Modern Hebrew	Afro-Asiatic	Eurasia	A	A
heg	Helong	Austronesian	Papunesia	D	D
heh	Hehe	Atlantic-Congo	Africa	A	A
hif	Fiji Hindi	Indo-European	Eurasia	A	A
hig	Kamwe	Afro-Asiatic	Eurasia	A	A
hil	Hiligaynon	Austronesian	Papunesia	C	C
hin	Hindi	Indo-European	Eurasia	A	A
hix	Hixkaryána	Cariban	South America	A	A
hla	Halia	Austronesian	Papunesia	A	A
hlt	Nga La	Sino-Tibetan	Eurasia	A	A
hmo	Hiri Motu	Pidgin	Australia	A	A
hmr	Hmar	Sino-Tibetan	Eurasia	D	D
hne	Chhattisgarhi	Indo-European	Eurasia	A	A
hnj	Hmong Njua	Hmong-Mien	Eurasia	A	A
hnn	Hanunoo	Austronesian	Papunesia	C	C
hns	Caribbean Hindustani	Indo-European	Eurasia	A	A
hop	Hopi	Uto-Aztecan	North America	D	D
hot	Hote	Austronesian	Papunesia	A	A
hra	Hrangkhoh	Sino-Tibetan	Eurasia	C	C

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
hrv	Croatian Standard	Indo-European	Eurasia	A	A
hto	Minica Huitoto	Huitotoan	South America	E	E
hub	Huambisa	Chicham	South America	A	A
hui	Huli	Nuclear Trans New Guinea	Papunesia	A	A
hun	Hungarian	Uralic	Eurasia	C	CxA
hus	Huastec	Mayan	North America	A	A
huu	Murui Huitoto	Huitotoan	South America	A	A
huv	San Mateo del Mar Huave	Huavean	North America	A	A
hvn	Hawu	Austronesian	Papunesia	D	DxX
hwc	Hawai'i Creole English	Indo-European	Eurasia	A	A
hye	Eastern Armenian	Indo-European	Eurasia	A	A
ian	Iatmul	Ndu	Papunesia	A	A
iba	Iban	Austronesian	Papunesia	A	A
ibo	Igbo	Atlantic-Congo	Africa	A	A
icr	San Andres Creole English	Indo-European	Eurasia	A	A
ifa	Amganad Ifugao	Austronesian	Papunesia	C	C
ifb	Batad Ifugao	Austronesian	Papunesia	B	B
ife	Ifè	Atlantic-Congo	Africa	C	C
ifk	Tuwali Ifugao	Austronesian	Papunesia	C	C3
ifu	Mayoyao Ifugao	Austronesian	Papunesia	B?C?D	B?C?D
ify	Keley-i Kallahan	Austronesian	Papunesia	D	D
ign	Ignaciano	Arawakan	South America	A	A
ikk	Ika	Atlantic-Congo	Africa	A	A
ikw	Ikwere	Atlantic-Congo	Africa	A	A
ilb	Ila	Atlantic-Congo	Africa	B	B
ilo	Iloko	Austronesian	Papunesia	C	C3
imo	Imbongu	Nuclear Trans New Guinea	Papunesia	B	B
inb	Colombian Inga	Quechuan	South America	A	A
ind	Standard Indonesian	Austronesian	Papunesia	C	C
ino	Inoke-Yate	Nuclear Trans New Guinea	Papunesia	A	A
iou	Tuma-Irumu	Nuclear Trans New Guinea	Papunesia	A	A
ipi	Ipili	Nuclear Trans New Guinea	Papunesia	B	B
iqw	Ikwo	Atlantic-Congo	Africa	B	B

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
iri	Irigwe	Atlantic-Congo	Africa	C	CxA
irk	Iraqw	Afro-Asiatic	Eurasia	E	E
iry	Iraya	Austronesian	Papunesia	D	DxX
isd	Isnag	Austronesian	Papunesia	C	C
isl	Icelandic	Indo-European	Eurasia	A	A
ita	Italian	Indo-European	Eurasia	A	A
itv	Itawit	Austronesian	Papunesia	C	C3
ium	Iu Mien	Hmong-Mien	Eurasia	B	B
ivb	Ibatan	Austronesian	Papunesia	C	C3
ivv	Itbayat	Austronesian	Papunesia	C	C
iws	Sepik Iwam	Sepik	Papunesia	Other	Other
ixl	Ixil	Mayan	North America	B	B
izr	Izere	Atlantic-Congo	Africa	C	C
izz	Izi	Atlantic-Congo	Africa	B	B
jac	Popti'	Mayan	North America	B	BxA
jae	Yabem	Austronesian	Papunesia	D	D
jam	Jamaican Creole English	Indo-European	Eurasia	A	A
jav	Javanese	Austronesian	Papunesia	D	D2a
jbu	Jukun Takum	Atlantic-Congo	Africa	E	ExA
jic	Tol	Jicaquean	North America	A	A
jiv	Shuar	Chicham	South America	A	A
jmc	Machame	Atlantic-Congo	Africa	B	B
jpn	Japanese	Japonic	Eurasia	A	A
jra	Jarai	Austronesian	Papunesia	E	E3
jvn	Caribbean Javanese	Austronesian	Papunesia	D	D
kaa	Kara-Kalpak	Turkic	Eurasia	A	A
kab	Kabyle	Afro-Asiatic	Eurasia	D	DxX
kac	Southern Jinghpaw	Sino-Tibetan	Eurasia	D	D
kal	Kalaallisut	Eskimo-Aleut	Eurasia	A	A
kan	Kannada	Dravidian	Eurasia	C	C
kao	Xaasongaxango	Mande	Africa	C	C
kaq	Capanahua	Pano-Tacanan	South America	A	A
kat	Georgian	Kartvelian	Eurasia	B	BxA
kaz	Kazakh	Turkic	Eurasia	A	A
kbc	Kadiwéu	Guaicuruan	South America	C	C
kbd	Kabardian	Abkhaz-Adyge	Eurasia	C	C
kbh	Camsá	Isolate	Eurasia	A	A
kbm	Iwal	Austronesian	Papunesia	D	D
kbp	Kabiyé	Atlantic-Congo	Africa	D	DxX
kbq	Kamano	Nuclear Trans New Guinea	Papunesia	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
kbr	Kafa	Ta-Ne-Omotic	Africa	A	A
kck	Kalanga	Atlantic-Congo	Africa	D	D2c
kdc	Kutu	Atlantic-Congo	Africa	A	A
kde	Makonde	Atlantic-Congo	Africa	A	A
kdh	Tem	Atlantic-Congo	Africa	C	C3
kdi	Kumam	Nilotic	Africa	C	C3
kdj	Karamojong	Nilotic	Africa	E	E
kdl	Tsikimba	Atlantic-Congo	Africa	D	DxX
kek	Kekchí	Mayan	North America	A	A
ken	Kenyang	Atlantic-Congo	Africa	B	B
ket	Ket	Yeniseian	Eurasia	A	A
kew	West Kewa	Nuclear Trans New Guinea	Papunesia	B	B
kez	Kukele	Atlantic-Congo	Africa	B	B
kff	Koya	Dravidian	Eurasia	A	A
kgf	Kulungfu- Yuanggeng-Tobo	Nuclear Trans New Guinea	Papunesia	B	B
kgk	Kaiwá	Tupian	South America	B	B
kgp	Kaingang	Nuclear-Macro-Je	South America	A	A
kha	Khasi	Austroasiatic	Eurasia	E	ExA
khk	Halh Mongolian	Mongolic-Khitán	Eurasia	A	A
khm	Central Khmer	Austroasiatic	Eurasia	B	AxB
khs	Kasua	Bosavi	Papunesia	A	A
khy	Kele-Foma	Atlantic-Congo	Africa	A	A
khz	Keapara	Austronesian	Papunesia	E	E
kia	Kim	Atlantic-Congo	Africa	A	A
kik	Kikuyu	Atlantic-Congo	Africa	A	A
kin	Kinyarwanda	Atlantic-Congo	Africa	B	B
kir	Kirghiz	Turkic	Eurasia	A	A
kix	Khiamniungan Naga	Sino-Tibetan	Eurasia	A	A
kjb	Q'anjob'al	Mayan	North America	A	A
kje	Kisar	Austronesian	Papunesia	D	DxX
kjh	Khakas	Turkic	Eurasia	A	A
kjs	East Kewa	Nuclear Trans New Guinea	Papunesia	B	B
kkc	Odoodee	East Strickland	Papunesia	A	A
kki	Kagulu	Atlantic-Congo	Africa	A	A
kkj	Kako	Atlantic-Congo	Africa	D	D
kkk	Kosarek Yale	Nuclear Trans New Guinea	Papunesia	A	A
klt	Nukna	Nuclear Trans New Guinea	Papunesia	B	B
klv	Maskelynes	Austronesian	Papunesia	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
kma	Konni	Atlantic-Congo	Africa	D	DxX
kmg	Kâte	Nuclear Trans New Guinea	Papunesia	A	A
kmh	Kalam	Nuclear Trans New Guinea	Papunesia	E	E
kmk	Limos Kalinga	Austronesian	Papunesia	D	DxX
kmm	Kom	Sino-Tibetan	Eurasia	A	A
kmo	Kwoma	Sepik	Papunesia	A	A
kmr	Northern Kurdish	Indo-European	Eurasia	B	AxB
kms	Kamasau	Nuclear Torricelli	Papunesia	B	B
kmu	Kanite	Nuclear Trans New Guinea	Papunesia	A	A
kne	Kankanaey	Austronesian	Papunesia	D	D
knf	Mankanya	Atlantic-Congo	Africa	C	C
kng	South-Central Koongo	Atlantic-Congo	Africa	A	A
knj	Akateko	Mayan	North America	A	A
knk	Kuranko	Mande	Africa	A	A
kno	Kono	Mande	Africa	D	DxX
knv	Tabo	Isolate	Eurasia	A	A
kog	Cogui	Chibchan	North America	A	A
kor	Korean	Koreanic	Eurasia	D	DxX
kpf	Komba	Nuclear Trans New Guinea	Papunesia	A	A
kpg	Kapingamarangi	Austronesian	Papunesia	A	A
kpj	Karajá	Nuclear-Macro-Je	South America	B	B
kpr	Korafe-Yegha	Nuclear Trans New Guinea	Papunesia	A	A
kpv	Komi-Zyrian	Uralic	Eurasia	B	B
kpw	Kobon	Nuclear Trans New Guinea	Papunesia	E	E
kpx	Mountain Koiali	Koiarian	Papunesia	A	A
kpz	Kupsabiny	Nilotic	Africa	B	B
kqc	Doromu-Koki	Manubaran	Papunesia	A	A
kqe	Kalagan	Austronesian	Papunesia	B?C?D	B?C?D
kqf	Kakabai	Austronesian	Papunesia	E	E
kql	Kyenele	Yuat	Papunesia	D	D
kqo	Konobo-Eastern Krahn	Kru	Africa	D	D
kqp	Kimre	Afro-Asiatic	Eurasia	A	A
kqs	Northern Kissi	Atlantic-Congo	Africa	D	DxX
kqw	Kandas	Austronesian	Papunesia	A	A
kqy	Koorete	Ta-Ne-Omotic	Africa	A	A
krc	Karachay-Balkar	Turkic	Eurasia	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
kri	Krio	Indo-European	Eurasia	E	ExA
krj	Kinaray-a	Austronesian	Papunesia	C	C
kru	Kurukh	Dravidian	Eurasia	A	A
ksb	Shambala	Atlantic-Congo	Africa	C	C
ksc	Bangad	Austronesian	Papunesia	D	D
ksd	Kuanua	Austronesian	Papunesia	A	A
ksf	Bafia	Atlantic-Congo	Africa	C	CxA
ksj	Uare	Kwalean	Papunesia	A	A
ksp	Kaba	Central Sudanic	Africa	E	E
ksr	Borong	Nuclear Trans New Guinea	Papunesia	A	A
kss	Southern Kisi	Atlantic-Congo	Africa	C	C
ksw	S'gaw Karen	Sino-Tibetan	Eurasia	E	ExA
ktb	Kambaata	Afro-Asiatic	Eurasia	A	A
ktj	Plapo Krumen	Kru	Africa	C	C
ktm	Kurti	Austronesian	Papunesia	C	C
kto	Kuot	Isolate	Eurasia	A	A
ktu	Kituba	Atlantic-Congo	Africa	A	A
kua	Kuanyama	Atlantic-Congo	Africa	C	C
kub	Kutep	Atlantic-Congo	Africa	E	AxE
kud	'Auhelawa	Austronesian	Papunesia	D	DxX
kue	Kuman	Nuclear Trans New Guinea	Papunesia	B	B
kuj	Kuria	Atlantic-Congo	Africa	E	ExA
kum	Kumyk	Turkic	Eurasia	B	B
kup	Kunimaipa	Kunimaipan	Papunesia	E	E
kus	Kusaal	Atlantic-Congo	Africa	B?C?D	B?C?D
kvj	Psikye	Afro-Asiatic	Eurasia	C	C
kvn	Border Kuna	Chibchan	North America	B	B
kwd	Kwaio	Austronesian	Papunesia	A	A
kwf	Kwara'ae	Austronesian	Papunesia	A	A
kwi	Awa-Cuaiquer	Barbacoan	South America	A	A
kwj	Kwanga	Sepik	Papunesia	E	E
kxc	Konso	Afro-Asiatic	Eurasia	C	C3
kxm	Northern Khmer	Austroasiatic	Eurasia	D	D
kxw	Konai	East Strickland	Papunesia	E	E
kyc	Kyaka	Nuclear Trans New Guinea	Papunesia	B	B
kyf	Kouya	Kru	Africa	A	A
kyg	Keyagana	Nuclear Trans New Guinea	Papunesia	A	A
kyq	Kenga	Central Sudanic	Africa	C	CxA
kyu	Western Kayah	Sino-Tibetan	Eurasia	A	A
kyz	Kayabí	Tupian	South America	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
kze	Kosena	Nuclear Trans New Guinea	Papunesia	A	A
kzf	Da'a Kaili	Austronesian	Papunesia	D	DxX
lac	Lacandon	Mayan	North America	C	C
lai	Lambya	Atlantic-Congo	Africa	C	CxA
laj	Lango	Nilotic	Africa	E	ExA
lam	Lamba	Atlantic-Congo	Africa	A	A
las	Lama	Atlantic-Congo	Africa	A	A
lat	Latin	Indo-European	Eurasia	A	A
lav	Latvian	Indo-European	Eurasia	A	A
lbb	Label	Austronesian	Papunesia	A	A
lbj	Leh Ladakhi	Sino-Tibetan	Eurasia	B	B
lbk	Central Bontoc	Austronesian	Papunesia	D	DxX
lcm	Tungag	Austronesian	Papunesia	C	C
ldi	Laari	Atlantic-Congo	Africa	A	A
lee	Lyélé	Atlantic-Congo	Africa	A	A
lef	Lelemi	Atlantic-Congo	Africa	D	DxX
leg	Enlhet Norte	Lengua-Mascoy	South America	B	B
leh	Lenje	Atlantic-Congo	Africa	C	C
lem	Nomaande	Atlantic-Congo	Africa	B	B
leu	Kara	Austronesian	Papunesia	A	A
lew	Ledo Kaili	Austronesian	Papunesia	C	C
lex	Luang	Austronesian	Papunesia	A?E	A?E
lgm	Lega-Mwenga	Atlantic-Congo	Africa	C	C
lhi	Lahu Shi	Sino-Tibetan	Eurasia	A	A
lhm	Lhomi	Sino-Tibetan	Eurasia	A	A
lhu	Lahu	Sino-Tibetan	Eurasia	B	B
lia	West-Central Limba	Atlantic-Congo	Africa	C	CxA
lid	Nyindrou	Austronesian	Papunesia	B?C?D	B?C?D
lif	Limbu	Sino-Tibetan	Eurasia	A	A
lin	Kinshasa Lingala	Atlantic-Congo	Africa	A	A
lip	Sekpele	Atlantic-Congo	Africa	D	D
lit	Lithuanian	Indo-European	Eurasia	B	B
ljp	Lampung Api	Austronesian	Papunesia	C	C
lmk	Lamkang	Sino-Tibetan	Eurasia	B	B
lmp	Limbum	Atlantic-Congo	Africa	BxE	BxE
lob	Lobi	Atlantic-Congo	Africa	A	A
log	Logo	Central Sudanic	Africa	D	D
lol	Mongo	Atlantic-Congo	Africa	C	C3
lom	Loma	Mande	Africa	C	C3
loq	Lobala	Atlantic-Congo	Africa	D	D
loz	Lozi	Atlantic-Congo	Africa	A	A
lsi	Lashi	Sino-Tibetan	Eurasia	BxE	BxE

