

# **A (Contact-)Grammar of Romeyka**

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

(Dr. phil.)

in der Fakultät Geistes- und Kulturwissenschaften

der Otto-Friedrich-Universität Bamberg

im Rahmen eines Cotutelle-Vertrags mit der

Universität Gent

vorgelegt von

Laurentia Schreiber

aus

Berlin

Bamberg 2024

Diese Arbeit hat der Fakultät Geistes- und Kulturwissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen. Die Arbeit wurde im Rahmen eines Joint PhD (/Cotutelle)-Verfahrens mit der Universität Gent, Belgien, verfasst und hat auch dort als Dissertation vorgelegen.

Gutachter: Universitätsprofessor Dr. Geoffrey Haig

Weiterer Gutachter: Universitätsprofessor Dr. Mark Janse (Universität Gent)

Tag der mündlichen Prüfung: 06. Februar 2023

Dieses Werk ist als freie Onlineversion über das Forschungsinformationssystem (FIS; <https://fis.uni-bamberg.de>) der Universität Bamberg sowie über die Academic Bibliography der Universität Gent erreichbar.

Dieses Werk ist durch das deutsche Urheberrecht geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die geltende Gesetzgebung zum deutschen Urheberrecht erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

URN: urn:nbn:de:bvb:473-irb-951155

DOI: <https://doi.org/10.20378/irb-95115>

## Acknowledgements

This publication is a minimally revised version of the original dissertation submitted in November 2022 with minor corrections and comments from the PhD committee incorporated. The publication of an extensively revised version as a grammar book is in planning (as is the publication of the data corpus). Due to the validity of data and interest of the topic, the study is made available already at this stage; comments are most welcome.

The present stage of this research would have not been possible without the support of many. First and foremost, I would like to thank the Romeyka/Rumca speakers who kindly welcomed me and spared their time to let me learn about their language. Without their benevolence and support, this dissertation was clearly not possible. So, it is my hope that this work may be an enrichment to the speech community rather than a matter of concern due to sensitive identificational links. If this work and its attempt to describe the language as it is currently spoken by (at least some) speakers would be of a certain value only to a single speaker, it would have been worth the effort. Having been granted the chance to become personally attached to the speech community, I felt the inner obligation to contribute something that possibly not too many others would have a chance to do and I very much hope that it is for the good of all speakers and those who feel emotionally attached to the language.

On the side of the academic endeavour, I would like to thank my supervisors Prof Geoffrey Haig and Prof Mark Janse for their ongoing support and guidance. Furthermore, I would like to thank Dr Hakan Özkan for his continued interest in my work on Romeyka over many years which often helped me to keep up my motivation. My gratitude is also due to the late Peter Mackridge for our academic exchange which granted me with important insights about Muslim Pontic Greek.

I acknowledge financial support from the Alexander-von-Humboldt Foundation within the framework of the project on “Post-predicate Elements in Iranian: Inheritance, Contact, and Information Structure” (PIs: Geoffrey Haig, Mohammad Rasekh-Mahand) for field research in Turkey; the German Academic Exchange Service (DAAD) for a one-months research scholarship for PhD students; and the University of Ghent for funding during my six-months research stay in Ghent. While I have benefitted greatly from wide support, it is needless to say that all errors are my own.



## Summary

The present thesis is the first thorough grammatical description of Romeyka, i.e., the Muslim variety of Pontic Greek still spoken in the indigenous setting in Trabzon province in north-eastern Turkey. While the language of the Christian speech communities of Pontus has been featured in earlier research during the last hundred years – both in Pontus prior to the population exchange in 1923, and after that in Greece –, less research is available on the variety of Pontic Greek as spoken by Muslims in the area at present. As an unrecognized minority language in a Turkish-dominant majority society and due to ambivalent attitudes and identificational links of the speakers towards their language, Romeyka is currently undergoing language shift to Turkish. While the language has been in contact with Turkish for several hundred years – which arguably resulted in a number of lexical and structural contact influences albeit without leading to a mixed acrolect –, the influence of Turkish on domains where traditionally Romeyka used to be spoken has grown following labour migration from the Black Sea area to larger cities in western Turkey since the 1960s posing now – dependent upon the speech community – a threat to the vitality of Romeyka. On the other hand, Pontic Greek – and in particular the Muslim variety – is known to be a conservative representative of Asia Minor Greek preserving archaisms dating back via Hellenistic and Medieval Greek to Ancient Greek, among them so notorious grammatical features as the Romeyka infinitive, but also complex negators and pronominal forms.

It is the aim of the present thesis to describe the synchronic grammar of Romeyka as currently spoken especially by elderly speakers in the Of valley, whose variety can be considered the most archaic. As a basis for profound grammatical research – and as a contribution towards language description and documentation at the same time –, a spoken language corpus has been compiled from the variety Romeyka of Of as spoken in Çaykara based on recordings of naturalistic oral texts produced by ten speakers mainly during fieldwork in the area in 2019. In order to contribute to further scholarly research, the corpus has been transcribed, translated and fully grammatically annotated and will be made publicly available upon the publication of the thesis. In the light of the ongoing language shift in Romeyka, the present thesis focuses on the actual use of the language by bilingual – and most often Turkish dominant – speakers. In order to disentangle the factors that led to the present grammatical shape of Romeyka, three research questions targeting potential causes for the grammatical developments have been considered: (I.) How strong is the Turkish influence on the grammar of Romeyka? (II.) Does the setting of Romeyka as a minority language in a Turkish-dominant society, namely the process of language shift, play a role in grammatical change? (III.) How persistent are the inherited Greek features and what is the role of language-internal changes? However, since the primary aim of the present thesis is the description and presentation of linguistic structures forming potential candidates for an ongoing language change compared to earlier varieties of Greek and to closely related varieties of Asia Minor Greek, the theoretical questions can only tentatively be answered at the present stage of research.

It can be nevertheless stated that Romeyka seems to have preserved its Greek core with a Greek basic lexicon, albeit with considerable lexical influences from Turkish in both content and functional word classes, but probably to a lesser extent compared to its close relative Cappadocian. Furthermore, Romeyka appears to preserve indeed a number of archaic features as well as certain particularities of Pontic such as a complex system of spatial deixis. On the other hand, there are considerable Turkish influences on the structural domain such as for example in progressives, verb serialization, adjunct and relative clauses and clause combination, that have evolved under long-time contact and appear consistently across different speakers. In addition, certain contact-induced structural features especially at the syntactic domain have been found to be sensitive to the bilingual profile of the individual speaker, i.e., among others, the dominant language, namely the borrowing of the Turkish question particle

*mI*, OV word orders in declarative clauses, and an increase in non-finite complementation strategies. This finding led to the assumption that – apart from long-term contact-induced changes which seem not to impair the grammatical integrity of Romeyka – language shift is a matter of individual multilingual speaker profiles with different degrees of Turkish dominance in both language competence and use. Finally, it is argued that the specific communicative setting in which Romeyka is spoken, and which determines the communicative practice of the speakers, is decisive to set the pragmatic frame for the language(s) used: in Turkish-dominant settings, speakers tend to be more inclined to apply multiple forms of code-switching and show stronger influences of Turkish structures.

## Samenvatting

Dit proefschrift is de eerste grondige grammaticale beschrijving van het Romeyka, d.w.z. de islamitische variant van het Pontisch Grieks, die nog steeds gesproken wordt in de authochthone omgeving in de provincie Trabzon in Noord-Oost Turkije. Hoewel de taal van de christelijke taalgemeenschappen van Pontus in eerder onderzoek de afgelopen honderd jaar al aan bod is gekomen – zowel in Pontus vóór de bevolkingsuitwisseling in 1923, als daarna in Griekenland – is er minder onderzoek gedaan naar de variëteit van het Pontisch Grieks die momenteel in Turkije wordt gesproken. Als een niet-erkende minderheidstaal in een door het Turks gedomineerde maatschappij en vanwege de ambivalente talige attitudes en identificatiebanden van de sprekers ten opzichte van hun taal ondergaat het Romeyka sinds de laatste decennia een taalverschuiving naar het Turks. Hoewel de taal al honderden jaren in contact is met het Turks – wat aantoonbaar heeft geresulteerd in een aantal lexicale en structurele contactinvloeden, zonder dat dit tot nu toe tot een gemengd acrolect heeft geleid –, is de invloed van het Turks in gebieden waar traditioneel het Romeyka werd gesproken, steeds groter geworden. De invloed van het Turks is gegroeid als gevolg van arbeidsmigratie van het Zwarte Zeegebied naar grotere steden in West-Turkije sinds de jaren zestig en vormt nu – afhankelijk van de taalgemeenschap – een bedreiging voor de vitaliteit van het Romeyka. Daarnaast staat het Pontisch Grieks – en in het bijzonder de islamitische variant – bekend als een conservatieve vertegenwoordiger van het Klein-Aziatisch Grieks die archaïsche structuren bewaart die via het Hellenistisch en Middeleeuws Grieks teruggaan tot het Oudgrieks, waaronder zo beruchte grammaticale kenmerken als de infinitief, maar ook een complex system van ontkenningsspartikels en voornaamwoorden.

Het is het doel van dit proefschrift om de synchrone grammatica van het Romeyka te beschrijven zoals die momenteel vooral door oudere sprekers in de Of-vallei wordt gesproken, wat als de meest archaïsche variant van het Romeyka kan worden beschouwd. Als basis voor diepgaand grammaticaal onderzoek – en tegelijkertijd als bijdrage aan taalbeschrijving en -documentatie – is in het kader van dit proefschrift tijdens veldwerk in Turkije in 2019 een gesproken taalcorpus samengesteld op basis van opnames van natuurlijke mondelinge data van tien sprekers van het dialect van Of zoals gesproken in Çaykara. Als een bijdrage aan verder wetenschappelijk onderzoek is het corpus getranscribeerd, vertaald en volledig grammaticaal geannoteerd en zal openbaar worden gemaakt bij de publicatie van het proefschrift. In het licht van de taalovergang van het Romeyka naar het Turks richt dit proefschrift zich in het bijzonder op het feitelijke gebruik van de taal door tweetalige sprekers die meestal dominant zijn in het Turks. Om de factoren te onderzoeken die hebben geleid tot de huidige grammaticale vorm van Romeyka zijn drie onderzoeksvragen overwogen die zich richten op mogelijke oorzaken van de grammaticale ontwikkelingen in het Romeyka: (I.) Hoe sterk is de Turkse invloed op de grammatica van Romeyka? (II.) Speelt de setting van Romeyka als minderheidstaal in een door het Turks gedomineerde maatschappij en met name het proces van taalovergang een rol bij de grammaticale verandering? (III.) Hoe persistent zijn de overgeërfde Griekse kenmerken en wat is de rol van taalinterne veranderingen? Aangezien het primaire doel van dit proefschrift echter de beschrijving is van taalstructuren die potentiële kandidaten vormen voor een taalverandering van het Romeyka in vergelijking met eerdere varianten van het Grieks alsook nauw verwante varianten van het Klein-Aziatisch Grieks, kunnen de theoretische vragen in de huidige onderzoeksfase slechts voorlopig worden beantwoord.

Desalniettemin kan worden gesteld dat het Romeyka zijn Griekse kern lijkt te hebben bewaard met een Grieks basislexicon, hoewel met aanzienlijke lexicale invloeden van het Turks in zowel lexicale als functionele woordklassen maar toch in mindere mate vergeleken met zijn naaste verwant, het Cappadocisch. Bovendien lijkt het Romeyka inderdaad een aantal archaïsche kenmerken te behouden, evenals bepaalde bijzonderheden van het Pontisch Grieks zoals een complex systeem van ruimtelijke deixis. Aan de andere kant zijn er aanzienlijke

Turkse invloeden op het structurele domein, zoals bijvoorbeeld in de progressieve tijd, seriële werkwoorden, bijwoordelijke en relatieve bijzinnen en “clause combining”, die zijn geëvolueerd onder langdurig contact en die consistent voorkomen bij verschillende sprekers. Bovendien blijken bepaalde contact-geïnduceerde structurele kenmerken, vooral op het syntactische vlak, gevoelig te zijn voor het tweetalige profiel van de individuele spreker en in het bijzonder diens dominante taal. Dit betreft fenomenen zoals het ontlenen van het Turkse vraagpartikel *mI*, OV-woordvolgorde in mededelende zinnen, en een toename van niet-finiëte complementatiestrategieën. Deze bevinding leidt tot de hypothese dat – afgezien van door langdurig taalcontact veroorzaakte veranderingen die de grammaticale integriteit van het Romeyka niet aantasten – taalverandering een kwestie is van individuele meertalige sprekersprofielen met verschillende gradaties van Turkse dominantie in zowel taalvaardigheid als taalgebruik. Ten slotte wordt beargumenteerd dat de specifieke communicatieve setting waarin het Romeyka wordt gesproken, en die de communicatieve praktijk van de sprekers bepaalt, een invloed heeft op pragmatische keuzes: in Turks-dominante omgevingen zijn sprekers meer geneigd om verschillende vormen van code-switching toe te passen en sterkere invloeden van Turkse structuren te tonen.



## Table of contents

1	Introduction .....	1
1.1	Sociolinguistics of the speech community .....	1
1.1.1	Ethnolinguistic vitality .....	1
1.1.2	Remarks on naming conventions .....	2
1.1.3	Historical situation & migration.....	4
1.1.4	Bilingualism & language shift.....	5
1.2	Typological profile .....	6
1.2.1	Genealogical classification.....	6
1.2.2	Dialectal variation .....	7
1.2.3	Language contact.....	8
1.3	State of research.....	8
1.4	Aims of the study & organisation.....	9
1.5	Methodology.....	12
1.5.1	Linguistic fieldwork .....	13
1.5.1.1	The speakers.....	13
1.5.1.2	Fieldwork materials .....	14
1.5.1.3	Fieldwork process .....	16
1.5.2	Data processing .....	16
1.5.2.1	Corpus building.....	16
1.5.2.2	Corpus design.....	17
1.5.3	Critical reflection of the methodology .....	19
1.5.4	Decisions in presentation .....	20
1.5.5	Pending steps in language description and archiving.....	24
2	Phonology.....	27
2.1	Segments.....	28
2.1.1	Vowels.....	28
2.1.2	Consonants .....	33
2.1.2.1	Voicing and aspiration on stops .....	36
2.1.2.2	Nasals .....	37
2.1.2.3	Tap .....	38
2.1.2.4	Approximants.....	38
2.1.2.5	Fricatives.....	39
2.1.2.6	Affricates.....	41
2.2	Phonotactics.....	42
2.2.1	Diphthongs and vowel sequences .....	42
2.2.2	Consonant clusters.....	44
2.2.3	Syllable structure and word shapes .....	46
2.3	Phonological processes and morphophonology.....	48
2.3.1	Palatalization and affrication.....	48
2.3.2	Assimilation .....	49
2.3.3	Final vowel lengthening .....	50
2.3.4	Variation in fricatives.....	52

2.3.5	Elision.....	52
2.3.5.1	Apheresis.....	52
2.3.5.2	Apocope .....	52
2.3.5.3	Crisis.....	53
2.3.5.4	Reduction /nd/ > /d/ .....	54
2.3.6	Consonant epenthesis .....	54
2.3.7	R-Metathesis.....	56
2.4	Prosody .....	56
2.4.1	Word stress .....	56
2.4.2	Sentence stress.....	58
2.5	Orthography.....	63
3	Word classes.....	65
3.1	Classes of content words .....	65
3.1.1	Nouns .....	65
3.1.1.1	Lexical nouns .....	65
3.1.1.2	Proper nouns .....	66
3.1.2	Adjectives.....	68
3.1.2.1	Attributive adjectives .....	69
3.1.2.2	Predicative adjectives.....	73
3.1.3	Numerals .....	75
3.1.4	Adverbs .....	78
3.1.4.1	Adverbs of degree .....	78
3.1.4.2	Adverbs of manner.....	80
3.1.4.3	Adverbs of time.....	82
3.1.4.4	Adverbs of place and spatial orientation.....	84
3.1.4.4.1	Simple adverbs of place.....	87
3.1.4.4.2	Post-posed localizers <i>-tšega, -tšes, -merea</i> .....	88
3.1.4.4.3	Pre- and suffixal demonstrative <i>ha</i> .....	89
3.1.4.4.4	Other complex spatial expressions .....	90
3.1.5	Verbs .....	91
3.2	Function word categories.....	92
3.2.1	Articles .....	92
3.2.1.1	Definite articles .....	93
3.2.1.2	Indefinite articles.....	98
3.2.2	Pronouns.....	99
3.2.2.1	Personal pronouns .....	100
3.2.2.1.1	Subject personal pronouns.....	102
3.2.2.1.2	Object personal pronouns .....	103
3.2.2.2	Demonstratives .....	107
3.2.2.3	Possessives .....	111
3.2.2.4	Reflexives .....	117
3.2.2.5	Reciprocals.....	118
3.2.2.6	Relativizers .....	119

3.2.2.7	Indefinite pronouns .....	122
3.2.2.8	Interrogatives .....	125
3.2.3	Quantifiers .....	129
3.2.4	Particles .....	135
3.2.4.1	Negation particles .....	136
3.2.4.2	Question/answer particles .....	136
3.2.4.3	Borrowed clause linkage markers .....	136
3.2.4.4	Modal particles.....	138
3.2.4.5	Discourse particles .....	141
3.2.4.5.1	Topic and focus particles .....	141
3.2.4.5.2	Other discourse markers .....	144
3.2.5	Prepositions .....	146
3.2.6	Conjunctions.....	150
3.2.6.1	Coordinating conjunctions .....	150
3.2.6.1.1	Coordinating conjunction <i>tš(e)</i> .....	150
3.2.6.1.2	Disjunctive conjunctions .....	154
3.2.6.1.3	Adversative conjunctions.....	155
3.2.6.1.4	Correlative conjunctions.....	156
3.2.6.2	Subordinating conjunctions .....	156
3.2.7	Interjections .....	159
4	Morphology.....	163
4.1	Word formation .....	163
4.1.1	Derivation.....	163
4.1.1.1	Nominal derivation .....	163
4.1.1.2	Adjectival derivation.....	164
4.1.1.3	Adverbial derivation .....	164
4.1.1.4	Verbal derivation .....	164
4.1.2	Compounding .....	165
4.1.3	Other strategies of word formation .....	166
4.1.4	Loan words.....	166
4.2	Nominal morphology.....	168
4.2.1	Nominal inflection.....	169
4.2.1.1	Gender.....	170
4.2.1.2	Number .....	171
4.2.1.3	Case.....	177
4.2.2	Possession.....	180
4.2.3	Definiteness .....	184
4.2.4	Diminution .....	192
4.3	Verbal morphology .....	192
4.3.1	Person agreement .....	193
4.3.2	Tense .....	197
4.3.2.1	Present.....	197
4.3.2.2	Aorist.....	198

4.3.2.3	Imperfect .....	200
4.3.2.4	Future .....	202
4.3.3	Aspect.....	203
4.3.3.1	Phasal aspect .....	203
4.3.3.2	Lexical aspect.....	204
4.3.3.3	Periphrastic progressive and continuative aspect .....	205
4.3.4	Mood .....	209
4.3.5	Voice .....	210
4.3.6	Negation and (irrealis) modality .....	211
4.3.6.1	Negation .....	211
4.3.6.2	Subjunctive mood .....	219
4.3.6.3	Optative mood.....	220
4.3.6.4	Expressions of wish, desire or intention .....	221
4.3.6.5	Potential constructions .....	225
4.3.6.6	Expressions of external obligation .....	228
4.3.6.7	Non- and antiveridical irrealis mood .....	229
4.3.7	(Non-)Finiteness.....	229
4.3.7.1	Infinitives .....	229
4.3.7.2	Nominalization.....	233
4.3.7.3	Participles.....	233
4.3.7.4	Other non-finite verb forms .....	235
5	Syntax.....	236
5.1	Structure of the noun phrase .....	236
5.2	Independent clauses .....	239
5.2.1	Simple declarative clauses .....	239
5.2.1.1	Word order in simple declarative clauses .....	240
5.2.1.2	Intransitive clauses and pro-drop .....	245
5.2.1.3	Transitive clauses .....	247
5.2.1.4	Ditransitive clauses .....	249
5.2.1.5	Predicatives .....	252
5.2.1.5.1	Predicate nominals.....	253
5.2.1.5.2	Predicate adjectives .....	254
5.2.1.5.3	Predicative locatives and other predicatives.....	255
5.2.1.5.4	Existentials.....	255
5.2.1.5.5	Predicative possession .....	257
5.2.1.6	Comparative constructions.....	259
5.2.2	Valency-changing processes .....	263
5.2.2.1	Causatives .....	263
5.2.2.2	Passives .....	264
5.2.3	Non-declarative clauses.....	265
5.2.3.1	Imperative clauses.....	265
5.2.3.2	Exclamatory clauses.....	266
5.2.3.3	Interrogative clauses .....	266