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
lsm	Saamia	Atlantic-Congo	Africa	A	A
lug	Ganda	Atlantic-Congo	Africa	C	CxA
luo	Luo	Nilotic	Africa	C	C
lus	Mizo	Sino-Tibetan	Eurasia	D	D
lwo	Luwo	Nilotic	Africa	C	C
lww	Lewo	Austronesian	Papunesia	A	A
lzh	Literary Chinese	Sino-Tibetan	Eurasia	A	A
maa	San Jerónimo Tecóatl Mazatec	Otomanguean	North America	A	A
mad	Madurese	Austronesian	Papunesia	C	C
maf	Mafa	Afro-Asiatic	Eurasia	A	A
mah	Marshallese	Austronesian	Papunesia	C	C
mai	Maithili	Indo-European	Eurasia	A	A
maj	Jalapa De Díaz Mazatec	Otomanguean	North America	A	A
mak	Makasar	Austronesian	Papunesia	C	C
mal	Malayalam	Dravidian	Eurasia	A	A
mam	Mam	Mayan	North America	A	A
maq	Chiquihuitlán Mazatec	Otomanguean	North America	A	A
mar	Marathi	Indo-European	Eurasia	B	B
mas	Masai	Nilotic	Africa	E	ExA
mau	Huautla Mazatec	Otomanguean	North America	A	A
mav	Sateré-Mawé	Tupian	South America	B	B
maw	Mampruli	Atlantic-Congo	Africa	C	C
maz	Central Mazahua	Otomanguean	North America	A	A
mbb	Western Bukidnon Manobo	Austronesian	Papunesia	C	C
mbc	Macushi	Cariban	South America	B	B
mbd	Dibabawon Manobo	Austronesian	Papunesia	C	C
mbf	Baba Malay	Austronesian	Papunesia	A	A
mbh	Mangseng	Austronesian	Papunesia	B?D?E	B?D?E
mbi	Ilianen Manobo	Austronesian	Papunesia	C	C
mbj	Nadëb	Naduhup	South America	D	D
mbk	Maxakalí	Nuclear-Macro-Je	South America	B	B
mbs	Sarangani Manobo	Austronesian	Papunesia	C	C
mbt	Matigsalug Manobo	Austronesian	Papunesia	C	C3
mca	Maca	Matacoan	South America	A	A
mcb	Machiguenga	Arawakan	South America	C	C3
mcd	Sharanahua	Pano-Tacanan	South America	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
mcf	Matsés	Pano-Tacanan	South America	A	A
mck	Mbunda	Atlantic-Congo	Africa	A	A
mcn	Masana	Afro-Asiatic	Eurasia	A	A
mco	Coatlán Mixe	Mixe-Zoque	North America	A	A
mcp	Makaa	Atlantic-Congo	Africa	A	A
mcq	Ese	Koiarian	Papunesia	B	B
mcu	Donga Mambila	Atlantic-Congo	Africa	C	C
mda	Mada	Atlantic-Congo	Africa	A	A
mdy	Male	Ta-Ne-Omotic	Africa	B	B
med	Melpa	Nuclear Trans New Guinea	Papunesia	E	E
mee	Mengen	Austronesian	Papunesia	B	BxA
mej	Meyah	East Bird's Head	Papunesia	B	B
mek	Mekeo	Austronesian	Papunesia	A	A
men	Mende	Mande	Africa	D	DxX
meq	Merey	Afro-Asiatic	Eurasia	B	B
meu	Motu	Austronesian	Papunesia	B	B
mfe	Morisyen	Indo-European	Eurasia	BxC	BxC
mfh	Matal	Afro-Asiatic	Eurasia	D	DxX
mfi	Wandala	Afro-Asiatic	Eurasia	D	D
mfk	North Mofu	Afro-Asiatic	Eurasia	C	CxA
mfq	Moba	Atlantic-Congo	Africa	A	A
mfy	Mayo	Uto-Aztecan	North America	A	A
mfz	Mabaan	Nilotic	Africa	C	CxA
mgc	Morokodo	Central Sudanic	Africa	C	C
mgh	Makhuwa-Meetto	Atlantic-Congo	Africa	C	C
mgr	Mambwe-Lungu	Atlantic-Congo	Africa	A	A
mhi	Ma'di	Central Sudanic	Africa	C	C
mhl	Mauwake	Nuclear Trans New Guinea	Papunesia	A	A
mhr	Eastern Mari	Uralic	Eurasia	D	D
mhx	Maru	Sino-Tibetan	Eurasia	E	E
mhy	Ma'anyan	Austronesian	Papunesia	C	C
mib	Atatláhuca Mixtec	Otomanguean	North America	D	D
mic	Mi'kmaq	Algic	North America	A	A
mie	Ocotepec Mixtec	Otomanguean	North America	A	A
mif	Mofu-Gudur	Afro-Asiatic	Eurasia	C	CxA
mig	San Miguel El Grande Mixtec	Otomanguean	North America	B?C?D	B?C?D
mih	Chayuco Mixtec	Otomanguean	North America	A	A
mil	Peñoles Mixtec	Otomanguean	North America	A	A
min	Minangkabau	Austronesian	Papunesia	D	DxX

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
mio	Pinotepa Nacional Mixtec	Otomanguean	North America	B?D?E	B?D?E
miq	Mískito	Misumalpan	North America	A	A
mir	Isthmus Mixe	Mixe-Zoque	North America	A	A
mit	Southern Puebla Mixtec	Otomanguean	North America	A	A
miy	Ayutla Mixtec	Otomanguean	North America	A	A
miz	Coatzospan Mixtec	Otomanguean	North America	D	DxX
mjc	San Juan Colorado Mixtec	Otomanguean	North America	A	A
mjw	Hills Karbi	Sino-Tibetan	Eurasia	B	B
mkd	Macedonian	Indo-European	Eurasia	B	BxA
mkl	Mokole	Atlantic-Congo	Africa	D	D
mkn	Kupang Malay	Austronesian	Papunesia	C	C
mks	Silacayoapan Mixtec	Otomanguean	North America	A	A
mlh	Mape	Nuclear Trans New Guinea	Papunesia	B	B
mlp	Bargam	Nuclear Trans New Guinea	Papunesia	A	A
mlt	Maltese	Afro-Asiatic	Eurasia	A	A
mmn	Mamanwa	Austronesian	Papunesia	A	A
mmo	Mangga Buang	Austronesian	Papunesia	A	A
mmx	Madak	Austronesian	Papunesia	D	D
mna	Mbula	Austronesian	Papunesia	D	D
mnb	Muna	Austronesian	Papunesia	B?C?D	B?C?D
mnf	Mundani	Atlantic-Congo	Africa	D	D
mnh	Mono	Atlantic-Congo	Africa	D	D
mnk	Mandinka	Mande	Africa	D	DxX
mnx	Sougb	East Bird's Head	Papunesia	A	A
moa	Mwan	Mande	Africa	D	D
moc	Mocoví	Guaicuruan	South America	C	C
mog	Mongondow	Austronesian	Papunesia	D	D
moh	Mohawk	Iroquoian	North America	CxE	CxE
mop	Mopán Maya	Mayan	North America	A	A2
mor	Moro	Heibanic	Africa	A	A
mos	Mossi	Atlantic-Congo	Africa	C	C
mox	Molima	Austronesian	Papunesia	A	A
mpg	Marba	Afro-Asiatic	Eurasia	C	CxA
mph	Mawng	Iwaidjan Proper	Australia	B	B
mpm	Yosondúa Mixtec	Otomanguean	North America	A	A
mpp	Migabac	Nuclear Trans New Guinea	Papunesia	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
mps	Dadibi	Teberan	Papunesia	D	D
mpt	Mian	Nuclear Trans New Guinea	Papunesia	E	E
mpx	Misima-Paneati	Austronesian	Papunesia	C	CxA
mqb	Mbukoko	Afro-Asiatic	Eurasia	B	B
mqj	Mamasa	Austronesian	Papunesia	C	C
mqy	Manggarai	Austronesian	Papunesia	C	C
mrg	Mising-Padam- Miri-Minyong	Sino-Tibetan	Eurasia	A	A
mri	Maori	Austronesian	Papunesia	B?C?D	B?C?D
mrw	Maranao	Austronesian	Papunesia	D	DxX
msa	Central Malay	Austronesian	Papunesia	D	DxX
msb	Masbatenyoo	Austronesian	Papunesia	C	C
msc	Sankaran Maninka	Mande	Africa	A	A
mse	Musey	Afro-Asiatic	Eurasia	E	E
msk	Mansaka	Austronesian	Papunesia	C	C
msm	Agusan Manobo	Austronesian	Papunesia	C	C
msy	Aruamu	Lower Sepik- Ramu	Papunesia	B	B
mta	Cotabato Manobo	Austronesian	Papunesia	C	C
mtg	Una	Nuclear Trans New Guinea	Papunesia	A	A
mti	Maiwa	Dagan	Papunesia	E	E
mtj	Moskona	East Bird's Head	Papunesia	B	B
mto	Totontepec Mixe	Mixe-Zoque	North America	A	A
mtp	Wichí Lhamtés Nocten	Matacoan	South America	B?C?D	B?C?D
mua	Mundang	Atlantic-Congo	Africa	A	A
mug	Musgu	Afro-Asiatic	Eurasia	A	A
muh	Mündü	Atlantic-Congo	Africa	A	A
mur	Murle	Surmic	Africa	C	C
mux	Bo-Ung	Nuclear Trans New Guinea	Papunesia	B	B
muy	Muyang	Afro-Asiatic	Eurasia	D	DxX
mva	Manam	Austronesian	Papunesia	A	A
mvn	Minaveha	Austronesian	Papunesia	D	D
mvp	Duri	Austronesian	Papunesia	C	C3
mwc	Are	Austronesian	Papunesia	A	A
mwf	Murriny Patha	Southern Daly	Australia	A	A
mwL	Mirandese	Indo-European	Eurasia	A	A
mwm	Sar	Central Sudanic	Africa	C	CxA
mwp	Kala Lagaw Ya	Pama-Nyungan	Australia	C	C

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
mwq	Mün Chin	Sino-Tibetan	Eurasia	A	A
mwv	Mentawai	Austronesian	Papunesia	D	D
mww	Hmong Daw	Hmong-Mien	Eurasia	A	A
mxb	Tezoatlán Mixtec	Otomanguean	North America	A	A
mxp	Tlahuitoltepec Mixe	Mixe-Zoque	North America	A	A
mxq	Juquila Mixe	Mixe-Zoque	North America	A	A
mxt	Jamiltepec Mixtec	Otomanguean	North America	B	B
mxv	Metlatónoc Mix- tec	Otomanguean	North America	A	A2
mya	Burmese	Sino-Tibetan	Eurasia	A	A
myb	Mbay	Central Sudanic	Africa	A	A
myk	Mamara Senoufo	Atlantic-Congo	Africa	E	E
myu	Mundurukú	Tupian	South America	C	C
myv	Erzya	Uralic	Eurasia	A	A
myw	Muyuw	Austronesian	Papunesia	C	C
myx	Masaaba	Atlantic-Congo	Africa	E	AxE
myy	Macuna	Tucanoan	South America	A	A
mza	Santa María Za- catepec Mixtec	Otomanguean	North America	A	A
mzh	Wichí Lhamtés Güisnay	Matacoan	South America	C	C3
mzk	Western Mambila	Atlantic-Congo	Africa	D	D
mzl	Mazatlán Mixe	Mixe-Zoque	North America	A	A
mzm	Mumuye	Atlantic-Congo	Africa	A	A
mzw	Deg	Atlantic-Congo	Africa	D	D
mzz	Maiadomu	Austronesian	Papunesia	A	A
nab	Southern Nambikuára	Nambiquaran	South America	A	A
naf	Nabak	Nuclear Trans New Guinea	Papunesia	A	A
nak	Nakanai	Austronesian	Papunesia	B?C?D	B?C?D
nan	Min Nan Chinese	Sino-Tibetan	Eurasia	B	B
naq	Nama	Khoe-Kwadi	Africa	D	D
nas	Naasioi	South Bougainville	Papunesia	A	A
nav	Navajo	Athabaskan- Eyak-Tlingit	North America	E	E
naw	Nawuri	Atlantic-Congo	Africa	D	D
nbc	Chang Naga	Sino-Tibetan	Eurasia	A	A
nbe	Konyak Naga	Sino-Tibetan	Eurasia	A	A
nbl	South Transvaal Ndebele	Atlantic-Congo	Africa	B	B

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
nbq	Nggem	Nuclear Trans New Guinea	Papunesia	A	A
nbu	Rongmei Naga	Sino-Tibetan	Eurasia	E	E
nca	Iyo	Nuclear Trans New Guinea	Papunesia	B	B
nch	Central Huasteca Nahuatl	Uto-Aztecan	North America	A	A
ncj	Northern Puebla Nahuatl	Uto-Aztecan	North America	A	A
ncl	Michoacán Nahu- atl	Uto-Aztecan	North America	C	CxA
nct	Chothe	Sino-Tibetan	Eurasia	E	E
ncu	Chumburung	Atlantic-Congo	Africa	D	D
ndc	Ndau	Atlantic-Congo	Africa	BxE	BxE
nde	Zimbabwean Nde- bele	Atlantic-Congo	Africa	B	B
ndi	Samba Leko	Atlantic-Congo	Africa	D	DxX
ndj	Ndamba	Atlantic-Congo	Africa	A	A
ndo	Ndonga	Atlantic-Congo	Africa	C	C3
ndp	Ndo	Central Sudanic	Africa	C	CxA
nds	Eastern Low Ger- man	Indo-European	Eurasia	C	C
ndz	Ndogo	Atlantic-Congo	Africa	D	D
neb	Toura	Mande	Africa	D	D
nep	Nepali	Indo-European	Eurasia	A	A
nfa	Dhao	Austronesian	Papunesia	D	D
nfr	Nafanan	Atlantic-Congo	Africa	C	C
ngc	Ngombe	Atlantic-Congo	Africa	C	C
ngp	Ngulu	Atlantic-Congo	Africa	D	D2c
ngu	Central Guerrero Nahuatl	Uto-Aztecan	North America	A	A
nhd	Chiripá	Tupian	South America	A	A
nhe	Eastern Huasteca Nahuatl	Uto-Aztecan	North America	A	A
nhg	Tetelcingo Nahu- atl	Uto-Aztecan	North America	A	A
nhi	Zacatlán- Ahuacatlán- Tepetzintla Nahuatl	Uto-Aztecan	North America	A	A
nho	Takuu	Austronesian	Papunesia	A	A
nhr	Naro	Khoe-Kwadi	Africa	B?C?D	B?C?D
nhu	Noone	Atlantic-Congo	Africa	A?B	A?B
nhw	Western Huasteca Nahuatl	Uto-Aztecan	North America	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
nhx	Isthmus-Mecayapan Nahuatl	Uto-Aztecan	North America	A	A
nhy	Northern Oaxaca Nahuatl	Uto-Aztecan	North America	A	A
nia	Nias	Austronesian	Papunesia	C	C
nif	Nek	Nuclear Trans New Guinea	Papunesia	A	A
nii	Nii	Nuclear Trans New Guinea	Papunesia	B	B
nij	Ngaju	Austronesian	Papunesia	C	C
nim	Nilamba	Atlantic-Congo	Africa	A	A
nin	Ninzo	Atlantic-Congo	Africa	C	CxA
niq	Nandi	Nilotic	Africa	B	B
niv	Amur Nivkh	Nivkh	Eurasia	A	A
niy	Ngiti	Central Sudanic	Africa	B?C?D	B?C?D
njb	Nocte Naga	Sino-Tibetan	Eurasia	A	A
njm	Angami Naga	Sino-Tibetan	Eurasia	B	B
njn	Liangmai Naga	Sino-Tibetan	Eurasia	E	E
njo	Ao Naga	Sino-Tibetan	Eurasia	A	A
njz	Nyishi-Hill Miri	Sino-Tibetan	Eurasia	B	B
nki	Thangal Naga	Sino-Tibetan	Eurasia	C	C3
nko	Nkonya	Atlantic-Congo	Africa	C	C
nlc	Nalca	Nuclear Trans New Guinea	Papunesia	D	D
nld	Dutch	Indo-European	Eurasia	C	C3
nma	Maram Naga	Sino-Tibetan	Eurasia	B?D?E	B?D?E
nmf	East-Central Tangkhul Naga	Sino-Tibetan	Eurasia	A	A
nmh	Monsang Naga	Sino-Tibetan	Eurasia	A	A
nmo	Moyon	Sino-Tibetan	Eurasia	A	A
nmw	Nimoa	Austronesian	Papunesia	C	CxA
nmz	Nawdm	Atlantic-Congo	Africa	D	D
nnb	Nande	Atlantic-Congo	Africa	A	A
nng	Maring Naga	Sino-Tibetan	Eurasia	A	A
nnh	Ngiemboon	Atlantic-Congo	Africa	E	E
nnl	Northern Rengma Naga	Sino-Tibetan	Eurasia	D	D
nno	Norwegian Nynorsk	Indo-European	Eurasia	C	C
nnp	Wancho Naga	Sino-Tibetan	Eurasia	D	DxX
nnq	Ngindo	Atlantic-Congo	Africa	D	D
nnw	Southern Nuni	Atlantic-Congo	Africa	A	A
noa	Woun Meu	Chocoan	South America	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
nob	Norwegian Bokmål	Indo-European	Eurasia	C	C
nog	Nogai	Turkic	Eurasia	A	A
nop	Numanggang	Nuclear Trans New Guinea	Papunesia	A	A
not	Nomatsiguenga	Arawakan	South America	A	A
nou	Ewage-Notu	Nuclear Trans New Guinea	Papunesia	A	A
nph	Phom Naga	Sino-Tibetan	Eurasia	A	A
npi	Nepali	Indo-European	Eurasia	C	CxA
npl	Nahuatl, Southeastern Puebla	Uto-Aztecan	North America	A	A
npo	Pochuri Naga	Sino-Tibetan	Eurasia	A	A
npv	Napu	Austronesian	Papunesia	C	C
nre	Southern Rengma Naga	Sino-Tibetan	Eurasia	A	A
nrf	Jerriais	Indo-European	Eurasia	A	A
nri	Chokri Naga	Sino-Tibetan	Eurasia	A	A
nsa	Sangtam Naga	Sino-Tibetan	Eurasia	A	A
nse	Nsenga	Atlantic-Congo	Africa	C	C
nsm	Sumi Naga	Sino-Tibetan	Eurasia	A	A
nsn	Nehan	Austronesian	Papunesia	A	A
nso	Pedi	Atlantic-Congo	Africa	A	A
nss	Nali	Austronesian	Papunesia	C	C
nst	Pangwa Naga	Sino-Tibetan	Eurasia	C	C
nsu	Sierra Negra Nahuatl	Uto-Aztecan	North America	A	A
ntp	Northern Tepehuan	Uto-Aztecan	North America	C	C
ntr	Delo	Atlantic-Congo	Africa	D	D
nus	Nuer	Nilotic	Africa	C	C
nuy	Wubuy	Gunwinyguan	Australia	C	C
nvm	Namiaie	Koiarian	Papunesia	A	A
nwb	Nyabwa	Kru	Africa	B?C?D	B?C?D
nwi	Southwest Tanna	Austronesian	Papunesia	D	DxX
nwx	Kathmandu Valley Newari	Sino-Tibetan	Eurasia	B	B
nxd	Ngando-Lalia	Atlantic-Congo	Africa	C	C
nya	Nyanja	Atlantic-Congo	Africa	A	A
nyf	Giryama	Atlantic-Congo	Africa	A	A
nyn	Nyankole	Atlantic-Congo	Africa	E	E3
nyo	Nyoro	Atlantic-Congo	Africa	A	A
nyu	Nyungwe	Atlantic-Congo	Africa	C	C

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
nyy	Nyakyusa-Ngonde	Atlantic-Congo	Africa	C	CxA
nzm	Zeme Naga	Sino-Tibetan	Eurasia	D	D
obo	Obo Manobo	Austronesian	Papunesia	C	C
obj	Northwestern Ojibwa	Algic	North America	A	A
oji	Chippewa	Algic	North America	D	DxX
ojs	Severn Ojibwa	Algic	North America	A	A
okv	Orokaiva	Nuclear Trans New Guinea	Papunesia	A	A
old	Mochi	Atlantic-Congo	Africa	D	D
omw	South Tairora	Nuclear Trans New Guinea	Papunesia	A	A
ong	Olo	Nuclear Torricelli	Papunesia	A	A
ons	Ono	Nuclear Trans New Guinea	Papunesia	E	E
ood	Tohono O'odham	Uto-Aztecian	North America	C	C
opm	Oksapmin	Nuclear Trans New Guinea	Papunesia	A	A
ory	Odia	Indo-European	Eurasia	Other	Other
oss	Iron Ossetian	Indo-European	Eurasia	A	A
ote	Mezquital Otomi	Otomanguean	North America	E	ExA
otm	Eastern Highland Otomi	Otomanguean	North America	C	CxA
otn	Tenango Otomi	Otomanguean	North America	B	B
otq	Querétaro Otomi	Otomanguean	North America	B	B
ots	Estado de México Otomi	Otomanguean	North America	A	A
oym	Wayampi	Tupian	South America	A	A
ozm	Koonzime	Atlantic-Congo	Africa	A	A
pab	Parecís	Arawakan	South America	D	D
pad	Paumari	Arawan	South America	A	A
pag	Pangasinan	Austronesian	Papunesia	C	C
pah	Tenharim-Parintintin-Diahoi	Tupian	South America	A	A
pam	Pampanga	Austronesian	Papunesia	C	C
pan	Eastern Panjabi	Indo-European	Eurasia	E	ExA
pao	Northern Paiute	Uto-Aztecian	North America	A	A
pap	Papiamento	Indo-European	Eurasia	E	ExA
pbb	Páez	Isolate	Eurasia	A	A
pbc	Patamona	Cariban	South America	A	A
pbi	Parkwa	Afro-Asiatic	Eurasia	C	C
pbl	Mak	Atlantic-Congo	Africa	E	E

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
pck	Paite Chin	Sino-Tibetan	Eurasia	D	D
pcm	Nigerian Pidgin	Indo-European	Eurasia	A	A
pdc	Pennsylvania German	Indo-European	Eurasia	C	C
pdt	Plautdietsch	Indo-European	Eurasia	C	C
pes	Western Farsi	Indo-European	Eurasia	E	ExA
pfe	Peere	Atlantic-Congo	Africa	E	E
pib	Yine	Arawakan	South America	A	A
pio	Piapoco	Arawakan	South America	A	A
pir	Wa'ikhana	Tucanoan	South America	A	A
pis	Pijin	Indo-European	Eurasia	A	A
pjt	Pitjantjatjara	Pama-Nyungan	Australia	C	C
pkb	Pokomo	Atlantic-Congo	Africa	A	A
plg	Pilagá	Guaicuruan	South America	D	DxX
pls	San Marcos Tlalcoyalco Popoloca	Otomanguean	North America	A	A
plt	Plateau Malagasy	Austronesian	Papunesia	C	C3
plu	Palikúr	Arawakan	South America	B	B
plw	Brooke's Point Palawano	Austronesian	Papunesia	C	C
pma	Paama	Austronesian	Papunesia	A	A
pmf	Pamona	Austronesian	Papunesia	D	DxX
pmx	Poumei Naga	Sino-Tibetan	Eurasia	A	A
pne	Western Penan	Austronesian	Papunesia	A	A
poe	San Juan Atzingo Popoloca	Otomanguean	North America	A	A
poh	Poqomchi'	Mayan	North America	Other	Other
poi	Highland Popoluca	Mixe-Zoque	North America	A	A
pol	Polish	Indo-European	Eurasia	A	A
pon	Pohnpeian	Austronesian	Papunesia	A	A
por	Portuguese	Indo-European	Eurasia	A	A
pot	Potawatomi	Algic	North America	A	A
poy	Pogolo	Atlantic-Congo	Africa	A	A
ppk	Uma	Austronesian	Papunesia	C	C
ppo	Folopa	Teberan	Papunesia	E	E
pps	San Luís Temalcayuca Popoloca	Otomanguean	North America	B	B
prf	Paranan	Austronesian	Papunesia	C	C
pri	Paicî	Austronesian	Papunesia	B?C?D	B?C?D
prk	South Wa	Austroasiatic	Eurasia	A	A
prs	Dari	Indo-European	Eurasia	A	A
pse	South Barisan Malay	Austronesian	Papunesia	D	D