5.2.3.3.1	Polar questions.....	267
5.2.3.3.2	Leading questions .....	268
5.2.3.3.3	Alternative questions .....	269
5.2.3.3.4	Content questions.....	269
5.2.3.3.5	Embedded interrogatives .....	271
5.3	Complex clauses and clause combining .....	273
5.3.1	Coordination.....	273
5.3.1.1	Clause coordination .....	273
5.3.1.2	Verb coordination (serial verbs) .....	275
5.3.2	Complementation .....	278
5.3.2.1	Finite complementation .....	278
5.3.2.1.1	Clausal complements lacking an overt complementiser.....	278
5.3.2.1.2	Clausal complements of verbs of saying .....	280
5.3.2.1.3	<i>na</i> -Clauses.....	283
5.3.2.2	Non-finite complementation .....	284
5.3.2.2.1	Infinitives.....	284
5.3.2.2.2	Deverbal nouns .....	287
5.3.3	Adjunct clauses .....	289
5.3.3.1	Adverbial clauses .....	289
5.3.3.1.1	Adjuncts of time .....	290
5.3.3.1.2	Adjuncts of location.....	294
5.3.3.1.3	Adjuncts of manner.....	295
5.3.3.1.4	Adjuncts of purpose.....	295
5.3.3.1.5	Adjuncts of reason .....	296
5.3.3.1.6	Circumstantial adjuncts .....	297
5.3.3.2	Conditional clauses .....	297
5.3.4	Relative clauses .....	301
6	Language contact.....	304
6.1	Pragmatic aspects of language contact: code-switching vs. lexical borrowing.....	304
6.1.1	Lexical borrowing .....	305
6.1.2	Code-switching.....	306
6.2	Contact-induced grammatical change.....	310
6.3	Language internal change and inheritance .....	312
6.4	Language shift-induced change .....	313
7	Conclusion.....	314
	References .....	316
	Appendix A: Corpus overview.....	A.i
	Appendix B: Phonological analyses.....	B.i
	Appendix C: Token frequencies.....	C.i
	Appendix D: Spatial adverbials.....	D.i

## List of Figures

Figure 1: Approximate locations of the dialect groupings of Romeyka .....	7
Figure 2: Annotation tiers of the Romeyka corpus in ELAN.....	17
Figure 3: The Romeyka vowel inventory .....	28
Figure 4: Vowel lengthening in a closed unstressed syllable (word <i>árgo:s</i> ‘bear’).....	51
Figure 5: Pitch and intensity in the spectrogram of the word <i>xóma</i> ‘earth’ .....	57
Figure 6: Intonation contour of a declarative clause .....	59
Figure 7: Intonation contour of coordinated phrases.....	59
Figure 8: “Climactic” intonation contour .....	60
Figure 9: Intonation contour of focus .....	60
Figure 10: Intonation contour of a polarity question.....	61
Figure 11: Intonation contour of <i>wh</i> -questions .....	61
Figure 12: Intonation contour of a topicalized <i>wh</i> -question .....	62
Figure 13: Intonation contour of a polarity question with Turkish interrogative particle.....	62
Figure 14: Intonation contour of an imperative clause.....	63
Figure 15: Complex intonation contour including focus, <i>wh</i> -element and imperative .....	63

## List of Tables

Table 1: Data (word lists) considered for phonological analysis .....	27
Table 2: The consonant inventory of Romeyka .....	34
Table 3: Consonant clusters in Romeyka .....	45
Table 4: Word patterns in Romeyka according to word class.....	47
Table 5: Word-initial [h] in Romeyka and Trabzon Turkish .....	56
Table 6: Word stress in Romeyka (poly-)syllables .....	57
Table 7: Some phoneme-grapheme correspondences in written Romeyka .....	64
Table 8: Absolute word form frequencies of native and loan adjectives .....	69
Table 9: Adjectival declension in Romeyka .....	70
Table 10: Cardinal numerals in the Romeyka corpus till value 10 .....	77
Table 11: Definite articles in Romeyka.....	94
Table 12: Pronouns in subject and object function in the corpus.....	100
Table 13: Subject and object pronouns in Romeyka.....	101
Table 14: Demonstrative pronouns in Romeyka.....	108
Table 15: Full and bound possessive pronouns in Romeyka .....	113
Table 16: Interrogatives in Romeyka .....	126
Table 17: Inflectional paradigm of quantifier <i>ulon</i> .....	131
Table 18: Prepositions in Romeyka .....	146
Table 19: Plural declension in Romeyka of Of.....	173
Table 20: Nominal case declension in Romeyka .....	178
Table 21: Examples of nominal declension in Romeyka .....	179
Table 22: Inflectional paradigm of <i>ime</i> ‘be’ .....	193
Table 23: Inflectional paradigm of <i>inume</i> ‘become’ .....	194
Table 24: Inflectional paradigm of <i>exo</i> ‘have’ .....	194
Table 25: Present tense inflectional paradigms of Romeyka verbs in <i>-o</i> and <i>-me</i> .....	195
Table 26: Tense paradigm of <i>tro(y)o</i> ‘eat’ .....	196
Table 27: Tense paradigm of <i>pa(y)o</i> ‘go’ .....	196
Table 28: Aorist tense inflectional paradigms of Romeyka verbs in <i>-o</i> and <i>-me</i> .....	198
Table 29: Aorist tense inflectional paradigm of <i>elebo</i> ‘see’ .....	199
Table 30: Aorist inflectional paradigm of <i>diyo</i> ‘give’ .....	199
Table 31: Imperfective endings in ROF.....	201
Table 32: Inflectional paradigm of the verb <i>kahume</i> ‘sit’ .....	201
Table 33: Periphrastic present progressive forms with <i>kahome</i> ‘sit’ and <i>steko/stekume</i> ‘stand’ on the example of <i>tšimume</i> ‘sleep’ .....	207
Table 34: Morphological imperatives in Romeyka.....	209
Table 35: Negation of indicative clauses in Romeyka .....	212
Table 36: Order of oblique and verb in Romeyka.....	243
Table 37: Non- and antiveridical mood in Romeyka .....	298

## Abbreviations

ACC	accusative
AOR	aorist
CL	clitic
CMP	comparative
COP	copula
COMP	complementiser
COND	conditional particle
DEM	demonstrative
DET	definite article
DIM	diminutive
EMPH	emphatic (pronoun)
F	feminine
GEN	genitive
GER	gerund
HUM	human
IMP	imperative
INDET	indefinite article
INF	infinitive
IPFV	imperfective
M	masculine
N	neuter
NEG	negation
NMZ	nominalization
NOM	nominative
OPN	object pronoun
OPT	optative
PASS	passive
PL	plural
POSS	possessive
PRS	present
PRT	modal particle
PTCP	participle
Q	question particle
REL	relativizer
SG	singular
TOP	topicalizer
1, 2, 3	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> person



## 1 Introduction

Romeyka is an endangered Indo-European language, belonging to the Hellenic branch. It is an indigenous minority language that has been spoken continuously at the southern Black Sea coast in Turkey (historically known as Pontus) since antiquity (the first inhabitancy of Pontus by Greeks is dated around the 7<sup>th</sup> c. BCE, Sitaridou 2014b) and is still spoken by Muslim communities *in situ*, as well as by diaspora speakers in Turkey's large cities and elsewhere abroad. Although Romeyka can be considered a modern Greek variety of the Pontic branch, it is assumed to differ structurally (see Section 1.2.1) from the Pontic Greek varieties as spoken in Greece and in other countries around the Black Sea (for example, Pontic Greek in Georgia, see Berikashvili 2017). As Oikonomidis (1908: VII) puts it:

“[...] [D]enn mit dem Sturze des Kaisertums von Trapezunt ist alles zerstört und vernichtet worden; gewiss von da ab steigerte sich besonders der Verfall des pontischen Dialekts und das beweist einerseits die verhältnismässig reinere Sprache der aus jener Zeit erhaltenen Volkslieder und andererseits das Idiom derjenigen Ophiten, die im 17. Jahrh. zur mohamedanischen Religion übergetreten sind und noch heute das Griechische in manchen Beziehungen reiner als die übrigen Pontier sprechen [...]”<sup>1</sup>

(Christian) Pontic Greek as spoken in present-day Greece has more speakers than Romeyka and is as a modern Greek dialect relatively well documented, while many structural aspects of Romeyka are still only poorly understood, though ongoing research is rapidly filling the gaps in our knowledge (Sitaridou 2013, 2014a, 2014b, 2016; Neocleous 2020; Schreiber 2018; Neocleous & Sitaridou 2022; Neocleous 2022; i.a.). Romeyka has not only on the one hand preserved more archaic structures than (Christian) Pontic Greek and lacks levelling with Modern Greek like the latter, but on the other hand, its sociolinguistic situation as a minority language in Turkey that has led to an intense language contact setting with Turkish and ultimately to language decline makes the study of present-day Romeyka in Turkey a fascinating real-time documentation of an ongoing language shift. A complex set of geographical and historical circumstances have – among others – led to the current sociolinguistic situation of Romeyka, which is briefly outlined in the following sections. This chapter aims to summarize the current state of knowledge and refers to the available literature for more detailed coverage. The sociolinguistic information in this chapter is largely going back to Schreiber (2016) and Schreiber & Sitaridou (2017); the presentation of information is closely based on Schreiber (2018).

### 1.1 Sociolinguistics of the speech community

#### 1.1.1 Ethnolinguistic vitality

Existing assessments of the sociolinguistic vitality of Romeyka vary between “definitely endangered” (Moseley 2010), “seriously/severely endangered” (Moseley 2007), “threatened” (Catalogue of Endangered Languages 2017) and “vigorous” (Lewis, Simons & Fennig 2016). As Schreiber (2016, and consequently Schreiber & Sitaridou 2017) have argued, these assessments are questionable because they (a) do often not differentiate between Romeyka and

---

<sup>1</sup> Engl. translation: “Because with the fall of the Empire of Trebizond everything was destroyed and annihilated; certainly from then, on the decline of the Pontic dialect increased in particular and this is proven on the one hand by the relatively purer language of the folk songs preserved from that time and on the other hand, the idiom of those Ophites, who converted to the Mohammedan religion in the 17<sup>th</sup> century and who still speak the Greek language in some respects more purely than the rest of the Pontians.” [translation LS]

Pontic Greek (hence speaker numbers are inaccurate), and (b) are based on insufficient data. Following the sociolinguistic vitality assessment of Schreiber (2016), the vitality of Romeyka as an unrecognized minority language in Turkey is much more vulnerable than previously assumed (see Schreiber & Sitaridou 2017 for an overview of the sociolinguistic variables threatening Romeyka).

The actual number of Romeyka speakers is not known. The last official number reports 4,535 speakers and stems from the 1965 Turkish general census (*Genel Nüfus Sayımı*), which was the last of its kind accounting for different mother tongues (reported in Mackridge 1987; Andrews 1989). Estimates based on the populations of the villages in Trabzon province where Romeyka is spoken run to approx. 300,000 speakers (Biliçi 2011), but this kind of estimate is obviously exceedingly coarse-grained (see Schreiber & Sitaridou 2017). In general, the estimation of speaker numbers is difficult due to (a) practical reasons such as internal and external migration and residence patterns, (b) a biased choice of Turkish as mother tongue, (c) mixed Turkish-Romeyka identity at least in the urban communities; and (d) patterns of language competence varying greatly between speakers which make it difficult to decide who can be considered a “speaker” as opposed to an “overhearer”, “rememberer”, etc. (for the terminology of varying language competences in language shift scenarios, see e.g., Grinevald & Bert 2022).

Romeyka speakers may harbour multiple identity constructs (for a sociological study on identities of Romeyka speakers in Trabzon district, see Sağlam 2017), such as Turkish national and citizenship identity, Muslim religious identity and, dependent on the speech community, a sense of cultural Romeyka identity (Sitaridou 2013). The latter is less pronounced in urban communities, which tend towards assimilation into a Turkish mainstream (Schreiber 2016). In general, Romeyka speakers deny any Rum or Greek identity (Sitaridou 2013) and often exhibit conservative Muslim and nationalistic identification links (Özkan 2013). The attitudes speakers hold towards their language have been investigated by Schreiber (2016, and consequently Schreiber & Sitaridou 2017) for the Istanbulite speech community and showed that the attitudes of this speech community towards Romeyka, its use and its maintenance are not very positive, although the sociolinguistic variables of age and gender proved to be decisive in this regard.

### 1.1.2 Remarks on naming conventions

Dealing with Romeyka as an indigenous minority language in Turkey, some remarks on the naming conventions are in order. Romeyka is the term that has become widespread in recent research (i.e., work undertaken by Sitaridou 2013 et pass. and Neocleous 2020, 2022), though “Muslim Pontic”, a term introduced by Mackridge (1987), is used as a technical term, too (namely by Özkan 2013; partially Brendemoen 2006; Bortone 2009; see also Schreiber & Sitaridou 2017 on glossonomy). Following Sitaridou (2013), the term Romeyka is in line with what many speakers call their language (for a terminological clarification of the language/dialect issue, see Section 1.2.1), though others use the term “Rumca”, especially when speaking in Turkish. As argued by Schreiber and Sitaridou (2017), the use of either the Greek term *Romeyka*, or the Turkish term *Rumca*, corresponds to the ethnolinguistic vitality of the particular speech community. Importantly, it should be noted that the Romeyka speakers who contributed to the present corpus of Romeyka appear to call their language predominantly by the Turkish term *Rumca* and not by the Greek term *Romeyka*, which seems to be even unfamiliar to many of the speakers. However, the language in which the fieldwork is administered may probably affect the terminology speakers use for their language. Further research on the preferred auto-glossonyms of the speakers is required. Interestingly, in the present corpus, speaker may refer to their language as *orumdža* with a pre-posed rounded close-mid back vowel, which resembles the form of the masculine definite article, e.g., *so Istanbul*

*op ine ul o rumdža u kserone* ‘The [speakers] who are in Istanbul don’t speak [lit. ‘know’] Rumca.’ (A1).<sup>2</sup>

In the following, the different glossonyms used for the Greek variety in question are briefly discussed in more detail:

- (i) Romeyka: Romeyka is the name of the Greek language of the Byzantine (i.e., East Roman) Empire (*ρωμαίικα*, lit. “Roman”) which is still being used by ideologists of the Megali Idea (‘Great Idea’ of a contiguous Greek nation state) in Greece. This connotation in Greece may makes the term somewhat problematic for speakers of the variety in Turkey, who certainly do not have such intensions although the reference to the “Byzantine” past is visible. The term Romeyka has been for hundreds of years used as auto-glossonym to denote the language of all speakers of Greek outside mainland Greece (except for southern Italy, P. Mackridge, p.c.); however, there seem to be apart from Pontus at present no other members of Asia Minor Greek varieties who use the term for their language; speakers of Cappadocian, for example, would rather use the name of the local community to refer to their variety (M. Janse, p.c.). Note furthermore, that there is a range of different variants in spelling of the term Romeyka which try to account for the implicit connotations of the word (see <https://www.romeyka.org/rediscovering-romeyka/>, accessed 25 Oct 2022) and have resulted in the spelling “Romeyka”. For example, the speakers of Romeyka in Uzungöl (Saráchos) are reported to refer to their variety as *Ρουμάικα* (P. Mackridge, p.c.). Note that the spelling of the term adopted on some Turkish websites about the language (e.g., <https://tr.wikipedia.org/wiki/Trabzon>, accessed 15/08/2022) and Tursun’s (2019) dictionary is “Romeika” (also Sağlam 2017).
- (ii) Pontiaka is another Greek term to refer to the Greek varieties of the historical area of Pontus at the Black Sea coast in today Turkey, albeit it has been primarily used as an auto-glossonym by the Christian Pontic speakers who were forced to leave the area in 1923. A Turkish translation of this term which is occasionally used in Turkish with implicit reference to the Pontic Empire is Pontosça.
- (iii) Rumca is the common Turkish term used for all Greek varieties in Asia Minor, which has been originally reserved for Christians and is recently mainly associated with the Istanbul Christian Greek-speaking community (for further discussion of the term see also Sağlam 2017: 92). In order to distinguish between the different Rumca communities, further spatial terms may be added, e.g., Trabzon Rumcası, Pontos Rumcası.
- (iv) The Turkish term Lazca is occasionally used in Turkey to informally refer to the Turkish dialect of the Black Sea coast, although it actually refers to the Kartvelian language Laz spoken in the Rize district of Turkey. In a few instances, the term may be also used extending to the different minority languages of the Black Sea area including Romeyka. In Greece, Christian Pontic Greek speakers may use the Greek equivalent Lazika to denote their language (M. Janse, p.c.).
- (v) Muslim Pontic Greek has been coined as a technical term by Peter Mackridge, which has been taken over by subsequent work, for example, by Özkan (2013) and (partly) Brendemoen (2002), but which has been also partly objected from others for the use of the religious term. Related technical terms that have been suggested by Peter Mackridge (p.c.) are “Pontic Greek as spoken in Turkey” or, more specifically, “Greek spoken in Çaykara province”.
- (vi) The term Ophitic is used to denote the Pontic Greek variety spoken in the Of district of Turkey in both the past and present by both Christians and Muslims (see Sitaridou

<sup>2</sup> This phenomenon is still to be explained. Note that in the Georgian population of Turkish-speaking Urum people (Skopeteas 2016), the pre-posed vowel in the language name Urum is inserted for phonological reasons to adhere to the syllable rules (S. Skopeteas, p.c.).

2014b: 25, Fn.1) but appears in recent research to be primarily associated with the language of Christian Pontic Greek speakers from the area of Of living in Greece since 1923 (see Revithiadou & Spyropoulos 2012). However, starting with his 1987 article, Peter Mackridge had also used the term Ophitic as to apply to the present-day Muslim Greek variety spoken in “the whole area south of the town of Of up to the border of Bayburt (i.e., including the administrative districts of Dernekpazarı and Çaykara)” (P. Mackridge, p.c.) to distinguish it from other regional varieties of Romeyka such as that of Tonya, which might have caused some confusion as to the coverage of the term Ophitic.

In sum, while Trabzon Rumcaşı is my preferred choice (see also Genç 2021) to reflect both the identity of the majority of my speakers, to acknowledge the reality of the dominance of Turkish as the speech community’s dominant language and on the other hand to be precise enough in distinguishing the variety from other (Christian) Greek varieties that have been spoken in Turkey, the term Romeyka is adopted throughout this thesis as a technical term to pay credit to the denomination used predominantly in recent research on the same varieties as this thesis and to acknowledge the reported auto-glossonym of speakers (Sitaridou 2013) albeit this is less common among the speakers I have worked with.

In order to differentiate between diatopic varieties of Romeyka as spoken in present-day Turkey, regional variants are (based on Sitaridou’s 2014b nomenclature) referred to with the name of the location in which the variety is spoken, e.g., Romeyka of Of (abbreviated ROf), Romeyka of Sürmene (abbreviated RSür). If no reference is made to a specific variety, Romeyka is abbreviated as R., for example as opposed to Modern Standard Turkish (abbreviated Tr.) or regional Trabzon Turkish (abbreviated TTr.). The term Pontic Greek (PG) is used to denote the language of Christian Pontic Greek speakers in Greece after 1923 (see, e.g., Drettas 1997). The variety of Pontic Greek spoken by Christians in Pontus since medieval times (Sitaridou 2014b: 25, Fn. 1) and before 1923 is referred to as Christian Pontic Greek. The following other abbreviations apply: Ancient Greek (AG) is used as a cover term for the Greek language of the Classical period between 5<sup>th</sup> c. BCE – 2<sup>nd</sup> c. BCE; the term Hellenistic Greek (HelGr) is used for the period between the 1<sup>st</sup> c. BCE – 5<sup>th</sup> c. CE. Medieval Greek (MedGr) refers to the Greek language of the period between 6<sup>th</sup> c. CE – 16<sup>th</sup> c. CE (periodization following Sitaridou 2014b: 37, Fn. 15). Cappadocian (short Capp.) refers to the variety of Cappadocian as spoken at present in certain villages in Greece (see Janse 2022). For a precise definition of the Romeyka variety studied in the present thesis, see Section 1.4 (point 4).

### 1.1.3 Historical situation & migration

Today Romeyka is spoken in approximately fifty villages in Trabzon province around the cities of Sürmene, Of, Çaykara and Tonya (Mackridge 1987; Andrews 1989; Özkan 2013; see Section 1.2.2). This is the remainder of what has been once a large area where Pontic Greek was spoken reaching from the Black Sea coast in the north to the Pontic Alps in the south. Christianisation of Pontus took place around the 4<sup>th</sup> c. CE and facilitated the expansion of Greek in the area (Sitaridou 2014b). With the Siege of Trebizond by the Ottomans in 1461, the history of Byzantine civilization in the area came to an end. Islamicization took place probably as of the 16<sup>th</sup> c. (Lowry 1977: 209–247; Vryonis 1986), whereby Christian and Muslim communities potentially cohabited for several hundred years. The demise to Pontic Greek as spoken by Christians in the Black Sea region was the relocation of Christian Pontic Greek speakers in Greece in the context of the Treaty of Lausanne in 1923. However, already in the 18<sup>th</sup> and 19<sup>th</sup> c., Christian Pontic Greeks had migrated in successive migration waves to Southern Russia and Caucasia (Moseley 2007) and later to the US, the Ukraine, and Greece (Brendemoen 2002). The remaining Muslim Pontic Greek speaking population became mobile when labour migration took place as of the 1950s, so, today there exist speech communities in urban centres

such as Ankara and Istanbul whose language vitality differs significantly from rural communities (Schreiber & Sitaridou 2017). Since 1960, diaspora communities exist in Germany and other central European countries (Özkan 2013). Still, weak economic opportunities in the area are a threatening factor for the vitality of Romeyka (Brendemoen 2002). It is assumed that the loss of traditional ways of lives and the loosening of the traditionally close social networks are a considerable factor in Romeyka endangerment (Schreiber 2016).

#### 1.1.4 Bilingualism & language shift

Today, all Romeyka speakers are (at least) bilingual in Romeyka and Turkish. Depending on the speech community (i.e., the village), the youngest generation may acquire Romeyka still as early L2 (Sitaridou 2013) or, respectively, as heritage language only (see Schreiber & Sitaridou 2017: 10, Table 4). While the oldest generation shows nearly simultaneous bilingualism, whereby many elderly speakers of the (great-)grandparent generation acquired Romeyka as L1 and learned Turkish only subsequently when entering school, the middle-aged (approx. 40–60 years) first parent generation shows additive bilingualism with Turkish as L1 and Romeyka as L2 in the home and informal domains. Especially in the urban speech communities (Schreiber 2016) but apparently also in many of the less remote villages *in situ*, children and adults below age 40 did not acquire Romeyka anymore, although they might know some words and partly have passive competences. This situation indicates that language shift is already completed (at least for the urban communities and many of the speech communities *in situ*). However, when it comes to individual language competences, the multilingual profiles of speakers appear to vary greatly (for some of the factors leading to different language competences even among siblings in the same family, see Schreiber 2016: 37–39). When differentiating between different linguistic competences in a language shift setting, the terminology to indicate different levels of productive and perceptive language competences ranges between terms such as semi-speakers, rememberers, or over-hearers (Grinevald & Bert 2011). In general, younger generations display often only perceptive language competence (if any at all), while the elder generations have still productive competences. Still, further research is required to account for very strong differences in individual language competences (Schreiber 2022) and to determine the factors causing this variation.

Alongside the lack of language transmission, the domains where Romeyka is used are diminished with an ongoing decline. The sociolinguistic attitudinal study by Schreiber (2016) has shown that Romeyka is (especially in the Istanbulite speech community) only used in the home and for informal conversation with neighbours and friends, as well as at some informal situations in the villages like the café. However, further detailed research on language use patterns of Romeyka in the villages *in situ* is in order; especially complementing self-reported attitudinal questionnaire data, with participant observation and other ethnographic means of field work. Furthermore, a larger coverage of different speech communities in the area would be necessary to account for different forms of multilingualism even between neighbouring villages (for a more comprehensive discussion of differences in language competence between different villages of Romeyka of Of as spoken in Çaykara, see Sağlam 2017: 94–97).

Romeyka has no literacy tradition and is predominantly a spoken variety. The development of a written form of Romeyka is not encouraged by Turkish language education policies, and it generally lacks the range of registers for more official domains. So far, no widely accepted orthography has been developed for Romeyka, although Romeyka appears recently in written form in social media and on some local websites using the Turkish alphabet (see Section 2.5).

## 1.2 Typological profile

### 1.2.1 Genealogical classification

Romeyka belongs to the Hellenic branch of the Indo-European language family. Within Hellenic, it is affiliated with the Attic-Ionic branch (i.e., in particular Ionic). It is as a variety of Pontic Greek part of the Asia Minor Greek group<sup>3</sup> where it is considered to form a sub-group with Cappadocian as opposed to other Asia Minor Greek varieties (for a detailed account of the genealogical classification of AMG see Karatsareas 2011: 40–55; cf. Sitaridou 2014b). Romeyka was isolated from other varieties of MedGr and Proto-Pontic Greek around the 14–16<sup>th</sup> c. CE (Sitaridou 2014b: 27). Several aspects of Romeyka’s origins remain controversial: it is unclear how far Romeyka was involved in the development of an Asia Minor *koiné* (Sitaridou 2014b: 29), and whether Medieval Pontic had diverged into Christian and Muslim varieties already after Islamization in the 16<sup>th</sup>/17<sup>th</sup> c., and thus before expulsion of the Christian speakers from Turkey in 1923, as argued by Sitaridou (2014b: 31). The latter would be a question of some relevance, though, as much more records are available of Christian Pontic Greek and the layer of evidence would be less thin for Muslim Pontic Greek, if the varieties were comparable in terms of their grammar. On the other hand, Christian Pontic Greek is assumed to have been in contact with other modern Greek varieties prior to 1923 and may in its written records display a different register compared to the spoken language. Especially after 1923, Pontic Greek in Greece underwent increasingly levelling with other Modern Greek varieties.

It is not clear, though, whether Romeyka owes its striking archaisms which it preserved unlike PG to the lack of contact with SMG after 1923 or already an earlier separation of the two varieties in Pontus. In literature, it is generally reported that the Christian Pontic Greek speakers settled prior to 1923 in villages closer to the sea (like in the case of the Ottoman province of Of, where most of the Christian villages were located in the north-east of the province, P. Mackridge, p.c.). This would be mere geographical distance and location make intense (language) contact between the Christian and Muslim communities less likely, although mixed communities must have existed as well. Indeed, Parcharidis (1880) reports from the mixed village of Bölümlü (Zisíno) lexical differences between the Christian and Muslim varieties. To shed more light on the crucial aspect of convergence or divergence between the Christian and Muslim Pontic Greek varieties, more historical sociolinguistic research is in order.

Finally, having dealt with the genealogical classification of Romeyka, some brief notes on the language vs. dialect discussion are in order: Technically speaking, Romeyka is a diatopic variety of Modern Greek which is located in Turkey as a language island and thus differentiated from other Modern Greek dialects not only in terms of its geographical location and history but also in terms of its sociolinguistic situation, with the Muslim faith of its speakers being an important identity factor. These are reasons to consider Romeyka in terms of its sociolinguistic characterisation as a minority language in Turkey, and thus a “language”, although this is probably not vindicable solely based on its linguistic features, despite the retention of archaisms. Note that the *Ethnologue* (Eberhard, Simons & Fennig 2023) considers “Pontic” a language of Greece (ISO 639-3 “pnt”) and the Muslim variety in Turkey as its dialect. Whether Romeyka forms based on its structural characteristics an own branch within the Asia Minor Greek language family cannot be decided here (but see Janse 2002 in favour of considering PG as a language rather than a Greek dialect; also cf. Drettas 1997). When Romeyka is referred to as a “language” throughout this thesis, this refers to its status as indigenous minority language in present-day Turkey.

---

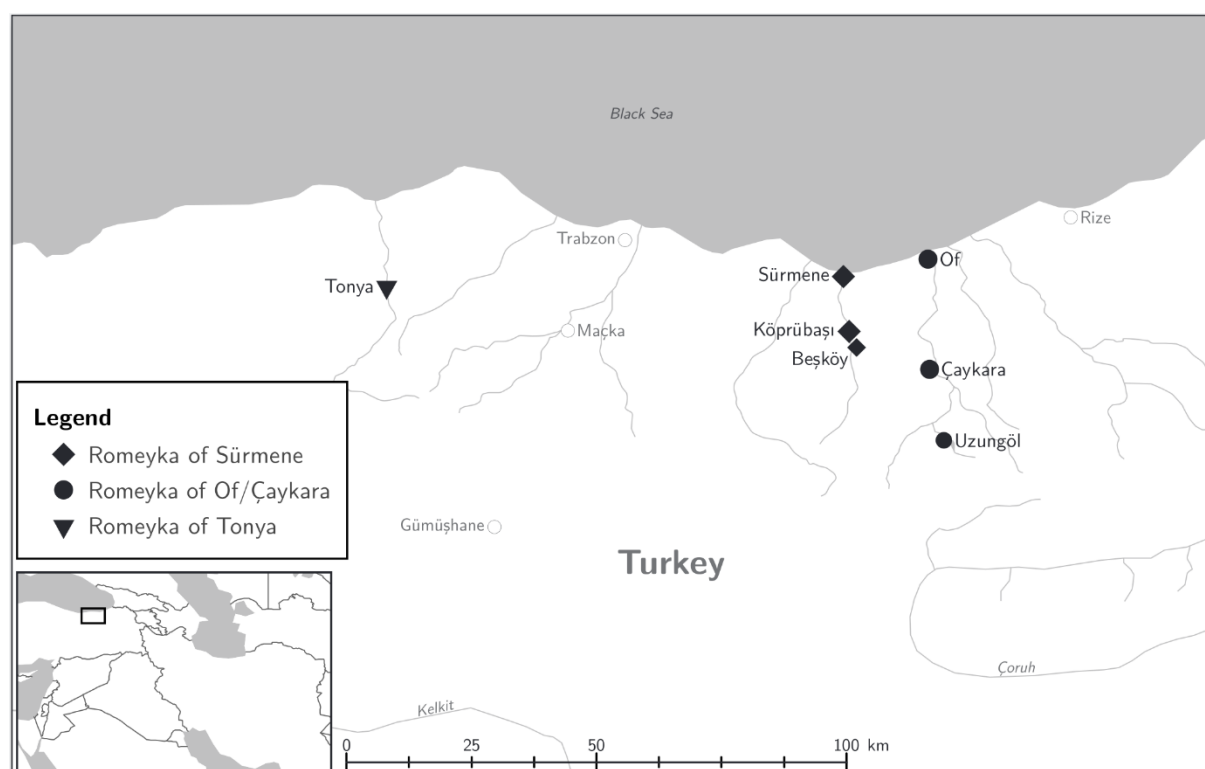
<sup>3</sup> More specifically, it might be considered to form part of a tentative “East Asia Minor Greek” branch (a notion going back to Dawkins, which has been also challenged, though), since other AMG varieties have been recently added to the west (see Janse 2020: 201–203).

### 1.2.2 Dialectal variation

Romeyka is spoken in a mountainous region at the Southern Black Sea coast in a stretch of land reaching from the mountains to the shore. The mountain range of the Pontic Alps reaches some 4,000 metres in altitude, and the rugged and poorly accessible terrain with close-knit communities and generally preserved traditional ways of life (at least until the end of the 20<sup>th</sup> c.) is likely a factor in determining regional variation and preservation of archaisms (Özkan 2013). Due to the semi-pastoral lifestyle in the villages and on the mountain pastures (Tr. *yayla*), a rich lifestyle-specific vocabulary in Romeyka remains, including terms for botany, work practices, and products (Schreiber & Sitaridou 2017).

Romeyka arguably consists of three sub-dialects which are mutually intelligible but differ in terms of phonology and lexicon as well as according to some morphosyntactic features (see Mackridge 1987). Figure 1 presents the three major sub-dialectal areas: Romeyka of Of (in previous research also referred to as “Ophitic”, henceforth RO<sub>Of</sub>; it is not clear whether Romeyka is still spoken in the city of Of nowadays, cf. Sitaridou 2014b: 32), including Romeyka of Çaykara (Katoxóri) which in turn includes the variety of Uzungöl (Saráchos), which had been described by Mackridge (1987; for a map of the Romeyka-speaking villages in the Çaykara administrative district, see Sağlam 2017: 93, Figure 1); Romeyka of Sürmene (Sürmena) (henceforth RSür); and Romeyka of Tonya. Several speakers also report another sub-dialect of Romeyka around Maçka (see also Brendemoen 2002: 32) which reportedly evolved due to immigration from the Sürmene area after a flood in 1932.

RO<sub>Of</sub> is argued to be the most archaic variety of Romeyka (Mackridge 1987), although also within each sub-dialect language competence differs between villages: “Especially in comparison to the Ogene variant, inhabitants of these villages in the lower sections of the Valley claim to speak a less archaic version of Romeyka with less fluency and a more restricted vocabulary.” (Sağlam 2017: 95). For the sub-dialect of Romeyka which is in the focus of the present study, see Section 1.4 (point 4).



**Figure 1:** Approximate locations of the dialect groupings of Romeyka (Schreiber 2018: 896, Fig. 1)

### 1.2.3 Language contact

Romeyka has been in contact with Turkish probably at least since the Islamization waves in the 17<sup>th</sup> c. (Brendemoen 2002; cf. Drettas 1997: 5–6, reporting contact with Turkish since the end of the 11<sup>th</sup> c.) but increasingly intense and facilitating language shift since internal migration started in the 1960s. Arabic elements were introduced to Romeyka via Ottoman Turkish. Intense language contact between Romeyka and Turkish (see Mackridge 1987; Brendemoen 2006) resulted in both lexical and grammatical borrowing and syntactic calquing (Sitaridou 2013).<sup>4</sup> Romeyka is also argued to have been in contact with Armenian Hemshin/Homshetsma, South-Kartvelian languages such as Georgian, and notably areal Laz, which is suspected of having exerted a substrate influence on Romeyka (Drettas 1997: 5–6; Brendemoen 2002; Sitaridou 2013). Furthermore, Kurdish guest workers must have stayed in the area at several points in history (Sitaridou 2013) though it is not clear whether this had any influence. Importantly, the vast majority of Romeyka speakers has not been in contact with any variety of Greek; an exception forming very few dedicated speakers of Romeyka who learned SMG and/or studied its diachrony.<sup>5</sup>

In terms of typological structures, Romeyka is located in an interesting contact setting between the Indo-European language family, namely its representative Greek, but also neighbouring Western Armenian Hemshin, and other Western Asian language types, namely the Turkic language family and Kartvelian. Several structural features clash here with word order in the syntactic domain being an especially noteworthy feature (Haig 2017; Haig, Noorlander & Schiborr, under review). Whether the location of Romeyka in this contact zone leads to specific grammatical patterns is a secondary issue that requires more scrutiny (Schreiber & Janse, in preparation).

### 1.3 State of research

There exist some early grammar(-fragments) focussing on Pontic Greek in Pontus, in particular Christian Pontic Greek (Deffner 1878; Parcharidis 1880, 1888; Dawkins 1931, 1937). However, as Romeyka is expected to have diverged from Christian Pontic Greek already before 1923 (Sitaridou 2014b), these descriptions cannot be uncritically adopted as a baseline for understanding the historical development of Romeyka.<sup>6</sup> Subsequent work has focused on PG as spoken in Greece after 1923 (Papadopoulos 1933, 1955, 1958; Tombaidis 1992, 1996; Drettas 1999; i.a.). A more recent comprehensive grammar of PG was presented by Drettas (1997). The most recent grammar of PG (Ophitic as spoken in Greece after 1923) has been provided in Greek language by Revithiadou & Spyropoulos (2012). We still lack a thorough typological description of many domains of the grammar of Romeyka – let alone a general grammatical description (but see Schreiber 2018) – although previous publications aimed at summarizing the features of Romeyka that differ from SMG or other PG dialects (Mackridge 1987, 1995, 1999; Bortone 2009; Özkan 2013) and during the last years, a considerable body of research has been carried out on the syntax of Romeyka (see Sitaridou 2014a, 2014b, et pass.; Neocleous 2020, 2022). The present thesis is the first attempt at providing a grammar of Romeyka as still spoken in present-day Turkey, although under considerable threat of language shift and ultimately, extinction (see Schreiber & Sitaridou 2017 on the vitality of Romeyka).

<sup>4</sup> For an investigation of mutual influences of Pontic Greek in Pontus and regional Turkish varieties, especially with regard to the phonological domain, see Brendemoen (2002).

<sup>5</sup> Importantly, linguistic data from those speakers being competent in SMG are either not included in the present study or are treated with extreme caution in order not to represent any form of levelling or purificational intervention.

<sup>6</sup> For this reason, the earlier work on (Christian) Pontic Greek, both prior to 1923 in Turkey and after 1923 in Greece, is considered in this thesis primarily as complementary evidence for specific issues rather than being generally included.



## 1.4 Aims of the study & organisation

In the light of the general lack of a basic grammatical description of Romeyka (or Muslim Pontic Greek) as spoken in Turkey and the advanced status of language endangerment and already progressing language shift to Turkish in many communities, it is the main aim of this study to provide a grammatical sketch of Romeyka reflecting the multilingual reality of the speakers and long-term language contact with Turkish; for some secondary theoretical aims of this thesis, see below in this Section. Crucially, it is believed that in order to be able to answer any theoretical questions, a profound description of the linguistic structures of Romeyka is needed as an initial step. Therefore, it is the primordial aim of this thesis to provide this grammatical account (i) as a documentation of an endangered indigenous variety, (ii) as a basis for further theoretical (contact-)linguistic reasoning, as well as (iii) other dialectological research.

In order to be able to depict the typological profile of Romeyka, as well as the degree of language contact and language shift, it is necessary to build on a sound empirical foundation, backed up by authentic and accountable corpus data. Despite extensive research on related varieties, there is still a paucity of such data on Romeyka. An initial major aim of this thesis was therefore to build up a corpus of language usage (Schreiber, in preparation), implementing the principles of language documentation (as outlined for example in Gippert, Himmelmann & Mosel 2006), which serves as the basis for a descriptive grammar of the language.

Within the aim of providing a grammatical description of Romeyka, there are certain aspects and obstacles that need to be taken into consideration:

1. It is the aim of this grammar to reflect a picture of Romeyka as spoken at the moment (i.e., in the 2010s and 2020s) in North-eastern Turkey in a certain dialectal area (see 4. below). The grammatical material of the present description has been provided in a field work setting by a limited number of bilingual speakers (for the methodology and its limitations see Section 1.5.3). It is assumed that any manipulation of these factors may lead to different results: Romeyka exhibits significant dialectal (see 2.) and idiolectal (see 3.) microvariation.
2. Romeyka shows a very high amount of micro-variation which is strongly dialectologically determined but also idiolectally. In addition, Romeyka displays a not insignificant amount of intra-speaker variation whereby the causation is either not clear or probably even ‘random’ in a sense that it is subject to the creativity of the speaker which he or she may exhibit as part of reaching the particular communicative goal of an utterance (see also Section 6.1). “Idiolectal” is used here as a general term summarizing linguistics variation which has arisen for example due to individual language acquisition patterns, migration patterns, marriage and other social factors. Micro-variation occurs at different parts of grammar and the lexicon, involving some striking features such as for example the retention of unstressed neuter noun ending going back to *-ion*, which is clearly dialectologically determined but seems also subject to intra-speaker variation. Another prominent example is the existence of the genuine infinitive only in some varieties with patterning differing from village to village, even of those villages that belong to the same municipality. The factors leading to morphosyntactic differences even between villages of the same administrative district are subject to further investigation; this subject could be related to the question whether Romeyka is spoken at all in a village today which differs from village to village. In general, language variation in Romeyka (of all sorts) is a topic certainly awaiting further in-depth research. The example in (1) depicts some micro-variation in responses of different speakers to the same translation task targeting predicative adjectives.

- (1)
    - a. *havu d=ospidi bola trano* (C1)
    - b. *au d=ospid trano* (B1)
    - c. *havu t=ospitin trano* (H3)
    - d. *avu t=ospidin tranon* (H2)
    - e. *havu t=ospidin bola tranon eni* (H1)

‘This house is big.’
3. In observing idiolectal language variation, and individual language competence in general, sociolinguistic and individual social factors of the speaker seem to play a crucial role. As has been argued in Schreiber (2016), individual language attitudes and the identity function of Romeyka correlate with language competence; the age of the speaker being a significant variable. In the present work, it is suggested that the influence of these sociolinguistic factors on language competence – as well as that of acquisitional factors – should be enriched by the pragmatic function Romeyka has for an individual speaker. In order to evaluate language production of an individual speaker, it should be considered in which functional domains and for which communicative purposes the speaker uses Romeyka in daily life: it seems that the language use patterns differ strongly between speakers and result in different language competence, even in comparable acquisitional settings.
4. The present work describes the grammatical properties of Romeyka as spoken today in Trabzon province (Tr. *Trabzon il*), namely the sub-variety Romeyka of Of (ROf). ‘Romeyka of Of’ is defined here as the variety of Romeyka spoken in the communities of the Solaklı river basin including the three administrative districts (Tr. *ilçe* ‘district’) of Of, Dernekpazarı and Çaykara. Since the majority of speakers from the present corpus stem from the administrative district of Çaykara, the focus is laid in the present work on ROf as spoken in the district (Tr. *ilçe*) of Çaykara. This is by no means restricted to the administrative city of Çaykara (Tr. *merkez* ‘center’) but in fact, Romeyka is spoken today probably more often in the different villages belonging to the administrative district of Çaykara. This comprises 32 different villages; for reasons pointed out with regard to the micro- and nano-variation in Romeyka under 2./3., the grammatical structures found in the present corpus should better not be generalized for all villages of the district of Çaykara, nor for the whole variety Romeyka of Of comprising the three districts of Çaykara, Dernekpazarı and Of. Therefore, whenever a micro-variation seems to be limited to (or could only be proven for) a certain village, this information is added in the grammatical description. In this case, the most accurate Turkish (and if available also the Romeyka) name of the respective location is given irrespective of the municipality (Tr. *belediye*), i.e., for example, the label Karaçam (Ótšena) refers to the village of Karaçam (Ótšena) rather to the municipality (Tr. *belediye*). For an overview of the home villages of the speakers of the present corpus, see Table A.1 in Appendix A. Note that apart from the majority of speakers from the district of Çaykara, comparison data are also available from a speaker from the district of Dernekpazarı, a speaker of Romeyka of Sürmene, which is spoken in the district of Sürmene in the neighbouring Manahoz valley, and three heritage speakers of Romeyka from Çaykara district but now living in Germany (for a detailed description of the methodology, see Section 1.5).
5. Despite the sub-dialectal focus of the present work, it is the intention of the present study to provide a (to a certain point comprehensive) reference for readers who are interested in the structures of present day Romeyka in Turkey. For this reason, reference is made to existing literature wherever possible (with an eye on feasibility and practicability), not only in relation to the Romeyka features found in the present corpus data but also as a report about features found in other varieties of Romeyka or that do for any reason not figure (extensively) in the present data (e.g., the infinitive). In

addition, the present work should not only be a resource for further scholarly research but should be also a meaningful reflection of Romeyka as being presently spoken in Turkey relevant for any interested non-linguistic readers; therefore, it has been chosen to present numerous linguistic examples where available (often two or three examples of one phenomenon), to allow for a profounder insight into the language beyond the described phenomena.