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
ptp	Patep	Austronesian	Papunesia	D	DxX
ptu	Bambam	Austronesian	Papunesia	D	D
pua	Western Highland Purepecha	Tarascan	North America	A	A
pwg	Gapapaiwa	Austronesian	Papunesia	B	B
pww	Pwo Northern Karen	Sino-Tibetan	Eurasia	B	B
qub	Huallaga Huánuco Quechua	Quechuan	South America	A	A
quc	K'iche'	Mayan	North America	A	A
quf	Lambayeque Quechua	Quechuan	South America	B	B
qug	Bolivar-North Chimborazo Highland Quichua	Quechuan	South America	A	A
quh	South Bolivian Quechua	Quechuan	South America	A	A
qul	North Bolivian Quechua	Quechuan	South America	A	A
qup	Southern Pastaza Quechua	Quechuan	South America	A	A
quw	Tena Lowland Quichua	Quechuan	South America	A	A
quy	Ayacucho Quechua	Quechuan	South America	A	A
quz	Cusco Quechua	Quechuan	South America	A	A
qva	Ambo-Pasco Quechua	Quechuan	South America	A	A
qvc	Cajamarca Quechua	Quechuan	South America	E	E
qve	Eastern Apurímac Quechua	Quechuan	South America	A	A
qvh	Huamalíes-Dos de Mayo Huánuco Quechua	Quechuan	South America	A	A
qvi	Imbabura Highland Quichua	Quechuan	South America	A	A
qvm	Margos- Yarowilca- Lauricocha Quechua	Quechuan	South America	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
qvn	North Junín Quechua	Quechuan	South America	B	B
qvo	Napo Lowland Quechua	Quechuan	South America	A	A
qvs	San Martín Quechua	Quechuan	South America	A	A
qvw	Huaylla Wanca Quechua	Quechuan	South America	A	A
qvz	Northern Pastaza Quichua	Quechuan	South America	A	A
qwh	Huaylas Ancash Quechua	Quechuan	South America	A	A
qxh	Panao Huánuco Quechua	Quechuan	South America	A	A
qxl	Tungurahua High- land Quichua	Quechuan	South America	A	A
qxn	Northern Conchucos Ancash Quechua	Quechuan	South America	A	A
qxo	Southern Conchucos Ancash Quechua	Quechuan	South America	A	A
qxr	Cañar-Azuay- South Chimbo- razo Highland Quichua	Quechuan	South America	A	A
rad	Rade	Austronesian	Papunesia	E	ExA
rai	Ramoaina	Austronesian	Papunesia	A	A
rar	Southern Cook Is- land Maori	Austronesian	Papunesia	D	D
ria	Riang	Sino-Tibetan	Eurasia	B	B
rim	Nyaturu	Atlantic-Congo	Africa	B	B3
rkb	Rikbaktsa	Nuclear-Macro-Je	South America	A	A
rmc	Central Romani	Indo-European	Eurasia	C	CxA
rme	Archaic Angloro- mani	Indo-European	Eurasia	A	A
rml	Baltic Romani	Indo-European	Eurasia	A	A
rmn	Balkan Romani	Indo-European	Eurasia	A	A
rmo	Sinte-Manus Ro- mani	Indo-European	Eurasia	C	C
rmq	Caló	Indo-European	Eurasia	A	A
rmy	Vlax Romani	Indo-European	Eurasia	A	A
ron	Romanian	Indo-European	Eurasia	A	A
roo	Rotokas	North Bougainville	Papunesia	D	D

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
rop	Kriol	Indo-European	Eurasia	A	A
rro	Waima	Austronesian	Papunesia	A	A
rub	Gungu	Atlantic-Congo	Africa	A	A
ruf	Luguru	Atlantic-Congo	Africa	A	A
run	Rundi	Atlantic-Congo	Africa	E	E
rup	Aromanian	Indo-European	Eurasia	A	A
rus	Russian	Indo-European	Eurasia	C	CxA
rwo	Rawa	Nuclear Trans New Guinea	Papunesia	A	A
sab	Buglere	Chibchan	North America	A	A
sag	Sango	Atlantic-Congo	Africa	E	E
sah	Sakha	Turkic	Eurasia	A	A
san	Sanskrit	Indo-European	Eurasia	B	B
sas	Sasak	Austronesian	Papunesia	D	DxX
sat	Santali	Austroasiatic	Eurasia	B	B3
sba	Ngambay	Central Sudanic	Africa	A	A
sbd	Southern Samo	Mande	Africa	A	A
sbe	Saliba	Austronesian	Papunesia	B	B
sbl	Botolan Sambal	Austronesian	Papunesia	C	C
sco	Scots	Indo-European	Eurasia	B	AxB
sda	Toraja-Sa'dan	Austronesian	Papunesia	D	D2a
seh	Sena	Atlantic-Congo	Africa	A	A
ses	Koyraboro Senni Songhai	Songhay	Africa	D	DxX
sey	Secoya	Tucanoan	South America	E	E
sgb	Mag-Anchi Ayta	Austronesian	Papunesia	C	C
sgw	Sebat Bet Gurage	Afro-Asiatic	Eurasia	A	A
sgz	Sursurunga	Austronesian	Papunesia	B	B
shi	Tachelhit	Afro-Asiatic	Eurasia	C	C3
shk	Shilluk	Nilotic	Africa	C	CxA
shn	Shan	Tai-Kadai	Eurasia	E	ExA
shp	Shipibo-Conibo	Pano-Tacanan	South America	A?B	A?B
shu	Chadian Arabic	Afro-Asiatic	Eurasia	C	CxA
sig	Paasaal	Atlantic-Congo	Africa	D	D2a
sil	Tumulung Sisaala	Atlantic-Congo	Africa	C	C
sim	Mende	Sepik	Papunesia	B	B
sin	Sinhala	Indo-European	Eurasia	A	A
sja	Epena	Chocoan	South America	A	A
sld	Sissala of Burkina Faso	Atlantic-Congo	Africa	C	C
slk	Slovak	Indo-European	Eurasia	A	A
sll	Salt-Yui	Nuclear Trans New Guinea	Papunesia	A	A
slv	Slovenian	Indo-European	Eurasia	C	CxA

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
sme	North Saami	Uralic	Eurasia	A	A
smk	Bolinao	Austronesian	Papunesia	C	C
sml	Central Sama	Austronesian	Papunesia	D	D
smo	Samoan	Austronesian	Papunesia	D	D-Other
smt	Simte	Sino-Tibetan	Eurasia	C	C3
sna	Shona	Atlantic-Congo	Africa	D	D2c
snc	Sinaugoro	Austronesian	Papunesia	A	A
snd	Sindhi	Indo-European	Eurasia	A	A
snf	Noon	Atlantic-Congo	Africa	D	D2c
snn	Siona-Tetete	Tucanoan	South America	C	C
snp	Siane	Nuclear Trans New Guinea	Papunesia	E	E
snw	Selee	Atlantic-Congo	Africa	C	C
sny	Saniyo-Hiyewe	Sepik	Papunesia	A	A
som	Somali	Afro-Asiatic	Eurasia	Other	Other
sop	Songe	Atlantic-Congo	Africa	A	A
soq	Kanasi	Dagan	Papunesia	A	A
sot	Southern Sotho	Atlantic-Congo	Africa	A	A
soy	Miyobe	Atlantic-Congo	Africa	D	DxX
spa0	Spanish	Indo-European	Eurasia	A	A
spl	Selepet	Nuclear Trans New Guinea	Papunesia	A	A
spm	Sepen	Lower Sepik- Ramu	Papunesia	A	A
spp	Supyire Senoufo	Atlantic-Congo	Africa	D	D
sps	Saposa	Austronesian	Papunesia	E	AxE
spy	Sabaot	Nilotic	Africa	B	B
sqi	Northern Tosk Al- banian	Indo-European	Eurasia	A	A
sri	Siriano	Tucanoan	South America	A	A
srm	Saramaccan	Indo-European	Eurasia	B?C?D	B?C?D
srn	Sranan Tongo	Indo-European	Eurasia	C	C
srp	Serbian Standard	Indo-European	Eurasia	A	A
srq	Sirionó	Tupian	South America	A	A
ssd	Siroi	Nuclear Trans New Guinea	Papunesia	A	A
ssg	Seimat	Austronesian	Papunesia	D	DxX
ssw	Swati	Atlantic-Congo	Africa	E	E
ssx	Samberigi	Nuclear Trans New Guinea	Papunesia	A	A
stn	Owa	Austronesian	Papunesia	A	A
stp	Southeastern Te- pehuan	Uto-Aztecan	North America	D	DxX

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
stq	Ems-Weser Frisian	Indo-European	Eurasia	C	C
sua	Sulka	Isolate	Eurasia	Other	Other
suc	Western Subanon	Austronesian	Papunesia	D	DxX
sue	Suena	Nuclear Trans New Guinea	Papunesia	A	A
suk	Sukuma	Atlantic-Congo	Africa	C	C
sun	Sundanese	Austronesian	Papunesia	B?C?D	B?C?D
sur	Mwaghavul	Afro-Asiatic	Eurasia	C	C
sus	Susu	Mande	Africa	D	DxX
suz	Sunwar	Sino-Tibetan	Eurasia	A	A
swe	Swedish	Indo-European	Eurasia	A	A
swg	Swabian	Indo-European	Eurasia	C	C
swh	Swahili	Atlantic-Congo	Africa	C	C
swk	Malawi Sena	Atlantic-Congo	Africa	B	B
swp	Suau	Austronesian	Papunesia	C	C
sxb	Suba	Atlantic-Congo	Africa	C	AxC3
sxn	Sangir	Austronesian	Papunesia	D	D
syb	Central Subanen	Austronesian	Papunesia	D	D
syc	Classical Syriac	Afro-Asiatic	Eurasia	D	D
szb	Ngalum	Nuclear Trans New Guinea	Papunesia	A	A
tab	Tabasaran	Nakh-Daghestanian	Eurasia	A	A
tac	Western Tarahumara	Uto-Aztecan	North America	A	A
taj	Eastern Tamang	Sino-Tibetan	Eurasia	B	B
tam	Tamil	Dravidian	Eurasia	C	C
taq	Tamasheq	Afro-Asiatic	Eurasia	C	C3
tar	Central Tarahumara	Uto-Aztecan	North America	A	A
tat	Tatar	Turkic	Eurasia	A	A
tav	Tatuyo	Tucanoan	South America	A	A
taw	Tai	Nuclear Trans New Guinea	Papunesia	A	A
tbc	Takia	Austronesian	Papunesia	A	A
tbg	North Tairora	Nuclear Trans New Guinea	Papunesia	A	A
tbk	Calamian Tagbanwa	Austronesian	Papunesia	D	D2a
tbl	Tboli	Austronesian	Papunesia	D	D
tbo	Tawala	Austronesian	Papunesia	A	A
tby	Tabaru	North Halmahera	Papunesia	D	DxX
tbz	Ditammari	Atlantic-Congo	Africa	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
tca	Ticuna	Ticuna-Yuri	South America	C	C
tcc	Barabaiiga-Gisamjanga	Nilotic	Africa	E	E
tcs	Torres Strait-Lockhart River Creole	Indo-European	Eurasia	A	A
tcz	Thado Chin	Sino-Tibetan	Eurasia	D	D
tdt	Tetun Dili	Austronesian	Papunesia	A	A
ted	Tepo Krumen	Kru	Africa	C	C
tee	Huehuetla Tepehua	Totonacan	North America	A	A
tel	Telugu	Dravidian	Eurasia	B	B
tem	Timne	Atlantic-Congo	Africa	C	CxA
teo	Teso	Nilotic	Africa	A	A
ter	Terena-Kinikinao-Chane	Arawakan	South America	D	D
tet	Tetum	Austronesian	Papunesia	C	C
tew	Rio Grande Tewa	Kiowa-Tanoan	North America	C	C
tfr	Teribe	Chibchan	North America	B	B
tgk	Tajik	Indo-European	Eurasia	A?E	A?E
tgl	Tagalog	Austronesian	Papunesia	D	DxX
tgo	Sudest	Austronesian	Papunesia	C	CxA
tgp	Movono	Austronesian	Papunesia	A	A
tha	Thai	Tai-Kadai	Eurasia	A	A
thk	Tharaka	Atlantic-Congo	Africa	B	B
tif	Tifal	Nuclear Trans New Guinea	Papunesia	A	A
tih	Timugon Murut	Austronesian	Papunesia	C	CxA
tik	Tikar	Atlantic-Congo	Africa	A	A
tim	Timbe	Nuclear Trans New Guinea	Papunesia	A	A
tir	Tigrinya	Afro-Asiatic	Eurasia	A	A
tiy	Tiruray	Austronesian	Papunesia	C	C3
tke	Takwane	Atlantic-Congo	Africa	A	A
tku	Upper Necaxa Totonac	Totonacan	North America	A	A
tlb	Tobelo	North Halmahera	Papunesia	C	C
tlf	Telefol	Nuclear Trans New Guinea	Papunesia	E	E
tlh	Klingon	Artificial Language	Papunesia	A	A
tlj	Talinga-Bwisi	Atlantic-Congo	Africa	A	A
tmd	Haruai	Piawi	Papunesia	E	E
tna	Tacana	Pano-Tacanan	South America	B	B