Apart from the primary aim to provide a grammatical description of Romeyka as spoken in present-day Turkey, certain contact-linguistic issues sparked my interest: When preparing the attitudinal study of Romeyka's language vitality (Schreiber 2016), I came across the statement of a respondent who stated that Romeyka is not a full language anymore due to intense mixture with Turkish (*Rumca Türkçe'yle karışmış bir dil* 'Rumca is a language that is mixed with Turkish', speaker M29 in Schreiber 2016). This is a negative attitude not uncommon among speakers of a minority language, among other negative attitudes such as "the language has no grammar" or "it is (structurally) not as beautiful as X (i.e., the majority language)". These kinds of negative attitudes arise under comparison with the standardized majority language – in European-model nation states often the only national language –, its functional spread and literary tradition. Although these negative attitudes are often held by speakers of minoritized languages, for example as outcome of negative attitudes of the majority society towards minority languages, and do by no means correlate with the linguistic reality, these negative attitudes towards the grammatical system of Romeyka sparked my interest: From a theoretical linguistic point of view, language contact can indeed lead to intense re-shaping of a language's typological structures (for example, as described by the concept of "metatypy" in contact linguistic literature) and – in a language shift scenario – gradual decrease in language use and functional domains can indeed result in a change in linguistic structures, often associated with simplification and loss (cf. Chapter 6).

Therefore, it is a secondary theoretical aim of the present study to assess the grammatical composition of Romeyka and to disentangle the nature of the factors that have contributed to its present shape, broadly considering the respective impacts of three types: (I) influence from the main contact language Turkish (involving different ways); (II) changes due to general factors operative in this kind of minority setting, for example language shift/ attrition / imperfect acquisition – in principle independent of the nature of the contact language; (III) the continuation of inherited processes of language change that have determined the history of the Greek language which can be identified through considering comparative evidence from related varieties of Greek, and the attested material from earlier stages of Greek. It has to be mentioned, though, that this latter aspect could be only qualified as tentative here, since the present research approach is not primarily from a historical-comparative perspective; nevertheless, it is hoped that the present data will inform the ongoing discussion on these issues. Finally, since Romeyka is embedded in a "transition zone" between different linguistic areas (see Haig 2017), different typological features – especially in the (morpho-)syntactic domain – clash here; if establishing contact relations between Romeyka and Turkish, the larger areal view may possibly also contribute to the assessment of the contact outcomes.

The underlying theoretical interest that inspired the grammatical description in this thesis can be subsumed under the following three research questions:

- I. How strong is the grammatical influence of Turkish on Romeyka?
- II. Are language shift and attrition impacting upon the grammatical structures of Romeyka?

These two research questions bring along with them a third question, namely: (III.) How much of the grammatical structures and "linguistic material" of Romeyka is Greek? As part of this question, one could ask: Which language internal changes set Romeyka apart from other

Modern Greek dialects? While the present work aims to provide a first tentative answer to the first two research questions based mainly upon comparison with Standard and regional Turkish and based on the theoretical contact linguistic literature and their findings about “vulnerable grammatical features” in language contact, the third question is – although in many respects highly interesting – clearly beyond the focus of the present study and remains subject for further (Greek dialectological and diachronic) research.

Since the aim to build a representative, naturalistic, and linguistically annotated spoken language corpus and to provide a grammatical description of the structural system of Romeyka is by no means a trivial task, the majority of this work is clearly focused upon the grammatical structures (Chapters 2–5). Chapter 2 presents the phonology of Romeyka including the most important phonological processes that set it apart from other modern Greek dialects. Chapter 3 presents an overview of the word classes in Romeyka; a focus is laid on the description of individual function words such as pronouns, particles, etc. Chapter 4 provides a general description of nominal and verbal morphology as well as some principles of word formation. Chapter 5 focuses on the syntax of independent and complex clauses. Some tentative answers to the two research questions stated above that are based upon the grammatical description in the focus of this study are given in Chapter 6: the first question (I.) will be dealt with in Sections 6.1, 6.2 and indirectly in 6.3; the second question (II.) will be targeted in Section 6.4. It has to be thereby born in mind that if any explanations are provided here for the development of certain structural features, they can at the present stage only be tentative; so, features listed in Chapter 6 need to be considered candidates for further in-depth research. Considerable work on related varieties exists that shows how intermingled the factors in language change and contact are (Janse 2009; Karatsareas 2011, 2013, 2014, 2016; Sitaridou 2016; Michelioudakis & Sitaridou 2020; Neocleous & Sitaridou 2022; i.a.) and which proves what a careful and comprehensive procedure is required in order to be able to make substantially grounded claims. This kind of in-depth research cannot be provided at the present stage, so the answers to the research questions given in Chapter 6 remain tentative. It is the primordial aim of this dissertation to provide a solid basis for any further theoretical research.

## 1.5 Methodology

The aim of the present section is to provide an overview over the process of data collection by means of linguistic fieldwork in Turkey, the steps in data processing and corpus building, together with a justification of the method applied and some critical remarks on the methodology as well as pending steps in making the present work a significant contribution to language documentation by means of publication and archiving of the data.

In order to be able to capture a realistic impression on how Romeyka is spoken in present-day Turkey, linguistic fieldwork is the optimal method to adopt for data collection. However, since the opportunities to collect first-hand data in field research in Turkey for the present study were not extensive (see Section 1.5.1.1), the data that have been gathered for the present study are complemented in particular by the following two resources (for other works from which linguistic data are cited see the references given in the respective linguistic examples and the reference list):

- a. A very informative and frequently consulted reference is the Romeyka dictionary by Tursun (2019), which includes data from different dialects of Romeyka. Furthermore, it does not only provide information on the lexical items themselves but also presents example clauses for the lemmata which are a welcome source of grammatical information. For selected lexical items, grammatical paradigms are provided as well. However, as many of the lexical items presented in Tursun’s comprehensive dictionary do not figure in the Romeyka corpus

– or figure differently there – lexical items that only occur in the dictionary are marked (TD) throughout this thesis.

- b. Linguistic data and field notes of Dr Hakan Özkan from Romeyka of Sürmene (RSür) have been occasionally used for the purpose of comparison. His data have been especially integrated in the Schreiber (2018) grammatical sketch of Romeyka, which forms both in structure and in content the basis for the present work. Data that stem from Dr Özkan's fieldwork are referenced in the present work as (Özkan, n.d.).

### 1.5.1 Linguistic fieldwork

#### 1.5.1.1 The speakers

The data for the present study have been in the main gathered during a one-week field trip to the area, accompanied by a community member who presently lives in Germany, in July 2019. However, in general research on the lexicon and grammar of Romeyka dates already back to 2013 and a single Istanbulite native speaker has been recorded repeatedly between 2013 and 2019.<sup>7</sup> The speakers from the communities *in situ* have been approached via acquaintances of the accompanying community member. As a general fact, it needs to be mentioned that it is difficult for non-community members to find speakers and to build the necessary trust for speakers to engage in linguistic research. One reason for this are probably the very close-knit social networks in a geographically remote area in which strangers immediately stand out. Furthermore, the history that has seen the expulsion of the Christian Pontic Greeks from the area might have an influence. In addition, official negative attitudes toward linguistic minorities in Turkey, especially in the second half of the 20<sup>th</sup> c., have contributed to a negative view of many speakers toward their own language as an element that sets them apart from the Turkish mainstream (see Schreiber 2016) which is likely to contribute also to a certain suspicion towards community outsiders paying interest to the Romeyka. Due to these difficult circumstances, the present set of speakers is to a large degree a convenience sample. The corpus consists by large of the texts (spoken language) of eight speakers; five females and three males (for an overview of the speakers, see Table A.1 in Appendix A). The reasons for the higher number of female speakers are that (a) due to the gender of the researcher, access to the same gender is easier in a still traditional Muslim community, and (b) women (especially of the older generations) are more accessible since the domain of life is still traditionally the house. Speakers of two age groups could be found: the grandparent generation, approximately around 70–80 years old, and the parent generation, approximately 45–55 years old. In general, these two age groups are probably those with the highest language competence; speakers of the grandparent generation being still often L1 speakers of Romeyka, whereby Turkish is often the L1 and Romeyka the L2 for the younger generations (for an attitudinal study of self-assessed language competence and bilingual profiles of speakers that differ considerably according to the speech community examined, see Schreiber & Sitaridou 2017). For comparative reasons, fieldwork material has been also collected from three heritage speakers<sup>8</sup> of Romeyka in Germany. Due to their

<sup>7</sup> Oral data have been gathered from an elderly native speaker in the Istanbulite speech community during a seven-months research stay in 2013, a subsequent one-month research fellowship in early 2014, and during field trips to either (or both) Istanbul and Çaykara in Summer 2014, February 2015, February 2016, in April and December 2017, and in July 2019.

<sup>8</sup> The term “heritage speaker” is used here to differentiate between the Romeyka and Turkish bilingual speakers living in Turkey, either in the Trabzon district or in Istanbul, and the Romeyka – Turkish – German trilingual speakers living in Germany. Therefore, the term “heritage speaker” is not used here in line with the general definition, for example as “[i]ndividuals who were raised in homes where a language other than the dominant community language was spoken, resulting in some degree of bilingualism in the heritage language and the

trilingual profile and as the data consists predominantly of elicited questionnaires, linguistic examples from this source are marked throughout the thesis by a speaker code (H1), (H2), (H3), and not by the unique identifier of the recording (e.g., 03\_21042018M). Furthermore, in order to distinguish between questionnaire data gained by means of a translation task from Turkish (see Section 1.5.3) and naturalistic texts, questionnaire data are throughout the thesis referenced by a letter and a number, i.e., A1, B1 and C1, instead of the unique recording identifier used to refer to a segment of naturalistic speech in the Romeyka corpus. The three codes correspond to the following speakers and recordings: A1= Speaker 9 in Table A.1 in Appendix A, corresponds to the recording numbers 55–57 in Table A.2 in Appendix A; B1= Speaker 2 in Table A.1 in Appendix A, corresponds to the recording numbers 39–40 in Table A.2 in Appendix A; C1= Speaker 3 in Table A.1 in Appendix A, corresponds to the recording numbers 44–46 in Table A.2 in Appendix A. In addition, H1= Speaker 10 in Table A.1 in Appendix A, corresponds to the recording number 52 in Table A.2 in Appendix A; H2= Speaker 11 in Table A.1 in Appendix A, corresponds to the recording number 51 in Table A.2 in Appendix A; H3= Speaker 12 in Table A.1 in Appendix A, corresponds to the recording number 50 in Table A.2 in Appendix A. Finally, additional recorded language data that have been collected by a community member have been considered in the Romeyka corpus.

Importantly, it needs to be mentioned that due to the difficulties in finding competent speakers and the restricted time during the fieldtrip of 2019, not all field materials (see Section 1.5.1.2) could be collected from all speakers. According to their availability, interest and competence, only selected materials have been used with particular speakers. A full set of materials only exists from the Istanbulite native speaker (speaker number 9 in Table A.1 in Appendix A).

A negative outcome of these limitations is that the individual multilingual profiles of speakers regarding their language acquisition, language use and biography could have not been monitored as closely as it is considered to be relevant for a synchronic study on language contact (for a discussion of methodological limitations of the present study, see Section 1.5.3). For example, a detailed monitoring of the full multilingual repertoire of a speaker together with his or her migration biography is relevant to obtain knowledge whether the speaker is competent in Standard Modern Greek or has knowledge about Ancient Greek. Competence in SMG has often proven to influence the Romeyka data produced by that speaker; in most cases not in first instance via levelling but often enough by the intention to fill gaps in the Romeyka lexicon or replace Turkish loans and calques by Modern Greek material and structure. With the aim to gather a record of Romeyka as currently spoken in Turkey, competence in Standard Modern Greek is not the default case for the majority of speakers, so data from bilingual speakers of Romeyka and Turkish (in either order of acquisition) are preferred in this study. However, it seems that different acquisitional profiles lead to different competences in Romeyka, so a monitoring of these factors is considered highly important. In addition, gaining information on basic biographical data such as place of birth and places of residence (with duration of the residence) are obviously important for the identification of the Romeyka dialect of a speaker.

### 1.5.1.2 Fieldwork materials

The present study is based upon different kinds of linguistic data that have been collected by means of the direct approach to fieldwork via sound recording of field sessions. The linguistic data mainly consist of: (i) questionnaire data including word lists; (ii) elicited naturalistic texts, that is, oral narratives and conversations between two speakers. It is the primary goal to gather naturalistic data as primary source to depict as natural language as possible in a fieldwork

---

dominant language” (Scontras, Fuchs & Polinsky 2015: 3), since some of the bilingual Romeyka speakers in Turkey could also count as “heritage speakers”, although the term is in present research predominantly used in migration settings rather than in indigenous language settings.

setting (see also Section 1.5.3), complemented by questionnaires and paradigm elicitations for a better coverage of information on the basic grammatical structures and the lexicon. For example, certain complex linguistic structures such as relative clauses or passives do not figure extensively in free speech, nor do full nominal or verbal inflectional paradigms.

The main body of data which were gathered during the field trip to Çaykara in 2019 and which form the core part of the Romeyka corpus (Section 1.5.2) have been elicited by means of the following field materials:

- i. a comprehensive morphosyntactic questionnaire covering basic grammatical domains and containing several paradigms
- ii. a 174-lexical item word list
- iii. two picture stories on cards (“Totem Field Storyboards”) to prompt quantifiers (“There are many... and more bears!”, Chen 2015) and conditional clauses (“The wood chopper”, TFS Working Group 2011) which have been selected culture-specifically
- iv. a list of questions to prompt free monologues for example about descriptions (recipes, directions, e.g., ‘How do you make butter?’) or narratives (happy, dangerous, etc. events, e.g., ‘What happened at the best wedding you ever went to?’)
- v. a prompt to initiate a naturalistic conversation between two speakers about daily life topics
- vi. supplementary materials such as a cultural-sensitive photo book featuring daily life activities in a rural environment to elicit descriptions; a colour sheet to elicit colour terms

The morphosyntactic questionnaire and the word list are based on the morphosyntactic questionnaire and wordlist developed for the Atlas of the Languages of Iran (ALI) Project (<https://carleton.ca/iran/questionnaires/>, accessed 17 August 2022). Both the questionnaire and the wordlist have been supplemented and adapted to the cultural situation of Romeyka. They have been piloted in an initial version in 2017 with the Istanbulite native speaker and in an improved version in 2018 with the heritage speakers from Germany (see Section 1.5.1.1). A finally amended version has been prepared for the fieldwork in Çaykara in 2019.

The field work materials furthermore comprise a personal information sheet in Turkish comprising questions on the social (gender, age, place of birth/youth/residence and the length of stay, education, profession) and sociolinguistic (that is, the multilingual profile including order of language acquisition, age of onset of acquisition, self-reported language competence in all languages, and frequency of language use) variables of the consultants. However, due to the complex circumstances during the field trip to Çaykara in 2019, these comprehensive speaker data have been only obtained from the Istanbulite speaker and the heritage speakers in Germany. In addition, to trace language shift between generations, monitoring of intergenerational variation in Romeyka would ideally include data from speakers of different generations within the same family; again, such an ‘ideal’ situation was only found in the German heritage speakers where data could be gathered from a father and his two sons together with a close monitoring of their multilingual profiles (see also Section 1.5.3).

Finally, the field work materials are completed by a consent sheet in Turkish informing the speakers about the aims of the research as well as their rights in contributing their speech data. The identity of the speakers is protected by a fully anonymized recording catalogue assigning a code to each speaker; only the following speaker information are revealed: place of residence and/or birth, age, gender (see Table A.1 in Appendix A).

### 1.5.1.3 Fieldwork process

The present data have been gathered either from a speaker with whom the researcher had a close personal relationship or in the presence of a female (semi-)speaker from the speech community who conducted the elicitation and conversed with the speakers, while the researcher was just present to assist with the elicitation materials and made the recordings. The speakers were approached solely via personal connections, so the sample is not representative (see Section 1.5.1.1). Individual fieldwork sessions started with a very brief explanation of the purpose of the research. Consequently, the speakers were prompted in Turkish by means of the field materials (Section 1.5.1.2; for critical notes on this methodology, see Section 1.5.3). The full session was recorded (sound recording).

## 1.5.2 Data processing

### 1.5.2.1 Corpus building

The data that have been gathered by the methodology described in Section 1.5.1 are sound recorded naturalistic and elicited spoken texts. Meta data indicating information about the field session (place, time, researcher and speaker, materials used, topics covered, additional persons present, etc.) are available for all sessions. After recording, the primary data have been catalogued according to the date of elicitation (see Table A.2 in Appendix A for the recording catalogue of the present Romeyka corpus). Nearly all available data out of a total corpus size of approximately five hours of speech recording are fully transcribed in a broad phonemic transcription in the annotation software ELAN, which is common in linguistic fieldwork and language documentation (for a justification of the type of transcription see Section 1.5.2.2 below). A few of the primary data are only partially transcribed, this concerns especially elicitations of word lists which include apart from the list of lexical items also some (at times lengthy) explanations. Out of a total of five hours of transcribed sound recording including all materials (elicited questionnaires, free narratives/descriptions, conversations), naturalistic texts of high quality, including monologues and dialogues, have been put together to form the Romeyka corpus of free naturalistic data. The texts that form part of the corpus have been selected based on the type of text, i.e., the elicitation method (data from the storytelling picture task were preferred over descriptions from the photo book; narratives were preferred over enumerations and explanatory comments) and the length of the recording. Transcribed (and translated) questionnaire data were kept apart as additional materials. Based on this procedure, a Romeyka corpus of free-spoken language data of approximately two hours of speech recording was composed. This corpus is fully transcribed, translated into Turkish and English and completely morphologically glossed in ELAN. This corpus of naturalistic spoken texts is referred to in the grammatical description below as “the present Romeyka corpus”. The Romeyka corpus comprises approximately 22,000 annotated tokens, which makes it a smaller corpus at the minimum level of the size corpora in language description usually have (for obstacles in the fieldwork process and data collection that causes the small corpus size, see Section 1.5.1 above).