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
tnc	Tanimuca-Retuarã	Tucanoan	South America	B	B
tnk	Kwamera	Austronesian	Papunesia	A	A
tnn	North Tanna	Austronesian	Papunesia	C	CxA
tnp	Whitesands	Austronesian	Papunesia	A	A
tob	Toba	Guaicuruan	South America	C	C
toc	Coyutla Totonac	Totonacan	North America	A	A
tod	Toma	Mande	Africa	A	A
toh	Gitonga	Atlantic-Congo	Africa	B	B
toi	Tonga	Atlantic-Congo	Africa	A	A
toj	Tojolabal	Mayan	North America	A	A
ton	Tonga	Austronesian	Papunesia	D	DxX
too	Xicotepec De Juárez Totonac	Totonacan	North America	A	A
top	Papantla Totonac	Totonacan	North America	A	A
tos	Highland Totonac	Totonacan	North America	A	A
tpa	Taupota	Austronesian	Papunesia	A	A
tpi	Tok Pisin	Indo-European	Eurasia	A	A
tpm	Tampulma	Atlantic-Congo	Africa	B	B
tpp	Pisaflores Tepehua	Totonacan	North America	A	A
tpt	Tlachichilco Tepehua	Totonacan	North America	A	A
tpz	Tinputz	Austronesian	Papunesia	B	B
tqb	Tenetehara	Tupian	South America	B	B
trc	Copala Triqui	Otomanguean	North America	C	C
trn	Trinitario-Javeriano-Loretano	Arawakan	South America	A	A
tro	Tarao	Sino-Tibetan	Eurasia	C	C
trp	Kok Borok	Sino-Tibetan	Eurasia	B	B
trq	San Martín Itunyoso Triqui	Otomanguean	North America	A	A
tsg	Tausug	Austronesian	Papunesia	D	D
tsn	Tswana	Atlantic-Congo	Africa	D	DxX
tso	Tsonga	Atlantic-Congo	Africa	A	A
tsw	Salka-Tsishingini	Atlantic-Congo	Africa	C	C3
tsz	Purepecha	Tarascan	North America	A	A
ttc	Tektiteko	Mayan	North America	D	D
tte	Bwanabwana	Austronesian	Papunesia	A	A
ttq	Tawallammat Tamajaq	Afro-Asiatic	Eurasia	B	AxB
tuc	Mutu	Austronesian	Papunesia	D	DxX
tue	Tuyuca	Tucanoan	South America	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
tuf	Central Tunebo	Chibchan	North America	A	A
tui	Tupuri	Atlantic-Congo	Africa	D	DxX
tuk	Turkmen	Turkic	Eurasia	A	A
tum	Tumbuka	Atlantic-Congo	Africa	E	ExA
tuo	Tucano	Tucanoan	South America	B	B
tur	Turkish	Turkic	Eurasia	C	C
tvk	Southeast Ambrym	Austronesian	Papunesia	A	A
twi	Twi	Atlantic-Congo	Africa	D	D
twu	Termanu	Austronesian	Papunesia	C	C
txq	Tii	Austronesian	Papunesia	D	D
txu	Kayapó	Nuclear-Macro-Je	South America	A	A
tyv	Tuvinian	Turkic	Eurasia	A	A
tzh	Tzeltal	Mayan	North America	A	A
tzj	Tz'utujil	Mayan	North America	A	A
tzo	Tzotzil	Mayan	North America	A	A
ubr	Ubir	Austronesian	Papunesia	A	A
ubu	Umbu-Ungu	Nuclear Trans New Guinea	Papunesia	A	A
udu	Uduk	Koman	Africa	A	A
uig	Uighur	Turkic	Eurasia	A	A
ukr	Ukrainian	Indo-European	Eurasia	B	AxB
upv	Uripiv-Wala-Rano-Atchin	Austronesian	Papunesia	A	A
ura	Urarina	Isolate	Eurasia	E	E
urb	Urubú-Kaapor	Tupian	South America	A	A
urd	Urdu	Indo-European	Eurasia	A	A
uri	Urim	Nuclear Torricelli	Papunesia	B	B
urk	Urak Lawoi'	Austronesian	Papunesia	C	CxA
urt	Urat	Nuclear Torricelli	Papunesia	A	A
usa	Usarufa	Nuclear Trans New Guinea	Papunesia	A	A
usp	Uspanteco	Mayan	North America	A	A
uvh	Uri	Nuclear Trans New Guinea	Papunesia	A	A
uvl	Lote	Austronesian	Papunesia	B	B
uzb	Northern Uzbek	Turkic	Eurasia	A	A
vag	Vagla	Atlantic-Congo	Africa	E	E
vap	Vaiphei	Sino-Tibetan	Eurasia	E	E
var	Huarijio	Uto-Aztecan	North America	A	A
ven	Venda	Atlantic-Congo	Africa	A	A
vid	Vidunda	Atlantic-Congo	Africa	C	C
vie	Vietnamese	Austroasiatic	Eurasia	A	A
viv	Iduna	Austronesian	Papunesia	B	B

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
vmy	Ayautla Mazatec	Otomanguean	North America	E	ExA
vun	Vunjo	Atlantic-Congo	Africa	A	A
vut	Vute	Atlantic-Congo	Africa	A	A
waj	Waffa	Nuclear Trans New Guinea	Papunesia	E	E
wal	Wolaytta	Ta-Ne-Omotic	Africa	A	A
wap	Wapishana	Arawakan	South America	C	C
war	Waray	Austronesian	Papunesia	C	C
wat	Kaninuwa	Austronesian	Papunesia	A	A
way	Wayana	Cariban	South America	B	B
wbm	Zhenkang Wa	Austroasiatic	Eurasia	A	A
wbp	Warlpiri	Pama-Nyungan	Australia	A	A
wca	Yanomám	Yanomamic	South America	C	C
wed	Wedau	Austronesian	Papunesia	D	DxX
wer	Weri	Kunimaipan	Papunesia	B	B
whk	Eastern Lowland Kenyah	Austronesian	Papunesia	C	C3
wib	Southern Toussian	Atlantic-Congo	Africa	E	E
wim	Wik-Mungkan	Pama-Nyungan	Australia	A	A
wiu	Wiru	Isolate	Eurasia	A	A
wmt	Walmajarri	Pama-Nyungan	Australia	A	A
wmw	Mwani	Atlantic-Congo	Africa	A	A
wnc	Wantoat	Nuclear Trans New Guinea	Papunesia	A	A
wnu	Usan	Nuclear Trans New Guinea	Papunesia	A	A
wob	Wobe-Wè North- ern	Kru	Africa	C	C
wol	Wolof	Atlantic-Congo	Africa	C	C
wos	Hanga Hundi	Ndu	Papunesia	E	E
wrk	Garrwa	GarrwanGarawa	Australia	C	C
wrs	Waris	Border	Papunesia	B	B
wsk	Waskia	Nuclear Trans New Guinea	Papunesia	A	A
wuv	Wuvulu-Aua	Austronesian	Papunesia	D	D
wwa	Waama	Atlantic-Congo	Africa	D	DxX
xal	Oirad-Kalmyk- Darkhat	Mongolic-Khitani	Eurasia	A	A
xav	Xavánte	Nuclear-Macro-Je	South America	B	B
xbi	Kombio	Nuclear Torricelli	Papunesia	B	B
xbr	Kambera	Austronesian	Papunesia	C	C
xed	Hdi	Afro-Asiatic	Eurasia	E	E
xho	Xhosa	Atlantic-Congo	Africa	C	C

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
xla	Kamula	Kamula-Elevala	Papunesia	A	A
xcl	Classical-Middle Armenian	Indo-European	Eurasia	B	B
xnn	Northern Kankanay	Austronesian	Papunesia	D	D
xon	Konkomba	Atlantic-Congo	Africa	D	D
xpe	Liberia Kpelle	Mande	Africa	A	A
xrb	Eastern Karaboro	Atlantic-Congo	Africa	A	A
xsb	Tinà Sambal	Austronesian	Papunesia	C	C
xsi	Sio	Austronesian	Papunesia	B	B
xsm	Kasem	Atlantic-Congo	Africa	E	E
xsr	Solu-Khumbu Sherpa	Sino-Tibetan	Eurasia	A	A
xsu	Sanumá	Yanomamic	South America	B	B
xtd	Diuxi-Tilantongo Mixtec	Otomanguean	North America	D	DxX
xtm	Magdalena Peñasco Mixtec	Otomanguean	North America	C	C
xtn	Northern Tlaxiaco Mixtec	Otomanguean	North America	A	A
xuo	Kuo	Atlantic-Congo	Africa	A	A
yaa	Yaminahua	Pano-Tacanan	South America	A	A
yad	Yagua	Peba-Yagua	South America	C	C
yal	Yalunka	Mande	Africa	B	BxA
yam	Yamba	Atlantic-Congo	Africa	D	DxX
yan	Mayangna	Misumalpan	North America	B	B
yao	Yao	Atlantic-Congo	Africa	D	D
yaq	Yaqui	Uto-Aztecan	North America	E	E
yby	Yaweyuha	Nuclear Trans New Guinea	Papunesia	A	A
ycn	Yucuna	Arawakan	South America	A	A
yim	Yimchungru Naga	Sino-Tibetan	Eurasia	A	A
yka	Yakan	Austronesian	Papunesia	C	C
ykg	Northern Yukaghir	Yukaghir	Eurasia	A	A
yle	Yele	Isolate	Eurasia	A	A
yli	Angguruk Yali	Nuclear Trans New Guinea	Papunesia	A	A
yml	Iamalele	Austronesian	Papunesia	A	A
yom	Yombe	Atlantic-Congo	Africa	B	AxB
yon	Yonggom	Nuclear Trans New Guinea	Papunesia	E	E
yor	Yoruba	Atlantic-Congo	Africa	A	A

Continued on next page

Table B.3 – Continued from previous page

ISO	name	family	macroarea	pattern	subpattern
yrb	Yareba	Yareban	Papunesia	A	A
yre	Yaouré	Mande	Africa	D	D-Other
yrk	Tundra Nenets	Uralic	Eurasia	A	A
yss	Yessan-Mayo	Sepik	Papunesia	C	C
yua	Yucatec Maya	Mayan	North America	C	C
yuj	Karkar-Yuri	Pauwasi	Papunesia	E	E
yut	Yopno	Nuclear Trans New Guinea	Papunesia	B	B
yuw	Yau-Nungon	Nuclear Trans New Guinea	Papunesia	A	A
yuz	Yuracaré	Isolate	Eurasia	A	A
yva	Yawa	Yawa-Saweru	Papunesia	A	A
zaa	Sierra de Juárez Zapotec	Otomanguean	North America	D	D
zab	Western Tlacolula Valley Zapotec	Otomanguean	North America	A	A
zac	Ocotlán Zapotec	Otomanguean	North America	A	A
zad	Cajonos Zapotec	Otomanguean	North America	D	DxX
zae	Yareni Zapotec	Otomanguean	North America	C	C
zai	Isthmus Zapotec	Otomanguean	North America	E	E
zam	Cuixtla-Xitla Za- potec	Otomanguean	North America	A	A
zao	Ozolotepec Zapotec	Otomanguean	North America	B	BxA
zar	Rincón Zapotec	Otomanguean	North America	A	A
zas	Santo Domingo Albarradas Zapotec	Otomanguean	North America	A	A
zat	Tabaa Zapotec	Otomanguean	North America	A	A
zav	Yatzachi Zapotec	Otomanguean	North America	A	A
zaw	Mitla Zapotec	Otomanguean	North America	A	A
zca	Coatecas Altas Zapotec	Otomanguean	North America	A	A
zho	Mandarin Chinese	Sino-Tibetan	Eurasia	E	E
zia	Zia	Nuclear Trans New Guinea	Papunesia	A	A
ziw	Zigula- Mushungulu	Atlantic-Congo	Africa	B	B3
zom	Zou	Sino-Tibetan	Eurasia	E	E
zos	Francisco León Zoque	Mixe-Zoque	North America	A	A
zpc	Choapan Zapotec	Otomanguean	North America	A	A

Continued on next page

Table B.3 – *Continued from previous page*

ISO	name	family	macroarea	pattern	subpattern
zpi	Santa María Quiegolani Zapotec	Otomanguean	North America	A	A
zpl	Lachixío Zapotec	Otomanguean	North America	A	A
zpm	Mixtepec Zapotec	Otomanguean	North America	A	A
zpo	Amatlán Zapotec	Otomanguean	North America	E	ExA
zpq	Zoogocho Zapotec	Otomanguean	North America	A	A
zpt	San Vicente Coatlán Zapotec	Otomanguean	North America	E	E
zpu	Yalálag Zapotec	Otomanguean	North America	C	C3
zpv	Chichicapan Zapotec	Otomanguean	North America	A	A
zpz	Texmelucan Zapotec	Otomanguean	North America	E	ExA
zsm	Standard Malay	Austronesian	Papunesia	A?E	A?E
zsr	Southern Rincon Zapotec	Otomanguean	North America	A	A
ztq	Quioquitani- Quieri Zapotec	Otomanguean	North America	E	E
zty	Yatee Zapotec	Otomanguean	North America	E	E
zul	Zulu	Atlantic-Congo	Africa	C	CxA
zyb	Yongbei Zhuang	Tai-Kadai	Eurasia	B	B
zyp	Zyphe	Sino-Tibetan	Eurasia	E	E

Table B.3: Classification of the languages in the massively parallel dataset into patterns and subpatterns.

Bibliography

- Aikhenvald, Alexandra & Pieter Muysken. 2010. *Multi-verb Constructions: A View from the Americas*. Leiden: Brill.
- Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6). 716–723. doi:10.1109/TAC.1974.1100705.
- Alekseev, Anatolij A. 1987. Participium activi v russoj letopisi: osobennosti funkcionirovanija. *Russian Linguistics* 59(11). 187–200.
- Altshuler, Daniel & Robert Truswell. 2022. Discourse structure, information structure, coordination, and extraction. In *Coordination and the Syntax – Discourse Interface*, Oxford University Press. doi:10.1093/oso/9780198804239.003.0006.
- AnderBois, Scott & Daniel Altshuler. 2022. Coordination, coherence and A’ingae clause linkage. *Proceedings of SALT* 32. 793–813.
- AnderBois, Scott, Daniel Altshuler & Wilson D. L. Silva. 2023. The Forms and Functions of Switch Reference in A’ingae. *Languages* 8(2). doi:10.3390/languages8020137.
- Andersen, Henning. 1969. The dative of subordination in Baltic and Slavic. In Thomas F. Magner & William R. Schmalstieg (eds.), *Baltic linguistics*, 1–9. Pennsylvania-London: Pennsylvania State University Press, University Park.
- Asher, Nicholas. 2004. Discourse topic. *Theoretical Linguistics* 30(2-3). 163–201.
- Asher, Nicholas & Alex Lascarides. 1998. The Semantics and Pragmatics of Presupposition. *Journal of Semantics* 15. 239–299.
- Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge: Cambridge University Press.
- Asher, Nicholas, Laurent Prévot & Laure Vieu. 2007. Setting the background in discourse. *Discours* 1. 1–29.
- Asher, Nicholas & Laure Vieu. 2005. Subordinating and Coordinating Discourse Relations. *Lingua* 115. 591–610. doi:10.1016/j.lingua.2003.09.017.
- Awoyale, Yiwola. 1987. Perspectives on verb serialization. *Niger-Congo Syntax and Semantics* 1. 3–36.
- Bah, Oumar. 1980. Dictionnaire Pular-Français, Avec un index français-pular. Ms.
- Bamgboṣe, Ayo. 1974. On serial verb constructions and verbal status. *Journal of West African Languages* 9. 17–48.
- Bary, Corien & Dag T.T. Haug. 2011. Temporal anaphora across and inside sentences: The function of participles. *Semantics and Pragmatics* 4. 1–56.