The collected language data were after transcription translated into Turkish by the (semi-)speaker with good comprehension skills in Romeyka who also elicited the data and consequently translated into English by the researcher under consideration of both the Turkish translation and the Romeyka data. Due to this procedure, in the corpus in ELAN the English translation may differ sometimes from the Turkish translation (see also Section 1.5.2.2). The part-of-speech tagging, and full morphological annotation was carried out by the researcher. Due to this methodology of data processing, the present Romeyka corpus contains some

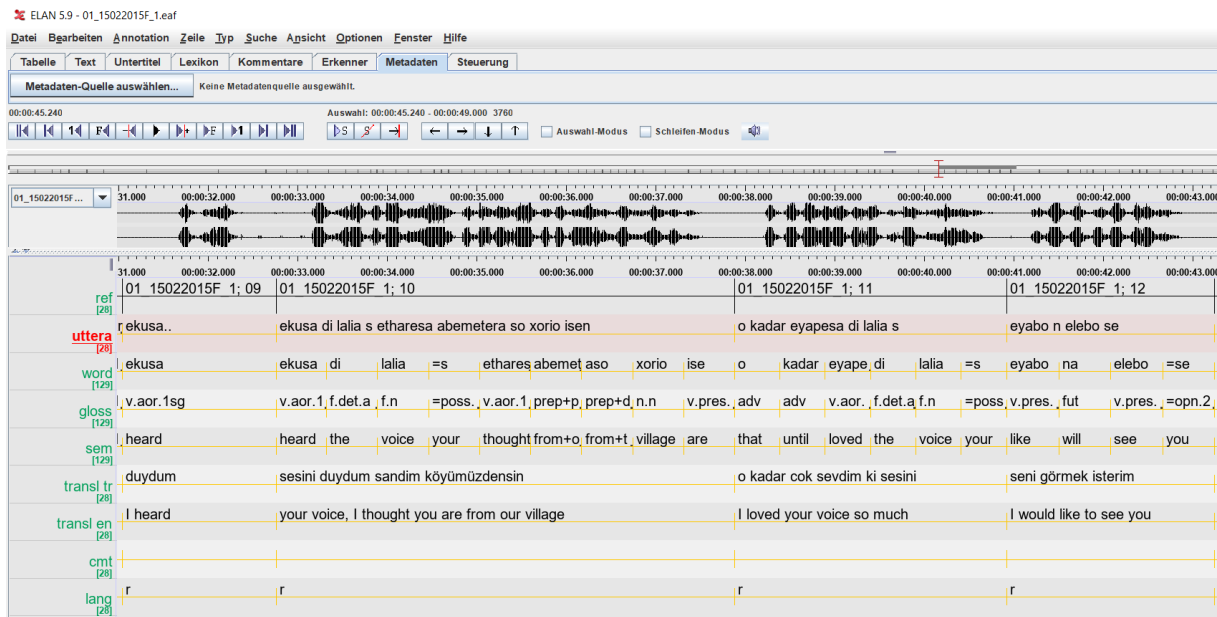


particularities based on initial decisions that have been made which are outlined in the following Section 1.5.2.2.

### 1.5.2.2 Corpus design

The Romeyka corpus in ELAN is based on the audio recording transcribed and annotated by means of the following tiers (see also Figure 2):

/ref	unique referential identifier for each segment in a session; the reference code is based on the date of the recording, the session number, and an indication of the gender of the respondent
/uttera	the segmented string of utterance as transcribed from the sound file
/word	partially normalized representation of individual phonological words
/gloss	part-of-speech glossing together with grammatical information like person, number, gender, case, etc.
/sem	semantic English equivalent of the word
/transl tr	Turkish translation provided per segment by a semi-speaker
/transl en	English translation by the researcher under consideration of the Turkish translation
/cmt	optional comments on grammatical phenomena, speakers, uncertainties, etc.
/lang	indication of the dominant language in a segment (r= Romeyka; t= Turkish)



**Figure 2:** Annotation tiers of the Romeyka corpus in ELAN

The utterance tier provides a broad phonemic transcription of the sound file. This methodology has been adopted for the sake of convenience to be able to provide transcriptions for the whole corpus (i.e., five hours of sound recording, and not only the selected two hours of naturalistic data that form the present Romeyka corpus used for this thesis), which allows a larger amount of linguistic data to be edited for analysis. However, the broad phonemic transcription may be not detailed enough when it comes to the phonological analysis; here, a narrow phonemic transcription in IPA would be helpful to distinguish more properly between different phonemic and allophonic representations. Furthermore, word stress is not indicated in the transcription since it was not clear at the beginning whether word stress would be a distinctive feature in Romeyka. However, it has proven that the study of word stress patterns would be indeed worth

pursuing; thus, this remains a desideratum for further steps (see Section 1.5.5). Moreover, it should be noted, that the transcription in the ELAN corpus uses for the sake of convenience digraphic elements for the following phonemes: /ð/ → ð, /ʃ/ → sh, /θ/ → th, /dʒ/ → dž, /tʃ/ → tš; also, /ɣ/ → y, /χ/ → x (for the conventions in transcription of these phonemes in the present grammar, see begin of Chapter 2). The digraphic elements do not occur in the target language as distinct phoneme sequences, so there is no confusion possible.

Finally, occasionally some discourse analytical information are included in the utterance tier: a change in speaker is indicated if an additional speaker other than the main interviewee takes turn (e.g., the interviewer, an accompanying person or a conversation partner in dialogues); in this case, the additional speaker's utterance is indicated in square brackets []. Note that in dialogues, the speakers are also indicated as "Sp1" or "Sp2" in the comment tier. Hesitations, interruptions, or breaks in the utterance are indicated by two periods following the last word in the utterance tier, e.g., *embro..* 'before..' (08\_04072019M\_2; 004).

Importantly, it should be noted that the segmentation of the sound file has been done basically oriented on speech pauses and not based on utterance units. This provided a faster progress in segmentation at the beginning but proved to be less fortunate at a later stage, where an indication of utterance units (for example by the symbol "#") would for example facilitate more fine-grained search options within the corpus. Unfortunately, this early decision is only to be corrected at the present stage with tremendous efforts (see also Section 1.5.5).

In the word tier, elements extracted from the utterance tier may be presented in a normalized form, i.e., for example, without idiolectal variation. This has been at the present stage not been carried out systematically, though, but only in cases where the phonemic transcription differed greatly from the target word, for example in contracted fast speech.

The morphological glossing based on the word tier is separated into two tiers: the gloss tier combines a part-of-speech glossing with morphological information such as tense, gender, person, numbering. This methodology has been preferred above the usual morphological glossing of the form "live.PRES.1SG" as it facilitates easier search options according to part-of-speech type in the corpus, which was considered important for the purpose of this corpus as a resource for grammar writing. For an overview of the abbreviations used in the ELAN corpus, see Table A.3 in Appendix A. The semantics tier presents the meaning of the word in English; note that not the English lemma form is indicated but the inflected English form conveying the very same meaning, e.g., "am" and "is" instead of "be", "left" for the past tense form of "leave". Clitic elements are both in the word tier and in the gloss tier presented as separate words (indicated by "=") to facilitate easier search patterns. Other segments are usually not distinguished. Certain combinations of prepositions and definite articles or pronouns are considered a phonological word, e.g., *abemetero* 'from our' (01\_15022015F\_1; 10); they are indicated in the gloss and semantics tier by "+", i.e., "prep+poss.1pl", "from+our". The semantic tier is left empty, if a target language item has no correspondence in English, like, for example, the Turkish question particle *mi* (glossed as "interr"). Question tags used in the gloss or semantics tiers indicate either an unclear grammatical gloss or meaning.

As indicated in Section 1.5.2.2, the English translation in the corpus may at times differ from the Turkish translation since the latter has been provided by the semi-speaker whereas the English translation has been provided by the researcher. Even if the translation of the researcher differed from that of the (semi-)speaker, the original Turkish translation has been kept as it may provide interesting information on how the speaker parsed the utterance. Note furthermore that the Turkish speaker translated the Romeyka text into Turkish in a way very close to the Romeyka original, leading often to uncommon word order in the Turkish translation. Furthermore, for the sake of convenience, the Turkish translation is typed with the basic Latin script without the language-specific diacritics of the Turkish script. This is a desideratum for a pending revision of the corpus (see Section 1.5.5).

Finally, the language tier has been found helpful in response to a high amount of code-mixing in the recording; it differentiates roughly between matrix languages (in the sense of Myers-Scotton (e.g., 2002) suggesting that the matrix language supplies the inflectional morphology for the clause which is principally visible on the predicate), although it is only partially useable to classify types of code-switching and is rather primarily intended as an indication of the dominant code. For example, Turkish elements that form part of a segment where Romeyka is identified as the matrix language are analysed as Romeyka elements; this may require some additional internal segmentation of words which is otherwise not done.

### 1.5.3 Critical reflection of the methodology

The earlier sections have already highlighted some of the shortcomings of the present method of data gathering which arose due to the very difficult field situation with the struggle to find speakers who are able and willing to participate, the researcher not being a member of the close-knit community (although equipped with a personal relation to the community), and the limited time schedule. In the following, some reflections on the methodology and the quality of data are presented that are relevant for understanding the present data.

- a. Although the field materials used (see Section 1.5.1.2) allow to obtain both free naturalistic data and questionnaire data for a comprehensive coverage of grammatical domains, the fact that the interviews have been carried out in Turkish as the basic means of communication is likely to affect the Romeyka data and in particular, the amount of code-switching and possibly also of lexical borrowings. Especially in a language contact setting with an indigenous language and a society's majority language, the use of the majority language – Turkish in the present case – is likely to affect the minority language data. In an ideal fieldwork situation, the communicative setting should be in the target language, that is, Romeyka. In the present situation, this was not possible since the community member carrying out the interviews were a semi-speaker with only comprehension skills in Romeyka.  
More specifically in a study examining language contact, translation tasks from the contact language need to be treated with extreme caution since they will likely facilitate considerable calquing and borrowing. (Although this data can also reveal interesting insights into how speaker parse Turkish constructions and transfer them into the Romeyka grammatical system.) For this reason, the core of the present Romeyka corpus consists only of free naturalistic texts (monologues and dialogues). Questionnaire data obtained by means of the translation task from Turkish are treated as secondary source and examples in the grammar sketch are indicated as “data from a translation task”. This concerns in particular the questionnaire data from the heritage speakers from Germany.
- b. The data of the present study have been elicited by the direct method of linguistic fieldwork. Although this has been handy in the field work procedure, productive and necessary to obtain comprehensive data, as a second step the collection of more data gathered by an indirect method would be valuable to get a realistic picture of how and when (linguistic functions, domains of use) Romeyka is actually used by the speakers. The approach of more recent documentary linguistics (e.g., Woodbury 2011; see also Dwyer 2006; Rice 2011 for the ethics of linguistic fieldwork) to train community members to record their own data, would be a viable method to gain more natural speech data, which also allow to examine the pragmatical context of language use.
- c. The role of the researcher in the fieldwork process should not be underestimated. Apart from well-known side effects of interview situations like the ‘observer’s paradox’ which happen at the direct approach to fieldwork of any researcher, a researcher who forms not part of the speech community can possibly obscure the data; especially, in sensitive sociolinguistic and political circumstances, trust in the integrity of the researcher is

important to gain naturalistic data. On the other hand, it is of crucial importance to grant protection to the speakers, ensure their legal rights and do as much as possible to prevent negative outcomes of the research. Research in documentary linguistics is likely to have an impact on the community itself and the consequences of this need to be considered. For example, the mere fact that a foreign researcher is interested in the indigenous language may on one hand lead to more positive attitudes of the speakers towards their language (see also Sitaridou 2013) and may possibly evoke a feeling of pride and may on the other hand lead to suspicion and an even stronger rejection of the indigenous language. Especially in the present sociolinguistic situation where the attitudes of speakers towards Romeyka seem to be not too positive (see Schreiber & Sitaridou 2017; Schreiber 2016), the will of the speakers needs to be considered before attempting any steps in language documentation. The idea of community-led documentation endeavours in recent documentary linguistic discourse (Woodbury 2011) is suitable to ensure that the speakers' interests are at the centre of any documentational attempts.

- d. The present data basis is with a 22,000 token-corpus relatively small for a robust grammar sketch. Although the present field materials allowed to cover the basic grammatical categories, it was not possible to obtain data for all linguistic structures and complete paradigms are only available for a restricted number of lexical items. For example, further data need to be elicited to describe causatives and passives in Romeyka. Full paradigms should be presented not only for verbal and nominal inflection but also for adjectival declination and some pronominal categories. Additionally, it would be valuable to prompt suitable speakers with the results of the present analysis to discuss possible alternatives and elicit grammaticality judgements.
- e. As pointed out above, comprehensive monitoring of sociolinguistic and acquisitional variables of speakers is crucial to allocate the speech data appropriately in terms of diatopic variation and language proficiency. A suitable means to test language proficiency (apart from self-assessment of the speakers and consideration of the order of language acquisition) would help to identify speakers whose language competence is representative for their generation. In research focusing on language shift, intergenerational data would be ideal to monitor the transmission process of the language. It is important to note that the present speaker sample is not representative and standardized since (a) the number of men and women is not equal, (b) nor are all age groups represented, (c) speakers from different dialects are included, (d) speakers possess different multilingual repertoires (e.g., competence in Modern Greek or German as L3). While it is for a qualitative study of less relevance that the sample is balanced, it would be crucial to monitor all these variables to be able to allocate the data in the correct way. For example, for female speakers it makes sense to compare the length of stay in their village of birth with that of their residence place (or other places of residence) to correctly estimate their dialect and competence. Since individual biographies play a large role in language competence and use, it is meaningful to control for sociolinguistic and acquisitional variables of the speakers.

#### 1.5.4 Decisions in presentation

The following decisions have been made in the presentation of original linguistic data in the present thesis (also applying largely to the corpus in ELAN):

- (i) As indicated already for the original corpus data in ELAN (Section 1.5.2.2), the examples throughout this thesis are provided in a broad phonemic transcription, e.g., *ineka* 'woman'. This is considered sufficient for morphosyntactic analysis, although a narrower transcription, i.e., [i'nɛka] 'woman', would have allowed more exact representation in the phonology section.

Word stress is only indicated in the Phonology chapter since it is not originally accounted for in the ELAN corpus and adding word stress throughout is unfortunately not feasible at this stage due to the high number of linguistic examples (but see Section 1.5.5).

(ii) The lack of voice distinction in plosives in Romeyka (see Section 2.1.2.1) forms a challenge to the representation of linguistic data from the corpus. Similar to the discussion of citation forms under (iii) below, both the corpus and the linguistic examples throughout this thesis contain still the full variation encountered in the sound recordings as no normalization has taken place yet (as could have been done based on the characterizations of the phonology chapter and the description of common phonological rules; cf. Section 1.5.5). The presentation of linguistic examples throughout this thesis is based on the individual tokens and not on a citation form. This means that linguistic examples in the present study are exactly in the same form as in the ELAN corpus, e.g., ‘woman’ can be represented in the data as *inega*, *ineka* (or *jinega*, *jineka*). However, considering all tokens (i.e., variants of a lexeme) that occur in the corpus, it appears that a certain variant is the most frequent; this is in the present Romeyka corpus predominantly the voiced realization of stops (the presentation of stops as voiced is also adopted by Özkan 2013, while Mackridge 1987 represents stops as voiceless). For the discussion of function word items in Chapter 3, the most frequent form of a word has been selected as citation form, e.g., *ba(l).TOP* and not *pa(l).TOP*, although from the phonological analysis it is evident that /p/ would be the underlying phoneme and that the voiced variant may arise following a phonological rule such as intervocalic voicing; the voicing of initial prevocalic unvoiced stops being another general phenomenon encountered frequently in the recordings. Till the present step it has not been possible to normalize the presentation of stops as voiceless throughout the corpus, although this has the following advantage that the full amount of variation is preserved which provides interesting information on diatopic or idiolectal variation and allows for further phonological research. If normalizing the tokens based on formal assumptions and the assumed etymological history of the Greek word, there is the fear of overgeneralizing certain distributions which do not actually apply to Romeyka and that may lead to a “levelled” picture of Romeyka. An advantage of providing citation forms is obviously more clarity and a more uniform representation, as long the underlying processes are sufficiently described. A disadvantage of keeping the phonological variation is that the corpus contains many doublets, which make type counts difficult. In addition, it may be difficult at times to correctly determine the voice/voiceless distinction for a stop where only a detailed phonetic analysis can reveal whether a stop is more or less voiced. A potential solution to this would be (a) to maintain the linguistic variation in the utterance tier of the corpus in ELAN (ideally even in a narrow transcription in IPA) but to normalize the forms in the word tier. In the thesis text, citation forms should be certainly provided for function items; if a normalized phonological representation is adopted throughout, it should be preceded by a comprehensive phonological discussion of all underlying rules. Since a full description of all phonological processes in Romeyka is beyond the scope of this study, the variation is maintained in the linguistic examples at this stage (see Section 1.5.5).

(iii) The citation forms of nouns, adjectives, pronouns, verbs, and other content words are basically derived from the Greek linguistic tradition, i.e., nouns and adjectives are presented in the sg.masc.nom. form and verbs in the first person singular. However, this general principle is superimposed in this thesis by the attempt to represent the naturalistic language data as accurately as possible in order not to make any premature normalizing generalizations that may cause a loss of actual variation visible in the naturalistic data. Therefore, the following two overarching principles apply: (a) the variation in naturalistic language data is generally preserved and lemma forms are not normalized, if no token is available for this form at the corpus. This applies in particular to nominal (and pronominal) endings, for example in plural

suffixes, e.g., *xorafæ* vs. *xorafe* ‘fields’, but also to apocope of verb endings (2.3.5.2). It is especially striking in the endings of masculine nouns in *-os*: If a masculine noun is attested in the corpus only with an ending in *-o*, the noun will be represented like this, even if from the etymology of the word it is clear that it is originally a noun in *-os*; if examples exist for the same lemma in *-o* and *-os*, the *s* is added in brackets, e.g., *fengo(s)* ‘moon’.<sup>9</sup> Note that since in these masculine nouns in *-os* usually in Greek linguistics *-os* is the masc.nom.sg. form and *-o* or *-on* the masc.acc.sg. form, in Romeyka words ending in *-o*, it cannot be clearly determined whether this form is nominative or accusative. The case is slightly different with neut.nom.sg. nouns which end originally in *-ion*: in this case, the lemma is presented in exactly the form as it has in the original data since the degree of retention of the full ending in *-ion* is a dialectological variable. In order to preserve this variation, final /n/ is not added to the lemma form in *-i*, even if the full form is attested with another speaker, e.g., the word for ‘house’ can be represented in the corpus by the following lemmata *ospi*, *ospit*, *ospiti*, *ospitin*. Masculine pronominal forms are generally presented as lemma ending in *-os*, e.g., *etšino(s)* ‘this’, *ato(s)* ‘he’. Moreover, the principle to maintain as much linguistic variation as possible leads also to the fact that different variants of the same word are maintained also apart from the endings, e.g., *(j)inega* ‘woman’.<sup>10</sup>

The second overarching principle (b) is that all word classes are presented in the least-marked form available, which equals in most cases the nominative singular indefinite form for nouns and adjectives and the first person singular indicative for verb lexemes. This principle leads to the representation of adjectives but also of other word classes such as numerals and quantifier with the shortest and most frequent form as citation form. For example, the adjective *kalos* ‘good’ is represented here by the lemma *kalo*, as this is the most frequent and least marked form in the corpus. Importantly, in other word classes like numerals and quantifier where masculine forms like *enas* ‘one.M’ or *olos* ‘most.M’ are practically not attested in Romeyka, the neuter (and accusative) ending in *-on* is the most frequent variant and is therefore adopted as citation form, i.e., *ena* ‘one’, *olon* ‘whole, entire’, *ulon* ‘all’.

Finally, it needs to be made explicit that in presenting the English translation of a single Romeyka lexeme, not all possible meanings of a lexeme can be given, but just the most prominent/frequent, e.g., *bola* ‘very’, but not *bola* ‘very, many, much’; a full account of the meanings of a lexeme becomes primarily visible in the different contexts in which words are used in the linguistic examples and where the respective relevant translation is indicated, e.g., *batsi* ‘girl’ or ‘daughter’.<sup>11</sup> Despite the wealth of etymological knowledge available for the diachronic development of many words in Romeyka, the meaning of a word is represented here in the form as elicited from a speaker, although this may cover only part of a word’s full historical meaning or even indicate a different meaning as would be commonly expected, e.g., *skoletši*, actually ‘worm’, is apparently used in the present data as a generalization of (some kind of) bug-like insect; also, *furnos* ‘toad’, seems to be used in Romeyka for both ‘toad’ and ‘frog’, equivalent to Tr. *kurbağa* ‘frog/toad’. Further research on semantic change in Romeyka would be certainly worthwhile, a glossary on frequent lexemes in Romeyka including their etymology would be a contribution to this (see Section 1.5.5).

(iv) Loanwords are with regard to their citation form largely treated like native words in (ii); however, in Chapter 2, word stress of loanwords is often left out, if not clearly identifiable without any further scrutiny. In the general grammatical description, loanwords are often addressed separately from native words, although largely the same rules apply to them. It needs

<sup>9</sup> This discussion applies especially to the Phonology Chapter, where the most citation forms are provided.

<sup>10</sup> In Chapter 2, also variants as *róómo* ‘road’ and *ðromo(s)* ‘road’ are presented, although in this case the latter is the citation form and the previous is just a variant corresponding to a token.

<sup>11</sup> Likewise, for reasons of space and readability in in-text examples in the phonology chapter, only the most basic gloss is provided in the translation of function words, e.g., *to* ‘the.N’ instead of *to* ‘the.N.SG.NOM/N/M.SG.ACC’.

to be highlighted, though, that it is to a certain degree difficult to distinguish “loanwords” from “native words” as they form part and parcel of the Romeyka vocabulary. However, it must be noted, that the amount of lexical loans in Romeyka is probably lower than that of Cappadocian, which makes it to a certain degree easier to draw a distinction between both groups, although theoretical research on lexical borrowing in Romeyka is still pending. A preliminary definition of loanwords in the present thesis is provided in Section 3.1.1.1 (for a tentative categorization of loan verbs in Romeyka, see Section 3.1.5)

(v) The cliticness of certain pronominal (and some other) items is debatable within Greek linguistics. With the scope of the present study, the different points of view of the debate could unfortunately not be sufficiently considered (yet; see also Section 1.5.5). In the present thesis, only enclitic pronominal elements of the NP, i.e., weak possessive and object pronouns, are considered clitics (for the reasons outlined in Section 3.2.2.1). These enclitic elements are indicated in the ELAN corpus by “=”. The following other function word categories are also discussed in (Greek) literature as candidates for clitics vs. suffixes: definite articles, the copula *ime* ‘be’ and certain particles like negation particles and potentially also the question particle *mi* that is borrowed from Turkish. Note, that in the linguistic examples throughout this thesis also function words such as the modal particle *na*, negation particle *utš*, coordinating conjunction *tše* ‘and’ and definite articles (including prepositions merged with definite articles) are indicated with “=”, if the form of the functional item is reduced for phonological reasons, e.g., *d=akšem(is)* ‘the evening’.

(vi) All linguistic examples presented without further notice throughout the thesis stem from the present Romeyka corpus (Schreiber, in preparation), which means they are usually from the dialect Romeyka of Of as spoken in Çaykara. Very few examples stem from a different dialect, this is indicated either in the text where reference to the example is made and/or in the brackets before the unique identifier code after the English translation in the linguistic example, e.g., (RSür; 05\_03072019M\_4; 04). If a linguistic example is from another source, this is often indicated in the text and at the end of the linguistic example, e.g., (Özkan, n.d., glosses added). Rarely, examples from Standard Turkish or Standard Modern Greek occur, which are presented without further reference. Occasionally, Turkish translations from the Romeyka corpus are presented alongside the Romeyka utterance; these are indicated in the text and referred to by the unique reference code of the utterance in the corpus.

(vii) The present grammar is roughly based on the first published grammar sketch by Schreiber (2018) which has been broadly extended. Due to this approach whereby the 2018 grammar sketch has so to say “dissolved” in the present text, the original grammar sketch is not extensively cited but only in cases where explicit reference is made to my earlier argumentation. Furthermore, the 2018 publication was primarily intended as a summary of information about the grammar of Romeyka that could be drawn from existing literature. In order to preserve the original source of the information, a piece of information stemming from another author is cited here by the original source reference and not by Schreiber (2018). Finally, large parts of the sociolinguistic background of Romeyka presented in the introduction to this thesis are based on the information presented in the introduction in Schreiber (2018) without explicit reference.

(viii) Throughout the thesis, the following formal decision in representation has been made: Arrowheads are used to indicate causal or temporal consequence. The right arrowhead “>” is used to indicate that the previous form becomes/turns into the latter or that the latter is derived from the first form, e.g., */\*fx/ > /fk/* (Section 2.2.2), *xorévo* ‘dance’ > *exórepsame* ‘we danced’ (Section 2.1.2.6). The left arrowhead “<” is used to indicate that the first form is derived

from/goes back to the latter; this is primarily used with diachronic forms to indicate the root to which a certain word goes back, e.g., *ónema* ‘name’ < AG *ὄνομα* ‘name’ (Section 2.4.1).

### 1.5.5 Pending steps in language description and archiving

As *bon ton* in documentary linguistics, it is intended to make both the present data corpus and the study available for further research and speakers. For publication and archiving of the corpus and to present a solid grammar of Romeyka, certain steps are still pending which could not be realized by now due to the complexity of the undertaking. These pending steps will be briefly summarized in the following.

(a) Steps to be completed before publication and archiving of the data corpus:

The present corpus has been built largely in line with common annotation practices in ELAN corpora from language documentation (see Prince & Nordhoff 2020). However, certain small improvements are in order: (i) it should be considered to indicate word stress in the corpus; (ii) the correct Turkish spelling should be provided for the Turkish translation; (iii) it should be considered whether the semi-speaker’s Turkish translation which is close to the Romeyka original should be standardized, e.g., to comply with correct Turkish word order; (iv) the glossing needs to be checked for certain word classes (see also Table A.3 in Appendix A), especially for past tense verbs (i.e., aorist vs. imperfective), for the modal particle *na* which is so far glossed in the corpus as FUT based on its most frequent function but should be better glossed as PRT.MOD or the like to represent its full functions, the glossing of *ba(l)* as focus particle (FOC) which represents an early stage of analysis needs to be updated and rendered as TOP (=topicalization marker), both the part-of-speech category “noun” and neuter gender are represented by the gloss N, which should be resolved, and, the cliticness status of certain functional items needs to be reconsidered (see Section 1.5.4); (v) if possible, the indication of utterance breaks would facilitate easier searches, although a re-segmentation of the whole corpus based on utterance breaks rather than the pause-oriented segments would probably not be feasible.

Finally, in order to make the corpus publicly available, it needs to be ensured that speaker’s rights are respected. This may entail that although consent was given for the use of the data for scholarly research and publication, a special consent may be sought for long-time archiving with varying degrees of data protection. For example, for religious women of the community it may be considered not appropriate to publicly display their voice. This must be respected and reflected in certain forms of data protection such as limited access rights.

(b) Intended revisions for the publication of a full-fledged grammar:

Due to the complexity of the present undertaking, the following desiderata remain in the presentation of the grammar that need to be addressed for a future publication:

(i) Certain theoretical questions of Greek linguistics need to be re-considered in more depth: at the centre, the debate about cliticness of certain function word categories like definite articles, possessive and object pronouns, the copula *ime* ‘be’ and certain particles like negation particles and potentially the negation particle *mi* that is borrowed from Turkish. Furthermore, the relevance of information structure seems to be much more important than initially expected, so further research is required on different forms of information structure, including the functions of some prominent discourse particles such as *ba(l)*. On the other hand, some sections touching upon critical theoretical issues that lead to far from the scope of the grammar sketch, like the role of veridicality in conditionals, would necessarily remain reduced to the general typological descriptive basics.

(ii) Some gaps in the present description should be filled leading to a more complete coverage of all grammatical domains. For example, further data would need to be elicited for



passives, causatives and reflexives. Furthermore, the elicitation of additional paradigms would be handy to pinpoint the details of nominal inflection in adjectival declension and in nominal case, gender and number inflection of all noun classes. In addition, further data could be gathered about deontic constructions and checking for the existence of a pluperfect tense. Although not crucial, some additional naturalistic data from further speakers from the same dialect area could help to confirm the present findings.

(iii) On the other hand, shortenings of the present discussion are in order where the existing data were described in too much detail in order to maintain clarity and proportionality of the grammar sketch.

(iv) The hierarchical structure of the grammar sketch would benefit from some reorganisation; the organisation of chapters has been set up at the beginning of the descriptive endeavour when it was not yet clear what the discussion of the individual features would reveal. Therefore, for example the Section on verb morphology (Section 4.3) appears to comprise a lot of syntax since it occurred that most of the verb morphology is analytic rather than morphological; topics like negation and non-finiteness that would traditionally belong to a syntax section are therefore dealt with in Chapter 4. In comparison to that, Chapter 5 on syntax deals primordially with clausal syntax.

(v) The following improvements would be desirable with regard to the presentation of linguistic examples in the grammar: First of all, the linguistic glosses that are available for the whole Romeyka corpus in ELAN should be added to all linguistic examples of the grammar. Secondly, it should be considered to normalize the existing variation in the original data, to use citation forms and to unify the voice distinction in plosives in linguistic examples after some general decision has been made. Thirdly, word stress should be ideally indicated in all linguistic examples. In the present thesis, word stress could only be indicated for reasons of feasibility in Chapter 2 on phonology. Although the indication of word stress may not be essential to the description of syntactic phenomena, word stress is relevant for morpho-phonological processes like the inflection of verbs, noun classes, and phonological processes such as vowel deletion in unstressed syllables. At last, even if the morpho-phonological information of word stress might not be relevant in all linguistic examples, the indication of word stress provides the reader with information about the pronunciation of a word. A desideratum in the research of word stress is the patterning of stress in Turkish loans where the placement of stress is often ambiguous and further light needs to be shed on the underlying patterns of this, also under consideration of word stress in Turkish.

(vi) For easier referencing, all linguistic examples should get a unique reference number (e.g., chapter number + example number) instead of consequent numbering per chapter. Furthermore, not few examples are repeated to illustrate different phenomena. These should be marked as re-occurring by repeating the original identifier together with a comment like “ex. X repeated here for the sake of convenience as X”.

(vii) The indication of lexeme frequencies in the Romeyka corpus provides some interesting information about the proportion of individual parts of speech. However, for the sake of practicability, only token frequencies could be considered in the present thesis. Instead, it would be worth checking for lexeme frequencies according to type, e.g., it would be interesting to have an indication of the overall number of verbal lexemes in the corpus. The estimation of amounts of loan words per part of speech category is also preliminary at the present stage, and a more solid methodology may be found to classify native and borrowed parts of speech.

(viii) Finally, a traditional grammar should contain some glossed example texts where coherent speech can be seen. Due to difficulties in exporting the glossed texts from the ELAN files to word, the present grammar sketch does unfortunately not contain any glossed texts, although they can be easily accessed in the attached corpus files. In addition to texts, it would be meaningful to provide wordlists or a lexicon of the language together with the grammar.

Although a wordlist of Romeyka can be extracted from the ELAN corpus and an additional lexicon file has been set up for Romeyka during fieldwork, the task is not trivial for Romeyka, as ways need to be found how to deal with the existing variation. Furthermore, the wealth of available etymological information on Greek words (see Section 1.5.4) would make it meaningful to represent the Romeyka lexicon in a glossary together with further information on semantic change, etc.. The compilation of such a file was unfortunately not feasible at the present stage of research.

## 2 Phonology

To date, no comprehensive analysis of Romeyka phonology or Romeyka-internal phonological variation is available (but for a basic phonological description see Schreiber 2018: 897–900 and Neocleous 2020: 24–33). Basic phonological features, in particular those diachronically divergent from SMG or other Greek varieties, have also been described by Mackridge (1987, 1999), Bortone (2009), and Özkan (2013). The phonology of Christian Pontic Greek is described in great detail by Oikonomidis (1908). Also, Drettas’ grammar of Pontic Greek (of the variety of Chaldia) includes a comprehensive chapter on phonology (1997: 41–106). Furthermore, Brendemoen (2002) describes the phonology of the Turkish dialects of Trabzon with reference to contact with Romeyka.

The phonological analysis presented in this chapter is essentially based on word lists of different speakers predominantly from ROF (Çaykara) (see Table 1) although the results are collated with the existing literature. Lexical data which are only attested in three male heritage speakers of ROF in Germany are marked throughout this chapter by (HS).

**Table 1:** Data (word lists) considered for phonological analysis

Corpus reference	Gender of speaker(s)	Age of speaker(s)	Dialect group	Residence
01_06042017F_0	Female	81	ROF (Çaykara)	Istanbul
04_01072019F_6 04_01072019F_7 04_01072019F_8	Female	49	ROF (Çaykara)	Çaykara
05_03072019M_1 05_03072019M_2	Male	ca. 40	RSür (Beşköy)	Sürmene
01_28062019F_4	Female	72	ROF (Çaykara)	Çaykara
04_21042018M_4	Male (3 speakers)	40, 45, 97	ROF (Çaykara)	Germany (H)

Furthermore, a complete word list of the Romeyka corpus has been extracted from the ELAN files and was considered for phonological analysis. Finally, Tursun’s (2019) dictionary, which comprises data from different dialects, has been used for the sake of comparison with other varieties or for finding words not occurring in the present corpus, for example to identify minimal pairs. Words from this dictionary which do not appear in the Romeyka corpus are marked throughout this chapter by (TD).

Although a narrow phonetic description would have been beneficial for the phonetic analysis (Chelliah & de Reuse 2011: 255–256; Meakins, Green & Turpin 2018: 99), a broad phonemic transcription has been applied in (nearly all of) the word lists for the following reasons: (i) since the focus of the present work is on the morphosyntactic description, the broad transcription was more convenient and efficient; (ii) the phonemic transcription allows for easy computerized searches in the corpus data; (iii) the corpus data are more easily accessible to a wider non-linguistic audience (cf. Chelliah & de Reuse 2011: 256).

The following phonemes (and marginal phonemes) used in the transcription of examples throughout the thesis differ from their representation in IPA: [ʃ] → *ş*, [tʃ] → *tş*, [ʒ] → *ž*,

[dʒ] → dʒ, [ç] → č, [n, ŋ] → n, [χ] → x, [œ] → ö,<sup>12</sup> and [y] → ü in Turkish loanwords.<sup>13</sup> The adopted orthography is in line with practice in recent research on Romeyka (Neocleous 2020; Sitaridou 2014b, i.a.). Although relevant for morpho-phonological processes such as vowel deletion or for example the occurrence of the plural suffix /æ/ only in nouns with an underlying unstressed plural in /ia/, word stress is for reasons of practicability only indicated in the present chapter on phonology (see also Section 1.2.5).

## 2.1 Segments

This section describes the phonological segmental inventory of Romeyka, that is, all distinctive sounds. Reference is made to phonemes that occur only in Turkish loanwords but due to the high level of bilingualism including productive code-switching, the phonological particularities of the regional Trabzon Turkish dialect can otherwise not be discussed (but see Brendemoen 2002).<sup>14</sup> Minimal pairs and near-minimal pairs are provided where available to verify phoneme status. The distribution of each sound is described for all environments, i.e., adjacent to vowels and consonants and in clusters; in all syllable and word positions: in word-initial, -medial, and -final position; and in all stress positions: stressed and unstressed. Furthermore, allophony and free variation are considered.

Note that the analysis of segments in this section is based on only roughly a dozen example words per phoneme, which were selected from the word lists with the objective of depicting a phoneme in as many different environments as possible. The distribution of these segments is summarized in the Tables B.1–8 in Appendix B.

### 2.1.1 Vowels

Romeyka has seven vowels which are distributed as shown in Figure 3. They are classified according to the qualitative aspects of openness, backness and roundedness. The historic quantitative feature of a phonemic vowel length distinction is lost today (Holton et al. 2019; Horrocks 2010: 160–170).

	Front	Central	Back
Close	i		u
Close-mid	e		o
Open-mid		œ	
Near-open		æ	
Open		a	

**Figure 3:** The Romeyka vowel inventory

In addition to the five vowels also present in SMG, /a/, /e/, /i/, /o/, /u/, Romeyka exhibits the secondary vowels /æ/ and, more rarely, /œ/ (Mackridge 1987; Özkan 2013; cf. Holton et al. 2019: 28). The sub-dialect RO<sub>f</sub> as spoken in Uzungöl (Saráchos) seems to be the only variety lacking the latter two vowels (Mackridge 1987: 131). In general, their phonemic status is dubious (see detailed description of the vowels below). The vowels /i/, /e/, /æ/, and /a/ are

<sup>12</sup> Except in Chapter 2 Phonology where word accent is indicated: [œ] → æ, á.

<sup>13</sup> The IPA value of the phoneme /j/ is [j], although Neocleous (2020: 27) lists it as [j̞]. There is probably dialectal and/or idiolectal variation.

<sup>14</sup> Neocleous (2020: 30) also confirms the different patterning of inherited and borrowed phonemes in RO<sub>f</sub>.

unrounded; /œ/, /o/, and /u/ are rounded; for examples see (1) below. Furthermore, Romeyka exhibits the following diphthongs: /au/, /eu/, /ai/, /ui/, /ei/ (see Section 2.2.1).

- |     |     |   |
|-----|-----|---|
| (1) | /i/ | <i>tširi</i> ‘father’   |
|     | /e/ | <i>(j)éma</i> ‘blood’, <i>peði</i> ‘child’                      |
|     | /æ/ | <i>omádæ</i> ‘eyes’   |
|     | /a/ | <i>mána</i> ‘mother’  |
|     | /œ/ | <i>gæni</i> ‘skin’, <i>gænksæ</i> ‘breasts’                     |
|     | /o/ | <i>stóma</i> ‘mouth’, <i>ónema</i> ‘name’, <i>šón(i)</i> ‘snow’ |
|     | /u/ | <i>furnó(s)</i> ‘frog’  |

The vowel phonemes /i/, /a/, and /u/ are realized like the respective cardinal vowels [i], [a], [u]. The vowels /e/ and /o/ are realized more open than the respective cardinal vowels [e] and [o] (see ex. 2a/b). /æ/ which only exists as plural suffix may be articulated in a more closed position thus tending towards [ɛ] as a variant of /æ/ (ex. 2c).<sup>15</sup> /œ/ appears more centralized than the cardinal vowel [œ] (ex. 2d). In some occurrences, the realization of the vowel /œ/ comes close to a diphthong /io/ which corresponds to the historical (written) form, e.g., *kióvi* ‘skin’ (ex. 2d).

- |     |    |           |                     |                       |
|-----|----|-----------|---------------------|-----------------------|
| (2) | a. | /e/ → [ɛ] | [‘jɛma] ‘blood’     | (04_01072019F_7; 027) |
|     | b. | /o/ → [ɔ] | [‘ɔnema] ‘name’     | (04_01072019F_7; 122) |
|     | c. | /æ/ → [æ] | [‘jæŋksæ] ‘breasts’ | (04_01072019F_7; 006) |
|     | d. | /œ/ → [œ] | [‘jœni] ‘skin’      | (04_01072019F_7; 017) |

In the following, the environment of each vowel phoneme is described in more detail. Importantly, it needs to be mentioned that the description of environments (for both vowels and consonants) is derived from a currently very small set of lexical items (approx. 10 example words per vowel phoneme; for the segmental analysis of vowels see Table B.1A/B in Appendix B). Therefore, it is inevitable that either environments are missed in which the vowel can also occur or that frequencies are not estimated correctly. So, the environments listed below are to be viewed as of hypothetical nature and need to be checked against more extensive data.

/i/

/i/ occurs in word-initial position as nucleus without onset both stressed and unstressed (*(j)inéga* ‘woman’, *ílo(s)* ‘sun’), in unstressed word-final position after different consonants (*ódi* ‘ear’, *gæni* ‘skin’, *rízi* ‘root’) and in both stressed and unstressed local environments frequently before sonorant consonants and fricatives but also plosives (*mídi* ‘nose’, *ftíra* ‘louse’, *ylitši* ‘milk’, *kariði* ‘walnut’, *filo* ‘leaf’, *tširi* ‘father’, *(j)inéga* ‘woman’, *ápsimo* ‘fire’, *liðári(n)* ‘stone’). It also occurs as the nucleus in stressed word-final syllables either preceding /a/ (*karðía* ‘heart’) or before a glide (*vizíja* ‘breasts’). In sum, there are hardly any restrictions on the distribution of /i/.

/e/

/e/ occurs in word-initial position in unstressed syllables before plosives (*ebán* ‘above’) and in stressed syllables before sonorant consonants (namely nasals and laterals) and plosives (*(j)éma* ‘blood’, *éla* ‘come.IMP.2SG’, *ébare* ‘take.IMP.2SG’). In this position, it occurs as the augment which expresses past (imperfect and aorist) tense (e.g., *pá(y)-o* ‘go’, ipfv. *e-péj-na*, aor. *e-píy-a* ‘I went’). It also occurs word-finally in unstressed syllables for example after laterals, nasals,

<sup>15</sup> Note that in ROF as spoken in Çaykara/Uzungöl (Sarachós), [æ] is generally replaced by /e/ (Peter Mackridge, p.c.).

and fricatives (*katsímale* ‘fog’, *ðós* ‘give.IMP.2SG’, *alévre* ‘flour’,<sup>16</sup> *stúðe* ‘bones’, also *ópse* ‘yesterday’). Furthermore, /e/ occurs between both sonorants and obstruents in unstressed (*ánemo(s)* ‘wind’, *pédi* ‘boy’, *xameléda* ‘mill’, *neró* ‘water’) and stressed syllables (*fénego(s)* ‘moon’, *thésó* ‘put.IMP.2SG’, *yarðéli* ‘child’, *šéri* ‘hand’, *(j)éma* ‘blood’, *skolétiši* ‘beetle’). The retention of the so-called “Ionic” [e] stemming from AGr [ɛ:] <η> mainly in unstressed syllables is one of the main characteristics of PG, e.g., *nífe* ‘bride’ < AGr *νύμφη* [nýmp<sup>h</sup>ɛ:], cf. MedGr *νύφη* [nífi]; *ílo(s)* ‘sun’ < AG *ἥλιος* [h<sup>h</sup>é:lios], cf. MedGr *ἥλιος* [ílios] (Sitaridou 2014b: 30; Dawkins 1916: 67; Papadopoulos 1955: 14–17; Oikonomidis 1908: 11–15; Mackridge 1987: 121–122, Holton et al. 2019: 27).