- Bauer, Brigitte. 2000. *Archaic Syntax in Indo-European: The Spread of Transitivity in Latin and French*. Berlin, New York: Mouton De Gruyter.
- Behrens, Bergljot & Cathrine Fabricius-Hansen. 2009. The Relation Accompanying Circumstance Across Languages: Conflict Between Linguistic Expression and Discourse Subordination? In Dingfang Shu & Ken Turner (eds.), *Contrasting meanings in languages of the East and West*, Peter Lang.
- Behrens, Bergljot, Cathrine Fabricius-Hansen & Kåre Solfeld. 2012. Competing structures: The discourse perspective. In Cathrine Fabricius-Hansen & Dag T.T. Haug (eds.), *Big events, small clauses*, 179–225. Berlin-Boston: De Gruyter.
- Bennett, David C. 1975. *Spatial and temporal uses of English prepositions*. London: Longman.
- Berent, G. P. 1973. Absolute Constructions as ‘Subordinate Clauses’. In Thomas C. Smith-Stark Ann Weiser, Claudia W. Corum (ed.), *You Take the High Node and I’ll Take the Low Node: Papers from the comparative syntax festival ‘The Differences between main and subordinate clauses’*, 147–154. Chicago: Chicago Linguistic Society.
- Bickel, Balthasar. 1991. *Typologische Grundlagen der Satzverkettung: Ein Beitrag zur Allgemeinen Grammatik der Satzverbindung und des Fährtenlegens*. Zürich: ASAS-Verlag.
- Bickel, Balthasar. 1999. From ergativus absolutus to topic marking in Kiranti: a typological perspective. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 25. 38–49.
- Birnbaum, Henrik. 1958. Zur Aussonderung der syntaktischen Gräzismen im Altkirchenslavischen. Einige methodische Bemerkungen. *Scando-Slavica* 4. 239–57.
- Bisang, Walter. 1995a. Verb serialization and converbs. Differences and similarities. In Martin Haspelmath & Ekkehard König (eds.), *The new psychology of language*, 137–188. Berlin-New York: Mouton de Gruyter.
- Bisang, Walter. 1995b. Verb serialization and converbs – differences and similarities. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective*, 137–188. Berlin, Boston: De Gruyter Mouton. doi:doi:10.1515/9783110884463-006.
- Blakemore, Diane & Robyn Carston. 2005. The pragmatics of sentential conjunction with *and*. *Lingua* 115. 569–589.
- Bonomi, Andrea. 1997. Aspect, quantification and when-clauses in Italian. *Linguistics and Philosophy* 20. 469–514.
- Borkovskij, Viktor I. & Petr S. Kuznecov. 1963. *Istoricheskaja grammatika russkogo jazyka*. Moscow: Moskovskij Univ.
- Bozzone, Chiara. 2014. *Constructions: A New Approach to Formularity, Discourse, and Syntax in Homer*. Los Angeles, CA: UCLA dissertation.
- Bras, Myriam, Anne Le Draoulec & Laure Vieu. 2001. Temporal Information And Discourse Relations In Narratives: The Role Of French Connectives ‘Puis’ And ‘Un Peu Plus Tard’. In *Proceedings of the ACL 2001 workshop on temporal and spatial information processing*, <https://aclanthology.org/W01-1307>.
- Brinton, Laurel J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin-New York: De Gruyter.
- Broadwell, George Aaron. 1997. Binding Theory and Switch Reference. In P. Pica & J. Rooryck (eds.), *Atomism and binding*, 31–49. Foris.

- Bybee, Joan L. & Östen Dahl. 1989. The Creation of Tense and Aspect Systems in the Languages of the World. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 13(1). 51–103. doi:10.1075/sl.13.1.03byb.
- Büring, Daniel. 2006. Focus projection and default prominence. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 321–346. Berlin, Boston: De Gruyter Mouton.
- Carlson, Lori & Daniel Marcu. 2001. Discourse Tagging Manual. ISI Tech Report ISI-TR-545 Information Sciences Institute, University of Southern California.
- Caudal, Patrick. 2005. Stage structure and stage salience for event semantics. In Paula Kempchinsky & Roumyana Slabakova (eds.), *Aspectual Inquiries*, 239–264. Dordrecht: Springer.
- Chafe, Wallace. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and Topic*, 27–55. New York: Academic Press.
- Charolles, Michel. 2005. Framing adverbials and their role in discourse cohesion from connexion to forward labelling. In Michel Aurnague, Myriame Bras, Anne Le Draoulec & Laure Vieu (eds.), *Sem-05 proceedings*, 13–30. Biarritz.
- Coelho, Sofia Vaz Pinto Simoes. 2020. *The Slavonic Life of Mary of Egypt in the Bđinski Sbornik. Introduction and Translation*. Oxford: University of Oxford MPhil Dissertation.
- Collins, Daniel E. 2004. Distance, subjecthood, and the early Slavic dative absolute. *Scando-Slavica* 50. 165–181.
- Collins, Daniel E. 2011. The pragmatics of “unruly” dative absolutes in Early Slavic. *Oslo Studies in Language* 3(3). 103–130.
- Comrie, Bernard. 1976. *Aspect*. London: Cambridge University Press.
- Comrie, Bernard. 1983. Switch-reference in Huichol: A typological study. In John Haiman & Pamela Munro (eds.), *Switch reference and universal grammar*, 17–38. Amsterdam/Philadelphia: John Benjamins.
- Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
- Comrie, Bernard. 2003. When agreement gets trigger-happy. *Transactions of the Philological Society* 101(2). 313–337.
- Conrad, Robert J. & Kepas Wogiga. 1991. *An Outline of Bukiyip Grammar*. Canberra: Pacific Linguistics.
- Corin, Andrew R. 1995. The dative absolute in Old Church Slavonic and Old East Slavic. *Die Welt der Slaven* 40. 251–284.
- Covington, Michael A. & Joe D. McFall. 2010. Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics* 17(2). 94–100. doi:10.1080/09296171003643098.
- Cristofaro, Sonia. 1998. Deranking and Balancing in Different Subordination Relations: a Typological Study. *Sprachtypologie und Universalienforschung* 51. 3–42.
- Cristofaro, Sonia. 2013a. Purpose Clauses. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/125>.

- Cristofaro, Sonia. 2013b. 'When' clauses. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/126>.
- Cristofaro, Sonia. 2019. *Subordination*. Oxford: Oxford University Press.
- Croft, William. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.
- Croft, William. 2003. Exemplar Semantics. <https://www.unm.edu/%7Ewcroft/WACpubs.html>.
- Croft, William & Keith T. Poole. 2008. Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34(1). 1–37. <https://doi.org/10.1515/THLI.2008.001>.
- Dahl, Östen & Bernhard Wälchli. 2016. Perfects and iamitives: two gram types in one grammatical space. *Letras de Hoje* 325–348. doi:10.15448/1984-7726.2016.3.25454.
- Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press Oxford.
- Declerck, Renaat H. C. 1997. *When-Clauses and Temporal Structure*. London: Routledge.
- Di Domenico, Elisa, Ioli Baroncini & Andrea Capotorti. 2020. Null and overt subject pronouns in topic continuity and topic shift. *Glossa* 5(1). 117.
- Dimitrova, Tsvetana. 2021. On the Diachrony of the Clitic Cluster in Bulgarian. *Balkanica et Slavia* 1(1). 67–95.
- Dooley, Robert A. 2010. Exploring Clause Chaining. *SIL Electronic Working Papers in Linguistics* .
- Dostál, Antonin (ed.). 1954. *Studie o vidovém systému v staroslověnině*. Prague: Státní pedagogické nakladatelství.
- Dowty, David R. 1972. *Studies in the Logic of Verb Aspect and Time Reference in English*: University of Texas dissertation.
- Dowty, David R. 1977. Toward a Semantic Analysis of Verb Aspect and the English 'Imperfective' Progressive. *Linguistics and Philosophy* 1(1). 45–77.
- Dowty, David R. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Dowty, David R. 1986. The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics. *Linguistics and Philosophy* 9(1). 37–62.
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info>.
- Dujčev, Ivan. 1972. *Bdinski zbornik. Ghent Slavonic Ms 408, A.D. 1360*. London: Variorum Reprints facsimile edn.
- Dyer, Chris, Victor Chahuneau & Noah A. Smith. 2013. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In *Proceedings of the 2013 conference of the north American chapter of the association for computational linguistics: Human language technologies*, 644–648. Atlanta, Georgia: Association for Computational Linguistics. <https://aclanthology.org/N13-1073>.
- Eckhoff, Hanne, Marek Majer, Eirik Welo & Dag Haug. 2009. Breaking down and putting back together: analysis and synthesis of New Testament Greek. *Journal of Greek Linguistics* 9(1). 56–92. doi:10.1163/156658409X12529372103308.

- Eckhoff, Hanne M. 2018. Quantifying syntactic influence: Word order, possession and definiteness in Old Church Slavonic and Greek. In Björn Hansen, Jasmina Grković-Major & Barbara Sonnenhauser (eds.), *Diachronic Slavonic syntax: The interplay between internal development, language contact and metalinguistic factors*, 29–62. Berlin-Boston: De Gruyter.
- Eckhoff, Hanne M. & Aleksandrs Berdičevskis. 2015. Linguistics vs. digital editions: The Tromsø Old Russian and OCS treebank. *Scripta & E-Scripta* 14-15. 9–25.
- Eckhoff, Hanne M. & Dag T.T. Haug. 2015. Aspect and prefixation in Old Church Slavonic. *Diachronica* 32(2).
- Eckhoff, Hanne M. & Laura A. Janda. 2014. Grammatical Profiles and Aspect in Old Church Slavonic. *Transactions of the Philological Society* 112(2).
- Evans, Barrie. 2017. A teaching grammar of Pular. Ms.
- Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 366–431. Oxford: Oxford University Press.
- Evans, Nicholas & Honoré Watanabe. 2016. The dynamics of insubordination: An overview. In Nicholas Evans & Honoré Watanabe (eds.), *Insubordination*, 1–38. Amsterdam/Philadelphia: John Benjamins.
- Fabricius-Hansen, Cathrine. 2007. On Manners and Circumstances. In Randi A. Nilsen, Nana A. A. Amfo & Kaja Borthen (eds.), *Interpreting Utterances: Pragmatics and its Interfaces*, 39–50. Oslo: Novus Press.
- Fabricius-Hansen, Cathrine & Dag T.T. Haug (eds.). 2012a. *Big events and small clauses*. Berlin-Boston: De Gruyter.
- Fabricius-Hansen, Cathrine & Dag T.T. Haug. 2012b. Co-eventive adjuncts: Main issues and clarifications. In Cathrine Fabricius-Hansen & Dag T.T. Haug (eds.), *Big events and small clauses*, 21–54. Berlin-Boston: De Gruyter.
- Fabricius-Hansen, Cathrine & Kjell Johan Sæbø. 1983. Das Chamäleon “wenn” und seine Umwelt. *Linguistische Berichte* 1–35.
- Finer, Daniel. 1984. *The Formal Grammar of Switch-reference*: University of Massachusetts, Amherst MA dissertation.
- Finer, Daniel. 1985. The syntax of switch-reference. *Linguistic Inquiry* 16. 35–55.
- Fischer, Rafael & Eva van Lier. 2011. Cofán subordinate clauses in a typology of subordination. In Rik van Gijn, Katharina Haude & Pieter Muysken (eds.), *Subordination in Native South American Languages*, vol. 2, 221–250. Amsterdam/Philadelphia: John Benjamins.
- Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Frascarelli, Mara. 2007. Subjects, topics and the interpretation of referential *pro*. *Natural Language & Linguistic Theory* 25. 691–734.
- Fuller, Lois K. 2006. The “Genitive Absolute” in New Testament/Hellenistic Greek: A Proposal for Clearer Understanding. *JGRChJ* 3. 142–167.
- Féry, Caroline. 2007. The prosody of topicalization. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form: Generalizations across languages*, 69–86. Amsterdam: John Benjamins.

- Féry, Caroline. 2008. Information structural notions and the fallacy of invariant correlates. *Acta Linguistica Hungarica* 55(3-4). 361–379.
- Gebert, Lucyna. 1987. Les constructions absolues en vieux russe. *Revue des études slaves* 59(3). 565–570.
- van Gijn, Rik. 2012. Switch-attention (aka switch-reference) in South-American temporal clauses. *Linguistic Discovery* 10(1). 112–27.
- van Gijn, Rik. 2016a. Switch reference: An overview. In Rik van Gijn & Jeremy Hammond (eds.), *Switch reference 2.0*, 1–54. Amsterdam/Philadelphia: John Benjamins.
- van Gijn, Rik. 2016b. Switch reference in Western South America. In Rik van Gijn & Jeremy Hammond (eds.), *Switch reference 2.0*, 153–206. Amsterdam/Philadelphia: John Benjamins.
- van Gijn, Rik, Katharina Haude & Pieter Muysken (eds.). 2011. *Subordination in native South American languages*. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy. 2001. *Syntax: a functional-typological introduction, Volume II*. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy. 2015. Chapter 7. Serial verbs and syntactic change: Niger-Congo. In *The diachrony of grammar*, 131–162. Amsterdam-Philadelphia: John Benjamins.
- Godley, A. D. (ed.). 1920. *Herodotus. The Persian Wars, Volume I: Books 1-2* (Loeb Classical Library 117). Cambridge, MA: Harvard University Press.
- Goldberg, Adele. 2005. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199268511.001.0001.
- Good, Jeff & Michael Cysouw. 2013. Languoid, Doculect and Glossonym: Formalizing the Notion 'Language'. *Language Documentation & Conservation* 7. 331–359.
- Goodwin, William W. 1997 [1930]. *A treatise on the grammar of New Testament Greek*. London: Bristol Classical.
- Gordon, Lynn. 1983. Switch-reference, clause order, and interclausal relationships in Maricopa. In John Haiman & Pamela Munro (eds.), *Switch-reference and universal grammar*, 83–104. Amsterdam: John Benjamins.
- Grenoble, Lenore. 2004. Theticals in Russian. *Journal of Pragmatics* 36(11). 1953–1974. doi:10.1016/j.pragma.2004.02.008.
- Grimes, Joseph E. 1975. *The thread of discourse*. The Hague: Mouton.
- Grković-Major, Jasmina. 2007. *Spisi iz istorijske lingvistike*. Novi Sad: Zoran Stojanovi.
- Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14(1). 39–100.
- Guérin, Valérie & Grant Aiton. 2019. Bridging constructions in typological perspective. In Valérie Guérin (ed.), *Bridging constructions*, 1–44. Berlin: Language Science Press.
- Haiman, John & Pamela Munro. 1983. Introduction. In John Haiman & Pamela Munro (eds.), *Switch reference and universal grammar*, ix–xv. Amsterdam: John Benjamins.
- Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244.

- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2023. *Glottolog* 4.8. Leipzig: Max Planck Institute for Evolutionary Anthropology. doi:10.5281/zenodo.8131084.
- Hansen, Maj-Britt Mosegaard. 1998. The semantic status of discourse markers. *Lingua* 104(3). 235–260. doi:[https://doi.org/10.1016/S0024-3841\(98\)00003-5](https://doi.org/10.1016/S0024-3841(98)00003-5).
- Hardy, Heather. 1982. Pragmatics and the syntax of switch reference in Tolkapaya. *South-western Journal of Linguistics* 5. 85–99.
- Hartmann, Iren, Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3). 463–484.
- Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective*, 1–56. Berlin, Boston: De Gruyter Mouton.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language*, vol. 2, 211–242. Mahwah, NJ: Lawrence Erlbaum.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. <http://www.jstor.org/stable/40961695>.
- Haspelmath, Martin & Ekkehard König. 1995. *Converbs in Cross-Linguistic Perspective. Structure and Meaning of Adverbial Verb Forms – Adverbial Participles, Gerunds –*. Berlin-New York: Mouton de Gruyter.
- Haug, Dag & Nilo Pedrazzini. 2023. The semantic map of *when* and its typological parallels. *Frontiers in Communication* 8. doi:10.3389/fcomm.2023.1163431.
- Haug, Dag T., Marius L. Jøhndal, Hanne M. Eckhoff, Eirik Welo, Mari J. B. Hertenberg & Angelika Müth. 2009. Computational and Linguistic Issues in Designing a Syntactically Annotated Parallel Corpus of Indo-European Languages. In *Traitement automatique des langues, volume 50, numéro 2 : Langues anciennes [ancient languages]*, 17–45. France: ATALA (Association pour le Traitement Automatique des Langues). <https://aclanthology.org/2009.tal-2.2>.
- Haug, Dag T. T. & Marius L. Jøhndal. 2008. Creating a Parallel Treebank of the Old Indo-European Bible Translations. In Caroline Sporleder & Kiril Ribarov (eds.), *Proceedings of the second workshop on language technology for cultural heritage data (latech 2008)*, 27–34. http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31.
- Haug, Dag T.T. 2011. Backward Control in Ancient Greek – Subsumption or Linearization? In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, 278–298. Stanford: CSLI Publications.
- Haug, Dag T.T. 2012. Open verb-based adjuncts in New Testament Greek and the Latin of the Vulgate. In Cathrine Fabricius-Hansen & Dag T.T. Haug (eds.), *Big events and small clauses*, 287–321. Berlin-Boston: De Gruyter.
- Haug, Dag T.T. 2017. Backward Control in Ancient Greek and Latin Participial Adjuncts. *Natural Language & Linguistic Theory* 35(1). 99–159.
- Heine, Bernd, Gunther Kaltenböck & Tania Kuteva. 2016. On insubordination and cooptation. In Nicholas Evans & Honoré Watanabe (eds.), *Insubordination*, 39–64. Amsterdam/Philadelphia: John Benjamins.

- Heinämäki, Orvokki. 1972. *Semantics of English temporal connectives*: University of Texas, Austin dissertation.
- Heusinger, Klaus von & Petra B. Schumacher. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154. 117–127.
- Himmelmann, Nikolaus & Beatrice Primus. 2015. Prominence beyond prosody - a first approximation. In Amedeo De Dominicis (ed.), *Prominence in linguistics. proceedings of the ps-prominences international conference*, 38–58. University of Tuscia: DISUCOM Press.
- Hinrichs, Erhard. 1986. Temporal Anaphora in Discourses of English. *Linguistics and Philosophy* 9(1). 63–82.
- Hobbs, Jerry R., Mark Stickel, Paul Martin & Douglas Appel. 1993. Interpretation as abduction. *Artificial Intelligence* 63(1–2). 69–142.
- Hogeterp, Albert & Adelbert Denaux. 2018. *Semitisms in Luke's Greek: A descriptive analysis of lexical and syntactical domains of Semitic language influence in Luke's Gospel*. Tübingen: Mohr Siebeck.
- Hopper, Paul J. 1979 [2020]. Aspect and foregrounding in discourse. In Talmy Givón (ed.), *Discourse and syntax* Syntax and Semantics (12), 211–241. Leiden: Brill.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56(2). 251–299.
- Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press 2nd edn.
- Isard, S. D. 1974. What would you have done if ... ? *Theoretical Linguistics* 1(1-3). 233–256.
- Janse, Mark. 2007. The Greek of the New Testament. In Anastassios-Fivos Christidis (ed.), *A history of Ancient Greek: From the beginnings to Late Antiquity*, 646–653. Cambridge: Cambridge University Press.
- Junczys-Downmunt, Marcin & Arkadiusz Szal. 2012. SyMGiza++: Symmetrized Word Alignment Models for Machine Translation. In Pascal Bouvry, Mieczysław A. Kłopotek, Franck Leprévost, Malgorzata Marciniak, Agnieszka Mykowiecka & Henryk Rybinski (eds.), *Security and intelligent information systems (siis)*, vol. 7053 Lecture Notes in Computer Science, 379–390. Warsaw, Poland: Springer.
<http://emjotde.github.io/publications/pdf/mjd2011siis.pdf>.
- Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory*. Dordrecht: Springer.
- Kamp, Hans, Josef Genabith & Uwe Reyle. 2011. Discourse Representation Theory. In Dov M. Gabbay & Franz Günthner (eds.), *Handbook of Philosophical Logic*, vol. 15, 125–394. Dordrecht: Springer. <https://doi.org/10.1007/978-94-009-7066-3>.
- Kamp, Hans & Christian Rohrer. 1983. Tense in Texts. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use and interpretation of language*, 250–269. Berlin-Boston: Mouton de Gruyter.
- Kamphuis, Jaap. 2020. *Verbal aspect in Old Church Slavonic. A corpus-based approach*. Leiden: Brill-Rodopi.
- Kaplan, David. 1989. Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from kaplan*, 481–563. Oxford University Press.

- Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.
- Keine, Stephan. 2013. Deconstructing switch-reference. *Natural Language and Linguistic Theory* 31. 767–826.
- Kortmann, Bernd. 1991. *Free adjuncts and absolutes in English: Problems of control and interpretation*. London-New York: Routledge.
- Krifka, Manfred. 2007. Basic notions of information structure. *Interdisciplinary Studies on Information Structure* 6. 13–56.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4). 243–276. doi:10.1556/aling.55.2008.3-4.2.
- Kurešević, Marina. 2006. Apsolutni dativ u srpskoj srednjovekovnoj pismenosti. *Zbornik Matice srpske za filologiju i lingvistiku* 49(1). 35–112.
- König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London-New York: Routledge.
- König, Ekkehard. 1995. The meaning of converb constructions. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective*, 57–96. Berlin, Boston: De Gruyter Mouton. doi:doi:10.1515/9783110884463-004.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Langdon, Margaret & Pamela Munro. 1979. Subject and (switch-)reference in Yuman. *Folia Linguistica* 13. 321–344.
- Levinsohn, Stephen H. 2000. *Discourse features of New Testament Greek: A coursebook on the information structure of New Testament Greek*. Dallas: Summer Institute of Linguistics.
- Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: The MIT Press. doi:10.7551/mitpress/5526.001.0001.
- Levshina, Natalia. 2015. European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2). 487–520. doi:doi:10.1515/flin-2015-0017.
- Levshina, Natalia. 2019. Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology* 23(3). 533–572.
- Levshina, Natalia. 2022. Corpus-based typology: applications, challenges and some solutions. *Linguistic Typology* 26(1). 129–160. <https://doi.org/10.1515/lingty-2020-0118>.
- Lindberg, Rebecka. 2013. *Definiteness in Old Church Slavonic: A Study of How Long and Short Form in Adjectives Reflect Information Status*. University of Oslo MA thesis. <https://www.duo.uio.no/handle/>.
- Longacre, Robert E. 2007. Sentences as combinations of clauses. In Timothy Shopen (ed.), *Language typology and syntactic description*, 372–420. Cambridge University Press.
- Lord, Carol. 1973. Serial verbs in transition. *Studies in African Linguistics* (4). 269–295.
- MacRobert, Mary C. 1986. Foreign, naturalized and native syntax in Old Church Slavonic. *Transactions of the Philological Society* 84(1). 142–166.

- Maschler, Yael & Deborah Schiffrin. 2015. Discourse markers: Language, meaning, and context. In Deborah Tannen, Heidi E. Hamilton & Deborah Schiffrin (eds.), *The Handbook of Discourse Analysis*, 189–221. Oxford: Wiley-Blackwell 2nd edn.
- Matisoff, James A. 2020. *The Dictionary of Lahu*. Berkeley: University of California Press.
- Mauri, Caterina & Anna Giacalone Ramat. 2012. The development of adversative connectives in Italian: Stages and factors at play. *Linguistics* 50(2). 191–239.
- Mayer, Thomas & Michael Cysouw. 2014. Creating a massively parallel Bible corpus. In *Proceedings of the ninth international conference on language resources and evaluation (LREC'14)*, 3158–3163. Reykjavik, Iceland: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/220_Paper.pdf.
- McGillivray, Barbara, Malvina Nissim, Nilo Pedrazzini, Saskia Peels-Matthey & Silvia Stopponi. Forthcoming. Evaluating Language Models for Ancient Greek: Design, Challenges, and Future Directions. *Diachronica* (Special issue “Demystifying new methods in historical linguistics”).
- McKenzie, Andrew. 2012. *The role of contextual restriction in reference-tracking*: Amherst: University of Massachusetts dissertation.
- McKenzie, Andrew. 2015a. A survey of switch-reference in North America. *International Journal of American Linguistics* 81. 409–448.
- McKenzie, Andrew. 2015b. Austinian situations and switch-reference: The role of context in reference-tracking. Ms.
- McLendon, Sally. 1978. Ergativity, case and transitivity in Eastern Pomo. *International Journal of American Linguistics* 44. 1–113.
- Mithun, Marianne. 1993. “Switch-reference”: Clause combining in Central Pomo. *International Journal of American Linguistics* 59(2). 119–136.
- Mithun, Marianne. 2008. The Extension of Dependency Beyond the Sentence. *Language* 84(1). 69–119.
- Moens, Marc. 1987. *Tense, aspect and temporal reference*: University of Edinburgh dissertation.
- Moens, Marc & Mark J. Steedman. 1987. Temporal Ontology in Natural Language. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics* 1–7.
- Morozov, Timofej S. (ed.). 1880. *Izbornik velikoga knjazja Svjatoslava Jaroslavoviča 1073 goda*. Peterburg: Obščestvo ljubitelej drevnej pis'mennosti.
- Musi, Elena. 2016. Semantic change from space-time to contrast: The case of Italian adversative connectives. *Folia Linguistica* 50(1). 1–30. doi:doi:10.1515/flin-2016-0001.
- Nahtigal, Rajko (ed.). 1942. *Euchologium Sinaiticum: Starocerkvenoslovanski glagolski spomenik*. Akademija Znanosti in Umetnosti.
- Nerbonne, John. 1986. Reference time and time in narration. *Linguistics and Philosophy* 9. 83–95.
- Nečásek, Ladislav. 1957. Staroslověnské dativní vazby participální a jejich předlohy v řeckém textu evangelíí. *Slavia* (26). 13–30.
- Nonato, Rafael. 2014. *Clause-chaining, switch-reference and coordination*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

- Nordhoff, Sebastian & Harald Hammarström. 2011. Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In *Proceedings of iswc 2011*, .
- Nychka, Douglas, Reinhard Furrer, John Paige & Stephan Sain. 2021. fields: Tools for spatial data. R package version 14.1. <https://github.com/dnychka/fieldsRPackage>.
- Och, Franz Josef & Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics* 29(1). 19–51.
- Omohundro, Stephen M. 1989. *Five balltree construction algorithms*. International Computer Science Institute Berkeley.
- Overall, Simon E. 2014. Clause-chaining, switch-reference and nominalisations in Aguaruna (Jivaroan). In Rik van Gijn, Jeremy Hammond, Dejan Matić, Saskia van Putten & Ana Vilacy Galucio (eds.), *Information structure and reference tracking in complex sentences*, 309–340. Amsterdam/Philadelphia: John Benjamins.
- Overall, Simon E. 2016. Switch-reference and case-marking in Aguaruna (Jivaroan) and beyond. In Rik van Gijn & Jeremy Hammond (eds.), *Switch reference 2.0*, 453–472. Amsterdam/Philadelphia: John Benjamins.
- Partee, Barbara H. 1984. Nominal and temporal anaphora. *Linguistics and Philosophy* 7. 243–286.
- Passonneau, Rebecca J. 1987. Situations and Intervals. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics* 16–24.
- Pedrazzini, Nilo. 2022. One question, different annotation depths: A case study in Early Slavic. *Journal of Historical Syntax* 6(4-11). 1–40. doi:10.18148/hs/2022.v6i4-11.96.
- Pedrazzini, Nilo. 2023. Early Slavic language models. doi:10.5281/zenodo.8414137. <https://doi.org/10.5281/zenodo.8414137>.
- Pedrazzini, Nilo & Hanne Martine Eckhoff. 2021. OldSlavNet: A scalable Early Slavic dependency parser trained on modern language data. *Software Impacts* 100063.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot & E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12. 2825–2830.
- Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). 149–188.
- Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi & Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the sixth international conference on language resources and evaluation (lrec'08)*, Marrakech, Morocco: European Language Resources Association (ELRA).
- Prévot, Laurent. 2004. *Structures sémantiques et pragmatiques pour la modélisation de la cohérence dans des dialogues finalisés*: Université Paul Sabatier dissertation.
- Reese, Brian, Julie Hunter, Nicholas Asher, Pascal Denis & Jason Baldridge. 2007. Reference Manual for the Analysis and Annotation of Rhetorical Structure (Version 1.0). Tech. rep. University of Texas at Austin University of Texas at Austin.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: Macmillan.