### /æ/

[æ] occurs in very limited environments; most prominently, as a plural suffix in the noun class of neuter nouns ending in unstressed word-final *-i*, e.g., *ospíti-a* > *ospíd-æ* ‘houses’, but *pedí-a* ‘children’. Additionally, it occurs after the voiced dental fricative /ð/, e.g., in the preposition *ðæ* ‘by, for’ (< MedGr *διὰ*) or in prefixed verbs like *ðævazo* ‘spend, read’ (< MedGr *διαβάζω*).<sup>17</sup> As a plural suffix, it can occur after sonorants as well as obstruents (*yarðélæ* ‘children’, *podáræ* ‘legs’, *ðóndæ* ‘teeth’, *aðélfæ* ‘siblings’, *gáenksæ* ‘breasts’). It alternates in free variation with /a/ and /e/ depending on the dialect (see below), e.g., *opsáræ*, *opsáre*, *opsár(j)a* ‘fish.PL’. /æ/ is historically derived via synizesis (i.e., the first high vowel in a vowel sequence becomes a palatal glide) from AG *-ia/* or *-ea/*, if /i/ resp. /e/ is unstressed (Mackridge 1987: 121; Özkan 2013: 142; Tursun 2019: 34; but see Oikonomidis (1908: 7–10) for examples of [æ] which cannot be derived from synizesis; cf. Holton et al. 2019: 7). It occurs in different varieties of Romeyka, except for an area of ROF (Çaykara) between the villages of Uzungöl (Saráchos), Karaçam (Ótsena) and Uzuntarla (Aliθinos) (Tursun 2019: 34); Özkan (2013: 132) notes it for RSür as spoken in Beşköy, as well as for Christian PG, and Mackridge (1987: 131) confirms its lack in ROF as spoken in Uzungöl (see also Fn. 15). Since /æ/ occurs in very limited environments and goes back to the vowel sequence /ia/, /æ/ is considered in the present thesis a marginal phoneme with uncertain status.

### /a/

/a/ occurs in word-initial position as nucleus without onset in both stressed (*ándra(s)* ‘husband’, *árgo(s)* ‘bear’) and unstressed (*aðelfó(s)* ‘brother’, *alévre* ‘flour’) environments. Furthermore, it occurs in unstressed word-final position after both sonorants and obstruents (*stóma* ‘mouth’, *ylóssa* ‘tongue, language’, *zá* ‘animals’). There it constitutes either a marker of a class of feminine nouns in *-a* (*(j)inéga* ‘woman’) or the nom/acc plural morpheme of neuter nouns in *-o* (*próvato(n)* ‘sheep’, pl. *próvata*). With the glide /j/ in word-final syllables it is sometimes separated from /i/, otherwise a vowel cluster /ia/ occurs, e.g., *mairí(j)a* ‘soup’ (see Section 2.2.1). /a/ occurs furthermore in all other environments in both stressed (*mána* ‘mother’, *mamí* ‘grandmother’, *potámi* ‘valley’, *kosárin* ‘chicken’, *káda* ‘cat’, *yála* ‘yoghurt’, *opsári(n)* ‘fish’, *podári* ‘leg, foot’, *ðáxtila* ‘fingers’, *aláti* ‘(tree)’ and unstressed syllables (*yónado(n)* ‘knee’, *malía* ‘hair, wool’, *batsí* ‘girl’, *álas* ‘salt’).

<sup>16</sup> Although *alévrae* is the plural form of the noun *alévri* ‘flour’ (< MedGr *ἀλευριάρι* ‘flours’), it seems in the present data that the plural form has become generalized also for the singular (cf., e.g., 08\_04072019M\_2; 150). Therefore, the form *alévre* is listed here as singular.

<sup>17</sup> These two occurrences of /æ/ were not striking in the present corpus (but cf. Tursun 2019: 232), where /æ/ is often indistinguishable from /e/, except for the plural ending *-æ*.

/œ/

/œ/ occurs in the Romeyka corpus only with a few examples and in very restricted environments mainly in stressed syllables after a velar voiced stop (*gœni* ‘skin’, *gœnksæ* ‘breasts’).<sup>18</sup> Its phonemic status is dubious; it is derived from synizesis, i.e., a merger, of the underlying vowel sequence /io/ or /eo/, if unstressed (Oikonomidis 1908: 23; Mackridge 1987: 121; Özkan 2013: 142; Holton et al. 2019: 7). /œ/ is considered a marginal phoneme here. It seems to exist also as a free variant of /e/ and /o/ in *tšǽrado* vs. *tšérado* ‘horn’, *tšǽrási* vs. *tšérási* ‘cherry (tree)’, *gǽlím̥bema* vs. *kolím̥bema* ‘swimming.NMZ’, and in the loan word *gemlég̃in* vs. Tr. *gömlek* ‘blouse’, that is, after the velar stops /k/(>/tš/) and /g/. Interestingly, /œ/ does not only undergo de-rounding from Turkish loans (as in *gemlég̃in* ‘blouse’) but also occurs as variant of underlying /o/ as in the case of *kolím̥bema* ‘swimming.NMZ’, that is, /œ/ is involved in bidirectional processes. At least in the case of *gǽlím̥bema* ‘swimming.NMZ’, association of the word with the Tr. word *göl* ‘lake’ might have an influence. As mentioned above, /œ/ can be realized in open syllables where the vowel is diphthongized to /io/. This also extends to Turkish loanwords like *gǽli* < Tr. *göl* ‘lake’ which can be realized as /kioli/.

/o/

/o/ occurs in word-initial position either stressed (*odí* ‘ear’, *ónema* ‘name’) or before the stress-bearing syllable (*omádi* ‘eye’, *opsári(n)* ‘fish’, *ospídi(n)* ‘house’) where it is realized slightly more openly. If unstressed, /o/ may be dropped in some varieties (possibly also on an idiolectal basis). In word-final position, it occurs either stressed (*aðelfó(s)* ‘brother’, *kamnó* ‘smoke’) or unstressed (*to* ‘the’, *zó* ‘animal’, *škilo(s)* ‘dog’, *fúrno(s)* ‘frog’, *lígo(s)* ‘wolf’) in different environments. In word-medial position in both open and closed syllables, it occurs in different environments both stressed (*stóma* ‘mouth’, *ylóssa* ‘tongue, language’, *ðóndæ* ‘teeth’, *rðómo* ‘road’, *yónado(n)* ‘knee’, *bolá* ‘very’, *próvado(n)* ‘sheep’, *psómi* ‘bread’, *xóma* ‘soil’) and unstressed (*áuro(s)* ‘man’, *próvado(n)* ‘sheep’, *poðári* ‘foot, leg’, *kosárin* ‘chicken’, *liftokári* ‘hazelnut’, *xorío* ‘village’).

/u/

/u/ occurs in word-initial position, mostly stressed (*úla* ‘all’) but also unstressed (*utš* ‘not’). It occurs in a limited number of occurrences in word-final position, for example in mono-syllabic function words (*pú* ‘where’, *kátu* ‘below’). In word-medial position, it occurs in both stressed (*stúði* ‘bone’, *krúyo* ‘hit’) and unstressed syllables (*furnó(s)* ‘frog’, *xuljéri* ‘spoon’, *káhume* ‘sit’, *ebugá* ‘under’, *tsubáði* ‘corn’) in different environments. Comparing the frequency of the other vowel phonemes, /u/ occurs less frequently and is subject to dialectal variation, as Özkan (2013: 139) reports: RSür as spoken in Beşköy substitutes /u/ by /o/ in some words, e.g., *ólo* ‘all’ (for a different analysis of *ólo(n)* and *úlo(n)*, see Section 3.2.3).

The phonemic status of vowels can be proven by the following minimal (and near-minimal) pairs:

/a/ ~ /e/	/p_ts/	<i>batsí</i>	‘girl’	~	<i>betsí</i>	‘skin’ (HS)
	/l_ɣ/	<i>láyo</i>	‘how’	~	<i>léyo</i>	‘say’
	/θ_#/	<i>érθa</i>	‘I came’	~	<i>érθe</i>	‘s/he came’
/a/ ~ /i/	/m_k/	<i>maḱrí</i>	‘far, long’	~	<i>mikrí</i>	‘small.M/F.NOM.PL’
	/š_#/	<i>raší</i>	‘mountain’	~	<i>ráša</i> <sup>19</sup>	‘back’

<sup>18</sup> Cf. also *šóni* ‘snow’ < MedGr. *χιόνι(v)* ‘snow’, which has probably arisen via [io] > [œ] > [o] following [ʃ] (Holton et al. 2019: 28); although an alternative evolution *χιόνι* [çóni] > [šóni] could be possible as well.

<sup>19</sup> This minimal pair goes back to the same root AG *πάχις* > MedGr *πάχια* [ráç(i(a))] ‘back’, with diminution in MedGr *πάχιο(v)* [raçio(n)] ‘mountain (rig)’ (M. Janse, p.c.).

/e/ ~ /i/	/p_ð/	<i>pēðí(n)</i> ‘boy’	~	<i>apīðí</i> ‘pear tree’
	/l_tš/	<i>skolétší</i> ‘bug’	~	<i>lītší</i> ‘wolves’
	/m_#/	<i>mē</i> ‘with’	~	<i>mī</i> ‘not (NEG)’
/i/ ~ /u/	/t_#/	<i>kátī</i> ‘something’	~	<i>kātu</i> ‘below’
	/#_l/	<i>ólon</i> ‘more (COMP)’	~	<i>úlo(n)</i> ‘whole, all’
	/t_#/	<i>tō</i> ‘the.N.NOM’	~	<i>tū</i> ‘the.N.GEN’
/o/ ~ /a/	/z_#/	<i>zō</i> ‘animal’	~	<i>zá</i> ‘animals’
	/#_n/	<i>ónema</i> ‘name’	~	<i>ánemo</i> ‘wind’
/o/ ~ /e/	/t_#/	<i>atō</i> ‘he’	~	<i>ate</i> ‘she’
	/p_s/	<i>(o)pīs</i> ‘behind’	~	<i>pós</i> ‘how’

Triplets and quadruples exist for example for:

/o/ ~ /i/ ~ /æ/	/f_#/	<i>aðelf<u>o</u>(s)</i> ‘brother’	~	<i>aðelf<u>i</u></i> ‘sister’	~	<i>aðelf<u>æ</u></i> ‘siblings’
/o/ ~ /u/ ~ /a/ ~ /i/	/t_#/	<i>kát<u>o</u>(s)</i> (TD) ‘male cat’	~	<i>kát<u>u</u></i> ‘down’	~	<i>kát<u>a</u></i> ‘female cat’ ~ <i>kát<u>i</u></i> ‘something’

For the marginal phoneme /æ/ no minimal pairs could be found apart from plural forms:

/æ/ ~ /i/	/d_#/	<i>omádi</i> ‘eye’	~	<i>omádæ</i> ‘eyes’
	/f_#/	<i>xoráfī</i> ‘field’	~	<i>xoráfæ</i> ‘fields’

Rather, there is free variation between:

/æ/ ~ /e/	/ð_#/	<i>nifáðæ</i> ‘brides’	~	<i>nifáðe</i> ‘brides’
/æ/ ~ /a/	/r_#/	<i>muskáræ</i> ‘calves’	~	<i>muskára</i> ‘calves’

No minimal pairs could be found for the marginal phoneme /œ/ but rather free variation:

/e/ ~ /œ/	/tš_r/	<i>tšérado</i> ‘horn’	~	<i>tšérado</i> ‘horn’
/o/ ~ /œ/	/k_l/	<i>gōlímema</i> ‘swimming.NMZ’	~	<i>gölímema</i> ‘swimming.NMZ’

It should be noted here that there exists a lot of variation regarding the vowel quality in many lexemes of Romeyka which leads to different pronunciations of the same lexical item (for examples, see below). Variation in vowels occurs often at the syllable boundaries (although not exclusively) and frequently in function words and nominal and verbal inflection. Further research must shed light on whether this variation is dialectal, phonologically conditioned or evidence for an ongoing change (internal or contact-induced). The diachrony of the lexemes is likely to provide cues as well.

/a/ ~ /e/	/#_ð/	<i>ēđá</i> ‘here’	~	<i>āđá</i> ‘here’
	/#_b/	<i>ēbán</i> ‘above’	~	<i>ābán</i> ‘above’
	/v_#/	<i>škévā</i> (TD) ‘dishes’	~	<i>škevæ</i> ‘dishes’
	/r_#/	<i>tomáre</i> (TD) ‘together’	~	<i>tomára</i> ‘together’ ~ <i>tomári</i> ‘together’



/a/ ~ /i/	/t_#/	<i>osát<u>a</u></i> (TD) ~ (o)sát( <i>i</i> ) (TD) ~ sít( <i>e</i> ) (TD) ‘when (function word)’
/a/ ~ /o/	/#_m/	<i>ámon</i> (TD) ‘like’ ~ <i>émon</i> ‘like’ ~ <i>ómon</i> ‘like’
	/#_n/	<i>ánda</i> ‘when’ ~ <i>óna</i> ‘when’
/e/ ~ /i/	/t_v/	<i>kat<u>e</u>véno</i> ‘descend’ ~ <i>kativéno</i> (TD) ‘descend’
/e/ ~ /o/	/#_m/	<i>émorfo</i> ‘beautiful’ ~ <i>ómorfo</i> ‘beautiful’
	/p_r/	<i>perpát<u>e</u>ma</i> (TD) ~ <i>porpát<u>e</u>ma</i> ‘walking.NMZ’
	/v_l/	<i>velóni</i> ‘needle’ ~ <i>volóni</i> ‘needle’
/i/ ~ /o/	/p_t/	<i>pítázo</i> (TD) ‘direct’ ~ <i>potázo</i> (TD) ‘direct’
	/#_f/	<i>ofétos</i> (TD) ‘this year’ ~ <i>ifétos</i> ‘this year’
/i/ ~ /u/	/s_r/	<i>síro</i> ‘pull’ ~ <i>súro</i> (TD) ‘pull’
/o/ ~ /u/	/k_v/	<i>ekoválena</i> ‘I carried’ ~ <i>ekuválena</i> (TD) ‘I carried’

In addition to the seven vowels described above, Mackridge (1987: 121) mentions that the vowels /u/ and /y/ are used in Turkish loans only (also Neocleous 2020: 24), whereby Özkan (2013: 142) points out that these vowels often figure differently in the Trabzon-Turkish dialect, possibly due to contact with Greek (see also Brendemoen 2002). While the frequency of the vowel /æ/ is only increased through Turkish loans (e.g., *şofæris* ‘driver’), the vowels /u/, /y/ and /æ/ are in the Romeyka corpus often realized differently from the Turkish standard and in line with regional Trabzon Turkish pronunciation:

Tr. <i>kızdık</i>	>	/egizévame/	‘we got annoyed’
Tr. <i>altı</i>	>	/altí/	‘six’
Tr. <i>göl</i>	>	/kióli/	‘lake’
Tr. <i>böyle</i>	>	/haboile/	‘like this’
Tr. <i>çörek</i>	>	/išurég/	‘cottage cheese’
Tr. <i>üç</i>	>	/udž/	‘three’
Tr. <i>yüz</i>	>	/juz/	‘hundred’
Tr. <i>üstelik</i>	>	/ustelúg/	‘on top’
Tr. <i>üstüne</i>	>	/ustiné/	‘above’
Tr. <i>çünkü</i>	>	/džingí/	‘because’, but:
Tr. <i>gün</i>	>	/gün/	‘day’
Tr. <i>değirmen</i>	>	/dæ:rmén/	‘mill’

### 2.1.2 Consonants

The consonantal inventory of Romeyka is presented in Table 2; the phonemic status of the grey-shaded consonants is dubious, or they are non-phonemic allophones such as [ŋ]; some of them are considered marginal phonemes because they occur in restricted environments and have a restricted functional load for reasons outlined below. Often, they are the result of phonological processes in the diachrony of Romeyka. All existing consonants, including the marginal ones, are described in the following sub-sections.

**Table 2:** The consonant inventory of Romeyka<sup>20</sup>

	Labial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Plosive	p b		t d			k g		
Nasal	m		n			ŋ		
Tap			r					
Fricative	f v	θ ð	s z	ʃ ʒ		ɣ	χ	h
Affricates	ps		ts	tʃ dʒ		ks		
Approximant					j			
Lateral			l					

The following minimal (and near minimal) pairs of phonemes which differ from each other in a single feature (e.g., voicing, place of articulation) are attested:

/n/ ~ /m/	/o_d/	<i>yomádo</i> ‘full’	~	<i>yónado</i> ‘knee’
	/e_a/	<i>éna</i> ‘one’	~	<i>(j)éma</i> ‘blood’
	/#_e/	<i>né</i> ‘yes’	~	<i>=me</i> ‘me.CL’
/f/ ~ /v/	/#_i/	<i>físi</i> (TD) ‘nature’	~	<i>vízi</i> ‘breast’
/θ/ ~ /ð/	/a_i/	<i>maθízo</i> ‘teach’	~	<i>maðízo</i> (TD) ‘argue’
		<i>áθiyo</i> ‘without’	~	<i>ðíyo</i> ‘give’
/s/ ~ /z/	/#_o/	<i>so</i> ‘at the.M/N.SG.ACC’	~	<i>zó</i> ‘animal’
/ɣ/ ~ /χ/	/e_o/	<i>eyó</i> ‘I’	~	<i>éxo</i> ‘have’
/p/ ~ /m/	/#_a/	<i>pal</i> .FOC	~	<i>malí</i> ‘hair.SG’
/t/ ~ /s/	/#_o/	<i>ton</i> ‘the.M.SG.ACC’	~	<i>so(n)</i> ‘to the.M.SG.ACC’
/t/ ~ /θ/	/(a)_e/	<i>atena</i> ‘her.OPN’	~	<i>θéna</i> ‘I put’
/θ/ ~ /s/	/(e)_e/	<i>théna</i> ‘I put’	~	<i>eséna</i> ‘you.OPN’
/ð/ ~ /z/	/e_i/	<i>peði(n)</i> ‘boy’	~	<i>pézi</i> ‘he plays’
/p/ ~ /k/	/#_a/	<i>pal</i> .FOC	~	<i>kal</i> ‘again’
/p/ ~ /t/	/a_e/	<i>apés</i> ‘inside’	~	<i>ates</i> ‘her.POSS.(CL)’
/k/ ~ /t/	/#_l/	<i>kaló</i> ‘good’	~	<i>t=álo</i> ‘the other’
/r/ ~ /l/	/#_ /	<i>roízo</i> ‘fall’	~	<i>laízo</i> ‘dangle’
/n/ ~ /l/	/e_o/	<i>thélo</i> ‘want’	~	<i>théno</i> ‘put’
/n/ ~ /r/	/#_o/	<i>noízo</i> (TD) ‘feel’	~	<i>roízo</i> ‘fall’
/ð/ ~ /l/	/(e)_i/	<i>ðíyo</i> ‘give’	~	<i>(e)líyo</i> ‘a little’
/ð/ ~ /r/	/#_i/	<i>ðísi</i> (TD) ‘west’	~	<i>rízi</i> ‘root’
/z/ ~ /l/	/#_a/	<i>zaró</i> (TD) ‘uneven’	~	<i>láyo</i> ‘how’

<sup>20</sup> Neocleous (2020: 27) notes also the palatal stops /c/, /j/ and palatal fricatives /ç/, /j/, whereby it is not clear whether he considers them to have phonemic status. These palatal phones do not figure in the present corpus but have been occasionally heard from other speakers. Probably, they are dialectal or idiolectal variants of (post)alveolar plosives and fricatives.