- Remneva, Marina. 1989. Konstrukcija “datel’nyj samostojatel’nyj” kak stilističeski markirovannoe sredstvo peredači vremennyx značenij. *Vestnik Moskovskogo universiteta, Serija 9: Filologija* 2(2). 33–41.
- Ritchie, Graeme D. 1979. Temporal clauses in English. *Theoretical Linguistics* 6. 89–115.
- Rizzi, Luigi. 2018. Subjects, topics and the interpretation of *pro*. In *From sounds to structures: Beyond the Veil of Maya*, 510–529. Berlin: De Gruyter Mouton.
- Roberts, John R. 1987. *Amele*. London: Croom Helm.
- Roberts, John R. 1988a. Amele switch-reference and the Theory of Grammar. *Linguistic Inquiry* 19(1). 45–63.
- Roberts, John R. 1988b. Switch-reference in Papuan languages: a syntactic or extrasyntactic device? *Australian Journal of Linguistics* 8. 75–117.
- Roberts, John R. 1997. Switch-reference in Papua New Guinea: A preliminary survey. In Andrew Pawley (ed.), *Papers in papuan linguistics*, 101–241. Canberra: Australian National University.
- Roberts, John R. 2017. A Typology of Switch Reference. In Alexandra Y. Aikhenvald & R. M. W. Editors Dixon (eds.), *The cambridge handbook of linguistic typology* Cambridge Handbooks in Language and Linguistics, 538–573. Cambridge University Press. doi:10.1017/9781316135716.017.
- Rodda, Martina Astrid. 2021. *A corpus study of formulaic variation and linguistic productivity in early Greek epic*. Oxford: University of Oxford dissertation.
- Rodda, Martina Astrid, Philomen Probert & Barbara McGillivray. 2019. Vector space models of Ancient Greek word meaning, and a case study on Homer. *Traitement Automatique des Langues* 60(3). 63–87.
- Růžička, Rudolf. 1958. Griechische Lehnsyntax im Altslawischen. *Zeitschrift für Slawistik* 3. 173–85.
- Růžička, Rudolf. 1961. Struktur und Echtheit des altslawischen dativus absolutus. *Zeitschrift für Slawistik* 6. 588–96.
- Růžička, Rudolf. 1963. *Das syntaktische System der altslawischen Partizipien und sein Verhältnis zum Griechischen*. Berlin: Akademie-Verlag.
- Sakharova, Anna. 2007. K voprosu o pragmatičeskich kriterijach raspredelenija predikacij na pričastnye i finitnye v drevnerusskoj letopisi. *Russkij jazyk v naučnom osveščeni* 85. 85–113.
- Sakharova, Anna. 2010. K voprosu o diskursivnyh funkcijach pričastnyh konstrukcij v russkoj letopisi. *Russian Linguistics* 34. 87–111.
- Sandström, Görel. 1993. *When-clauses and the Temporal Interpretation of Narrative Discourse*. Umeå: University of Umeå dissertation.
- Schaeken, Jos. 2018a. *Voices on Birchbark: Everyday Communication in Medieval Russia*. Cambridge: Boston.
- Schaeken, Jos. 2018b. *Voices on Birchbark: Everyday Communication in Medieval Russia*. Leiden, The Netherlands: Brill. doi:https://doi.org/10.1163/9789004389427.
- Scharpe, Jan L. & Frans Vyncke. 1973. *Bdinski Zbornik. An Old-Slavonic menologium of women saints. (Ghent University Library Ms. 408, A.D. 1360)*. Brugge: De Tempel.

- Scherrer, Yves & Achim Rabus. 2019. Neural morphosyntactic tagging for Rusyn. *Natural Language Engineering* 25(5). 633–650.
- Scherrer, Yves, Achim Rabus & Susanne Mocken. 2018. New Developments in Tagging Pre-modern Orthodox Slavic Texts. *Scripta & E-Scripta* 18. 9–33.
- Schwarz, Gideon. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6(2). 461–464. <https://doi.org/10.1214/aos/1176344136>.
- Senft, Gunter. 2004. What do we really know about serial verb constructions in Austronesian and Papuan languages? In Isabelle Bril & Françoise Ozanne-Rivierre (eds.), *Complex Predicates in Oceanic Languages*, 49–64. Berlin-Boston: De Gruyter Mouton. doi:doi:10.1515/9783110913286.49.
- Skupskij, Boris I. 1993. Proisxoždenie datel'nogo samostojatel'nogo v pervonačal'nom slavjanskom perevode evangelija. *Slovo* 41–43. 49–59.
- Slater, Carol. 1977. The semantics of switch-reference in Kwtsaan. *Berkeley Linguistics Society* 3. 24–36.
- Smith, Carlota S. 1978. The syntax and interpretation of temporal expressions in English. *Linguistics and Philosophy* 2. 43–99.
- Song, Sanghoun. 2017. *Modeling information structure in a cross-linguistic perspective* (Topics at the Grammar-Discourse Interface 1). Berlin: Language Science Press.
- Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Wiley-Blackwell.
- Stahlke, Herbert. 1970. Serial verbs. *Studies in African Linguistics* (1). 60–99.
- Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford: Basil Blackwell.
- Steedman, Mark J. 1977. Verbs, Time, and Modality. *Cognitive Science* 1(2). 216–234. doi:https://doi.org/10.1207/s15516709cog0102_4.
- Steedman, Mark J. 1982. Reference to Past Time. In R. Jarvella & W. Klein (eds.), *Speech, place, and action*, 125–157. New York, NY: Wiley.
- Steiner, Richard C. 2000. Does the Biblical Hebrew Conjunction $\text{—}\mathbf{1}$ Have Many Meanings, One Meaning, or No Meaning at All? *Journal of Biblical Literature* 119(2). 249–267.
- Stern, Dieter. 2013. Izučaja istočniki Bdinskogo sbornika: Slovo ob Avraamii Kidunskom i ego plemjannice Marii. *Starob lgarska Literatura* 74–91.
- Stern, Dieter. 2015. Studien zum Bdinskij Sbornik: die slavische Erzählung des Abraham von Qidun und seiner Nichte Maria und ihre griechische Vorlage. *Cyrrillomethodianum* 20. 257–287.
- Stern, Dieter. 2016. Translations as Text Redactions: Observations on Hagiographic Translations (taken from the Bdinski Sbornik). *Slavica Ambrosiana* 6. 23–41.
- Stern, Dieter. 2018. Copying Greek into Slavonic? The Slavonic branch of the Greek tradition of the life of Abraham of Qidun. In Lara Sels, Jürgen Fuchsbauer, Vittorio Tomelleri & Ilse De Vos (eds.), *Editing medieval texts from a different angle: Slavonic and multilingual traditions* Orientalia Lovaniensia Analecta (Vol. 276), 225–250. Leuven, Paris, Bristol/CT: Peeters.
- Stirling, Lesley. 1993. *Switch-Reference and Discourse Representation*. Cambridge University Press. doi:10.1017/CBO9780511597886.

- Stopponi, Silvia, Nilo Pedrazzini, Saskia Peels, Barbara McGillivray & Malvina Nissim. 2023a. Evaluation of distributional semantic models of Ancient Greek: Preliminary results and a road map for future work. In Adam Anderson, Shai Gordin, Bin Li, Yudong Liu & Marco C. Passarotti (eds.), *Proceedings of the Ancient Language Processing Workshop*, 49–58. Varna, Bulgaria: INCOMA Ltd., Shoumen, Bulgaria. <https://aclanthology.org/2023.alp-1.6>.
- Stopponi, Silvia, Nilo Pedrazzini, Saskia Peels-Matthey, Barbara McGillivray & Malvina Nissim. 2023b. Ancient Greek language models. doi:10.5281/zenodo.8369516.
- Stump, Gregory. 1985. *The Semantic Variability of Absolute Constructions*. Dordrecht: Springer.
- de Swart, Henriëtte. 1995. Contraintes aspectuelles et réinterprétation contextuelle. *Sémiotiques* 9. 89–115.
- de Swart, Henriëtte. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16. 347–385.
- Sæbø, Kjell J. 2011a. Adverbial clauses. In Claudia Maienborn, Klaus Heusinger & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 2, 1420–1441. Berlin-Boston: Mouton de Gruyter.
- Sæbø, Kjell J. 2012. Adverbial clauses. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 2, 1420–1441. Berlin-Boston: Mouton de Gruyter.
- Sæbø, Kjell Johan. 2011b. Adverbial clauses. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics. An International Handbook of Natural Language Meaning. Volume 2*, 1420–1441. Berlin: Mouton de Gruyter.
- Talmy, Leonard. 1972. *Semantic structures in English and Atsugewi*: University of California, Berkeley dissertation.
- Talmy, Leonard. 1975. Semantics and syntax of motion. In John Kimball (ed.), *Syntax and Semantics*, vol. IV, 181–238. New York: Academic Press.
- Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description. Vol. III: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. Path to realization: a typology of event integration. *Buffalo Working Papers in Linguistics* 91(01). 147–187.
- Talmy, Leonard. 2000. *Toward a cognitive semantics. Vol. 2: Typology and process in concept structuring*. Cambridge, Mass: MIT Press.
- Team, R Core. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Thompson, Sandra A., Robert E. Longacre & Shin Ja J. Hwang. 2007. Adverbial clauses. In Timothy Shopen (ed.), *Language typology and syntactic description: Complex constructions*, vol. 2, 237–300. Cambridge: Cambridge University Press.
- Thurman, Robert C. 1975. Chuave medial verbs. *Anthropological Linguistics* 17(7). 342–352.
- von Tischendorf, Constantin. 1869–1872. *Novum Testamentum Graece. 8th edn*. Leipzig: Hinrichs.

- Traugott, Elisabeth C. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Perspectives on historical linguistics*, 245–271. Amsterdam-Philadelphia: Benjamins.
- Traugott, Elisabeth C. 2003. Constructions in grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, 624–647. Oxford: Wiley.
- Traugott, Elisabeth C. & Richard B. Dasher. 2002a. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Traugott, Elisabeth C. 2004. Historical pragmatics. In Laurence R. Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*, 538–561. Oxford-Malden, MA: Blackwell.
- Traugott, Elisabeth C. 2008. "All that he endeavoured to prove was...": On the emergence of grammatical constructions in dialogic contexts. In Robin Cooper & Ruth Kempson (eds.), *Language in Flux: Dialogue coordination, language variation, change and evolution*, 143–177. London: Kings College Publications.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2002b. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs & Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. *Approaches to Grammaticalization* 1(1). 189–218.
- Trost, Pavel. 1958. O dativu absolutní. In Karel Horálek & Josef Kurz (eds.), *Sborník slavistických prací, věnovaných IV. Mezinárodnímu sjezdu slavistů v Moskvě*, 33–35. Prague: Universita Karlova ve Státním pedagogickém nakl.
- Truckenbrodt, Hubert. 1995. *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. Massachusetts Institute of Technology: MIT working papers in linguistics.
- Truckenbrodt, Hubert. 2007. The syntax–phonology interface. In Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, 435–456. Cambridge: Cambridge University Press.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*, vol. 66 2. Ithaca, NY: Cornell University Press.
- Verkuyl, Henk J. 1972. *On the Compositional Nature of the Aspects*. Dordrecht: Reidel.
- Večerka, Radoslav. 1961. *Syntax aktivních participií v staroslověněštině*. Prague: Statní Pedagogické Nakl.earlství.
- Večerka, Radoslav. 1996. *Altkirchenslavische (Altbulgarische) Syntax. Vol. 3: Die Satztypen: Der einfache Satz*. Freiburg. Br.: Weiher.
- Večerka, Radoslav. 1997. The influence of Greek on Old Church Slavonic. *Byzantinoslavica* 58. 363–84.
- de Vries, Lourens. 2005. Towards a typology of tail-head linkage in Papuan languages. *Studies in Language* 29(2). 363–384.
- Waltke, Bruce K. & Michael P. O'Connor. 1990. *An introduction to Biblical Hebrew syntax*. Winona Lake, IN: Eisenbrauns.
- Webber, Bonnie L. 1983. So what can we talk about now? In M. Brady & R. C. Berwick (eds.), *Computational models of discourse*, 331–371. Cambridge, MA: The MIT Press.
- Webber, Bonnie L. 1987. The Interpretation of Tense in Discourse. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics* 147–154.
- Webber, Bonnie L. 1988. Tense as discourse anaphor. *Computational Linguistics* 14(2). 61–73.

- Wilkins, David P. 1988. Switch-reference in Mparntwe Arrernte (Aranda): Form, Function, and Problems of Identity. In Peter Austin (ed.), *Complex sentence constructions in Australian languages*, 141–176. Amsterdam/Philadelphia: John Benjamins.
- Winer, George B. 2001 [1882]. *A treatise on the grammar of New Testament Greek*. Eugene, OR: Wipf & Stock [repr. of T. & T. Clark, 1882].
- Winter, Werner. 1976. Switch-reference in Yuman languages. In Margaret Langdon & Shirley Silver (eds.), *Hokan studies*, 165–174. The Hague: Mouton.
- Worth, Dean S. 1994. The dative absolute in the Primary Chronicle: Some observations. *Harvard Ukrainian Studies* 18. 29–46.
- Wälchli, Bernhard & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50. 671–710.
<https://doi.org/10.1515/ling-2012-0021>.
- Zaliznyak, Andrey A. 2004. *Drevnenovgorodskij dialekt*. Moscow: Jazyki slavjanskoj kul'tury.