/z/ ~ /r/	/i_o/	<i>sízo</i> (TD) ‘swing’	~	<i>síro</i> ‘pull’
/z/ ~ /n/	/a_V/	<i>maθízo</i> ‘teach’	~	<i>maθáno</i> ‘learn’
	/i_o/	<i>xtízo</i> ‘build’	~	<i>xtíno</i> ‘cow’
/ks/ ~ /ps/	/#_i/	<i>ksílo</i> ‘wood’	~	<i>psílo</i> ‘thin’
	/#_e/	<i>kséno</i> (TD) ‘foreign’	~	<i>pséno</i> ‘cook’
/ks/ ~ /ts/	/e_#/	<i>eks-</i> ‘out (prefix)’	~	<i>aets</i> ‘like this’
/ts/ ~ /ps/	/o_(e)/	<i>óts</i> ‘who(ever)’	~	<i>opsé</i> ‘yesterday’

Other minimal pairs are:

/k/ ~ /ɣ/	/#_ /	<i>kalá</i> ‘good’	~	<i>yála</i> ‘yoghurt’
/k/ ~ /χ/	/#_l/	<i>kláði</i> ‘branch’	~	<i>xláði</i> (TD) ‘weed’
/t/ ~ /z/	/#_o/	<i>to</i> ‘the.M’	~	<i>zó</i> ‘animal’
/t/ ~ /n/	/#_e/	<i>téro</i> ‘look’	~	<i>neró</i> ‘water’
/t/ ~ /v/	/#_r/	<i>trész</i> ‘s/he runs’	~	<i>vrész</i> ‘it rains’
/t/ ~ /f/	/#_e/	<i>téro</i> ‘look’	~	<i>féro</i> ‘bring’
/p/ ~ /f/	/e_i/	<i>epítše</i> ‘s/he did’	~	<i>efítše</i> ‘s/he left’
/s/ ~ /l/	/a_o/	<i>aso</i> ‘from the’	~	<i>álo</i> ‘more, other’
/s/ ~ /r/	/e_o/	<i>tesó</i> ‘your’	~	<i>téro</i> ‘look’
	/i_o/	<i>íso</i> (TD) ‘plain’	~	<i>iró</i> (TD) ‘wet’
/s/ ~ /ð/	/#_e/	<i>sevéno</i> ‘enter’	~	<i>ðevéno</i> ‘pass’
/z/ ~ /θ/		<i>Zéno</i> ‘village (Ulucami)’	~	<i>θéno</i> ‘put’
/m/ ~ /s/	/e_e/	<i>eména</i> ‘me’	~	<i>eséna</i> ‘you.OPN’
/m/ ~ /l/	/#_a/	<i>malía</i> ‘hair.PL’	~	<i>lalía</i> ‘voice, sound’

Some triplets to quintuplets exist for:

/#_ilo/	<i>mílo</i> ‘apple’	<i>filo</i> ‘leaf’	<i>psílo</i> ‘thin’	<i>ksílo</i> ‘wood’	<i>škílo(s)</i> ‘dog’
/#_ero/	<i>téro</i> ‘look’	<i>féro</i> ‘bring’	<i>neró</i> ‘water’	<i>(e)péro</i> ‘take’	<i>kséro</i> ‘know’
/i_a/	<i>ípa</i> ‘I said’	<i>íxa</i> ‘I had’	<i>iða</i> ‘I saw’		
/#_ulia/	<i>ðulía</i> ‘work’	<i>ɣulía</i> ‘cabbage dish’		<i>kulía</i> (TD) ‘goat without horn’	

In essence, the consonant system of Romeyka is Greek. However, the frequency of phonemes that already existed (or developed) in the Greek system may be increased through Turkish (and other) loanwords, e.g., /v/ (*veja* ‘or’, *duvári* ‘wall’), /χ/ (in Tr. *-lik* suffixes, e.g., Tr. *bolluk* ‘abundance’ > *boluxi*), /h/ (*sahat* ‘hour’). Also, the frequency of the postalveolar fricative [ʃ] (*beš* ‘five’) and the palatal affricates [tʃ] and [dʒ] increases via Turkish loans (*džami* ‘mosque’, *tšureg* ‘cottage cheese’). Furthermore, considering the distinction in voicing of stops in Turkish loanwords, /b, d, g/ could be regarded as phonemic as well (*bile* ‘even’, *tek* ‘only’, *pek* ‘quite’).

## 2.1.2.1 Voicing and aspiration on stops

In Romeyka, it is widely assumed that voicing is not a distinctive feature for stops. Furthermore, it is widely assumed that stops are underlyingly voiceless. Therefore, the following three phonemes are tentatively assumed here: /p/, /t/, /k/. They are nevertheless often realized as voiced (cf. Oikonomidis 1908: 32), although that seems to hold true to a lesser extent for the velar plosive /k/. Mackridge (1987: 123) notes for PG that there is, apart from loan words, no distinction of voice in the stops, which are generally voiceless (this does not apply to ROF as spoken in Uzungöl, though, Mackridge 1987: 132). According to Özkan (2013: 140), however, stops are considered voiced with fortis articulation and become lenis after a nasal (voiced consonants appear to be aspirated as in the regional Turkish varieties, Özkan 2013: 140). In addition, intervocalic voicing often takes place (Section 2.3.2). The lack of a clear distinction regarding voicing is also reflected in the phonology of the regional Turkish variety (Brendemoen 2002). The parallels between regional Turkish pronunciation of plosives and their realization in Romeyka is striking: “In most parts of Trabzon, including the nucleus districts, [in Turkish] aspirated unvoiced stops – /p/, /t/, and /k/ – only occur word-initially in emphasized position; in other cases, stops are either voiced – /b/, /d/, /g/ – or unvoiced unaspirated – /p/, /t/, /k/” (Brendemoen 2006: 68–69). According to Tursun (2019: 21), the lack of voice contrasts with stops appears in all Greek dialects but is un-systematic and depends in Romeyka on the “multilingualism” of the speaker, i.e., it seems that speakers with competence in SMG adhere more to the voiceless pronunciation of the stops.

The bilabial stop is /p/. In contrast to what Drettas (1997: 52–53) describes for PG, bilabial stops are not aspirated in Romeyka. According to Neocleous (2020: 30), only stops in loanwords are aspirated. The phoneme /p/ is realized as [p], [p̚], [b], [b̚]. Based on frequency and occurrence in most contexts, it is assumed that [p] is the phoneme with the variants [p̚], [b̚] and [b]. It is, however, difficult to establish a consistent set of phonological rules for the distribution of the more voiced variants. Often the degree of voicing seems to depend on the speaker without a particular sociolinguistic variable at work. Drettas (1997: 54) describes for PG that in his data, the variant of /p/ [b̚] occurs in speakers younger than age 35. Possibly this change has gone further in Romeyka and led to a general spread of the more voiced variants. As for the distribution of the variants of /p/, the phoneme is realized as [p] in a cluster with /s/, i.e., what might be considered an affricate /ps/ but is in Ancient Greek quantitative meters considered a consonant cluster ([ˈpsomi] ‘bread’). Furthermore, the following rule can be established (see also Mackridge 1987: 123; Özkan 2013):

/p/ → [b] / n, m\_ e.g., /embró/ [ɛmˈbrɔ] ‘before’, /(e)mbéno/ [(ɛ)mˈbɛno] ‘enter’,  
/koʃímbela/ [ɡɔdʒimbela] ‘plums’

This rule seems not necessarily to extend beyond word boundaries, though, e.g., /ton pésko/ → [ðon ˈpɛʃko] ‘the.ACC stove’. Furthermore, an only tentative rule is the following, which determines the degree of voicing of the bilabial plosive, although unpredictable variation exists:

/p/ → [b] / \_u, o, a e.g., /puθén/ → [buˈθen] ‘anywhere’, /polá/ → [boˈla] ‘very’, /pal/ → [bal].FOC, /páyo/ → [ˈbayo] ‘go’  
(not exclusively /#\_/: but cf. [ˈpu] ‘where’, [pɔˈla] ‘very’, [ˈpayo] ‘go’)  
/p/ → [b̚, p̚] / \_e, i e.g., /peði/ → [b̚ɛˈði] ‘child, son’, /peðás/ → [p̚ɛˈðas] ‘son’, /perméno/ → [p̚ɛrˈmɛno] ‘wait’, /epíjen/ → [ɛˈpijen] ‘he went’,

i.e., the further back the place of articulation of the vowel, the more voiced the bilabial plosive. These two rules described just above do not exclusively apply to syllable onsets – although frequently; there is a general tendency for voicing of voiceless stops in intervocalic

environments (voicing assimilation/intervocalic voicing, see Section 2.3.2; cf. also Holton et al. 2019: 205):

/p/ → [b, ɸ, p] / V\_V e.g., /epíʎa/ → [ɛ'biʎa] ‘I went’ ~ /epíra/ → [ɛ'ɸira] ‘I took’, [eɸebu'ga] ‘from below’, /ípe/ → ['ibe]/['iɸe]/['ipe] ‘s/he said’; but cf. voiceless /eʎápume/ → [ɛ'ʎapume] ‘we loved’ (01\_04022016F\_1; 091)

In word-final position, /p/ is realized slightly less voiced than usually, namely as [p] ([e'lep] ‘he sees’ 01\_06042017F\_1; cf. Drettas 1997: 52). This should not give rise to assuming final devoicing at work, though. Rather, the phonological environment beyond word boundaries seems to play a role. More phonological research is required here.

The alveolar stop /t/, realized as [ɖ], [d], [t̪], [t], shows also different degrees of voicing. Like with the bilabial stop, it is difficult to establish a pattern in the distribution of the voiced variants of the alveolar stop. Tentatively, there seems to be a tendency that [t] → [d, ɖ] / [ \_a], in an unstressed syllable (/pórta/ → ['borda] ‘door’, /xameléta/ → [ɣame'leða] ‘mill’, /káta/ → ['kaða] ‘cat’, also in plural forms in -æ, /ospítæ/ → [ɔ'spidæ] ‘houses’, /ðónt/ ‘tooth’ → ['ðɔndæ] ‘teeth’, /omátæ/ → [o'maðæ] ‘eyes) and more generally in (unstressed) intervocalic environments (/ospíti/ → [ɔ'spidi] ‘house’, /aláti/ → [a'laði] ‘tree’, /vútero/ → ['vuðero] ‘butter’, /ɣomáto/ → [ɣo'maðo] ‘filled’, but /kátu/ → ['kaɖu] ‘under’). This extends also to word-initial and -final contexts (/tíri/ → [ði'ri] ‘cheese’, /tranó/ → [dra'no] ‘big’, /tópos/ → ['dobɔs] ‘place’, /ospít/ → [ɔ'spid] ‘house’). [t̪] may dominate in stressed syllables (/potámi/ → [po'tami] ‘valley’). Voiceless [t] occurs predominately in affricates with the voiceless fricative /s/ (/tsupaði/ → [tsu'ɸaði] ‘corn’), variably also [ʃ] (/tše/ → [ʃɛ] ‘and’, /utš/ → [uʃ] ‘not’, /tširi/ → [ʃiri] ‘father’ but also [ɖɛ] ‘and’, [uɖɔ] ‘not’, [ɖɛiri] ‘father’), and in consonant clusters with /s/ (/patsí/ → [pa'tsi] ‘girl’, /stétš/ → [stɛʃ] ‘he stands’, /nestía/ → [nɛ'stia] ‘fasting’). In the neuter definite article, [t] and [ɖ] alternate, /to/ → [to], [ɖo] ‘the.N’, /ta/ → [ta], [ɖa] ‘the.N.PL’.

The velar stop /k/ is realized as [k̟], [g̟], [k], [g]. For historical reasons, it is assumed that /k/ is the phoneme (not [g̟], which occurs almost as frequently as /k/). As a tendency and in coincidence with the distribution of the other stops, a voiced variant of /k/ appears frequently in intervocalic environments (/líko(s)/ → ['ligo(s)] ‘wolf’, /árko(s)/ → ['argo(s)] ‘bear’, /(j)inéka/ → [(j)i'neka] ‘woman’) and after a nasal (/fén̄go(s)/ → ['fɛŋgo(s)] ‘moon’; also see Neocleous 2020: 29). A more voiceless variant appears in word-initial position (/kariði/ → [ka'riði], ‘walnut’, /káhume/ → ['kahume] ‘sit’), especially in stressed consonant clusters (/krúʎo/ → ['kruʎo] ‘hit’, /ksílo/ → ['ksilo] ‘wood’). Word-finally, (a variant of) /k/ appears only in loan words (/tšurég/ → [ʃu'rɛg] ‘cottage cheese’). /k/ has the allophone /ʃ/ / [ \_e, i] (see also Sections 2.1.2.6 and 2.3.1).

Neocleous (2020: 27–28) notes the existence of the palatal plosive [c] / [ \_e, i] in Turkish loanwords, which is, however, subject to micro-dialectal variation even within ROF as spoken in Çaykara (cf. Oikonomidis 1908: 33).

#### 2.1.2.2 Nasals

Romeyka exhibits two nasals: /n/ and /m/. Velar [ŋ] is an allophone of /n/ occurring before velar stops in codas (*fén̄go(s)* ‘moon’), and also word-initially (*ŋgózo* ‘?’<sup>21</sup> (04\_01072019F\_2; 151)). /n/ occurs in all syllable-positions adjacent to vowels (*nífe* ‘bride’, *ánemo(s)* ‘wind’, *próvado(n)* ‘sheep’) and consonants, including nasals, especially after a syllable break (*em.názo*

<sup>21</sup> The translation is not quite clear here; it might be a passive form going back to MedGr ἀγγίζω /angízo/ ‘touch’ or MedGr ἐγγίζω /engízo/ ‘approach’ (M. Janse, p.c.).

‘resemble’<sup>22</sup>, *án.drás* ‘husband’). A striking feature is the occurrence of /n/ in word-initial position preceding the alveolar stop /t/, e.g., [ndo] ‘what’ where it is distinctive, cf. *to* ‘the.N.SG.NOM/ACC’ (Mackridge 1987: 123). /m/ occurs most frequently adjacent to vowels in syllable-initial position. It can also occur in codas, though (*kam.nó(s)* ‘smoke’, *em.bró* ‘before’). In word-final position, it occurs only as a (reduced) inflection ending mostly of 1PL or as clitic pronoun (namely, 1SG weak possessive).

### 2.1.2.3 Tap

Alveolar /r/ is realized as a tap [ɾ]. It occurs word-initially before vowels (*raší* ‘mountain’, *rízi* ‘root’) and [ɾ] (*rðómo* ‘road’). In the latter case r-metathesis took place from *ðrómos* ‘road’ (see Section 2.3.7). It occurs in onsets alone and in consonant clusters (*ka.rí.ði* ‘walnut’, *a.lé.vre* ‘flour’, *ðró.mo(s)* ‘road, *brin* ‘before’, see also Section 2.2.2) and in codas (*ormí* ‘river’, *ár.go(s)* ‘bear’, *bór.da* ‘door’, *kar.ðía* ‘heart’). It may also occur word-finally, especially when final unstressed /in/ nominal endings of neuter nouns are (partly) dropped (see Section 2.3.5.2), e.g., *liftokár(i)* ‘hazelnut’.

### 2.1.2.4 Approximants

The alveolar lateral approximant /l/ occurs in all syllable environments. In onsets, it occurs alone (*lí.ko(s)* ‘wolf’, *ma.lía* ‘hair’) and in a cluster with /ɣ/ and /k/ (*ɣli.tší* ‘milk’, *klí.ði* ‘key’, see also Section 2.2.2), while it appears in codas alone (*bal.FOC*) and only across syllable boundaries adjacent to fricatives, nasals, and the glide /j/ (*aðel.fó(s)* ‘brother’, *al.mé.yo* ‘milk’, *xul.jéri* ‘spoon’).

The palatal approximant [j] has doubtful phonemic status and is considered a marginal phoneme here. Historically, it evolved from the voiced velar fricative /ɣ/ [ɣ] to a voiced palatal fricative /j/ [j] (M. Janse, p.c.) but it also evolved as a semi-vowel via synizesis of /i/ preceding another vowel – a phenomenon, which is usually not relevant for Romeyka (Holton et al. 2019: 15). It occurs (i) in word-initial position diachronically as an allophone of the velar fricative /ɣ/ / [ɣ\_i], but from a synchronous point of view it could be considered phonemic here (*(j)inéka* ‘woman’ < AG *γυνή* [gyˈnɛ:] > MedGr *γυναῖκα* [jiˈneka] ‘woman’, *ja* ‘for’ < MedGr *για* /ja/ < MedGr *δια* /ðja/ ‘for’, *ejéndune* ‘s/he became’ < MedGr *γίνομαι* ‘become’, *fají* ‘food’ < MedGr *φαγί* ‘food’; see also Oikonomidis 1908: 56); (ii) word-initially in Turkish loanwords before the vowels /a/, /o/, /u/ (*jok* ‘no’, *jan-im* ‘my side’, *juz* ‘hundred’). (iii) It is optionally inserted in onsets of stressed syllables before /e/ (‘*(j)éma* ‘blood’, *xul.(j)éri* ‘spoon’). (iv) It is also inserted as a glide between vowel-sequences, most strikingly /ia/ in plurals but also /ai/, /ie/ (*vizi(j)a* ‘breasts’, *škili(j)a* ‘dogs’, *mijés* ‘once’; cf. Oikonomidis 1908: 57; *n’éz.ma* ‘cow dung’ is the result of (partial) synizesis < MedGr *véασμα* /néazma/ or *νέασμα* /njázma/ ‘manure’, M. Janse, p.c.). (v) In 3<sup>rd</sup> person singular verb inflection of some verbs, it may occur word-finally after /e/ as a reduced form of the otherwise occurring diphthong /ei/ but at the same time mirroring a velar fricative in the verb stem (e.g., *léyo* ‘say’ > *lej* ‘s/he says’, *eftáyo* ‘make’ > *eft(j)é(j)* ‘s/he makes’). For determining the phonemic status of /j/, it should be differentiated between the first occurrence (i) going back to /ɣ/, where /j/ is phonemic and the other occurrences (iii–v), where it is mostly phonologically conditioned and appears as the glide /j/. For a more detailed description see Section 2.3.6.

<sup>22</sup> Also *emnέzo* in Tursun (2019: 246), < MedGr *(é)μοιάζω* /e(m)jázol/, *(ó)μοιάζω* /o(m)jázol/ (M. Janse, p.c.).

### 2.1.2.5 Fricatives

Romeyka exhibits the following fricatives: with a voice distinction, labiodental /f/, /v/, dental /θ/, /ð/, alveolar /s/, /z/, and postalveolar /ʃ/, /ʒ/; as well as velar voiced /ɣ/ and uvular voiceless /χ/. Glottal unvoiced [h] occurs in restricted domains and is therefore considered a marginal phoneme here.

The labiodental voiceless fricative /f/ occurs frequently word-initially and syllable-initially before all vowels (*fī.la* ‘leaves’, *fur.nó(s)* ‘frog’, *fē.ro* ‘put’, *tšo.fāl(in)* ‘head’) and also in a cluster [ \_ t] (*ftí.ra* ‘louse’, *kóft* ‘s/he cuts’). It also occurs in syllable-final (*lif.to.kári* ‘hazelnut’) and word-final position, the latter especially after deletion of the inherited noun endings (*xoráf(i)* ‘field’; see also Section 2.3.5.2).

The voiced labiodental fricative /v/ occurs also word-initially and syllable-initially (*vú.dero* ‘butter’, *vari* ‘heavy’, *vizí.ja* ‘breasts’, *ó.va* ‘eggs’, *pró.va.da* ‘sheep.PL’). It may appear in onsets in a cluster with sonorant consonants (*vroší* ‘rain’) and in codas before sonorants (*a.lé.vre* ‘flour’, *a.vlé.a* ‘courtyard’). It occurs rarely in word-final position and mostly in 3<sup>rd</sup> person verbal inflection of loanverbs integrated with -évo, e.g., *arajév* ‘search.3SG’ (for a comment on final devoicing see Section 2.3.2).

The voiceless dental fricative /θ/ occurs in word-initial position before front vowels and /a/ (*θélo* ‘want’, *θaðía* ‘?’). It occurs in syllable-initial position before all vowels (*li.θá.ri* ‘stone’, *ma.θí.zo* ‘teach’, *ka.lá.θo* ‘basket’) and in a cluster [ \_ r] (*án.θro.bo(s)* ‘man’). It may also occur word-finally after deletion of the final noun ending (*parakáθ(i)* ‘conversation’).

The voiced dental fricative /ð/ occurs word-initially before all vowels (*ðío* ‘two’, *ðí.yo* ‘give’, *ðulía* ‘work’, *ðavázo* ‘pass (by)’, *ðáxtila* ‘fingers’, *ðónda* ‘teeth’) as well as syllable-initially (*ɣar.ðé.li* ‘child’, *tsu.bá.ði* ‘corn’, *pe.ði* ‘son’, *a.ðel.fō* ‘brother’, *stú.ði* ‘bone’, *ha.ðá* ‘here’, *po.ðá.ri* ‘leg’). It may appear in a cluster with /r/ (*ðró.mo(s)* ‘road’, *án.ðro.bo(s)* ‘man’). In word-final position of verbs or nouns, it occurs after deletion of the final ending (*apíð(i)* ‘pear tree’).

The voiceless alveolar fricative /s/ occurs in word-initial position (*sú.ro* ‘pull’, *sindišénume* ‘we chat’, *só.yun* ‘at least’), albeit less frequently than postalveolar [ʃ], and often in a cluster with a stop (*stó.ma* ‘mouth’, *stró.ma* ‘bed’, *skolétši* ‘bug’, *ksílo* ‘wood’, *spaxtó* ‘goat’ (HS)). It occurs in onsets (*ko.sá.ra* ‘chickens’, *gáenk.sæ* ‘breasts’, *ɣlóssa* ‘tongue’, *opsári* ‘fish’); hardly in codas in word-medial position (*istera* ‘later’), but frequently word-finally (*áuros* ‘man’, *ílos* ‘sun’, *álas* ‘salt’, *árgo(s)* ‘bear’; for final devoicing cf. Section 2.3.2). Adjacent to vowels in word-medial position, it occurs frequently after the back vowel /o/ which is due to the grammatical function of the -os ending in masculine gender.

The voiced alveolar fricative /z/ occurs both word- and syllable-initially (*zó* ‘animal’, *rí.zi* ‘root’, *vi.zí.ja* ‘breasts’, in the verbal suffix -íz- e.g., *ey.no.rí.zo* ‘know’, *ko.lí.zo* ‘burn’) but less often in codas (*kózmōs* ‘world’ (TD) < MedGr *κόσμος* [kózmōs] and word-finally only incidentally in 3<sup>rd</sup> person singular verb endings, e.g., *škíz* ‘he splits’, cf. also Section 2.3.2). Adjacent to vowels, it occurs predominantly after the high front vowel /i/, especially due to the frequency of the verbal suffix -íz-. /z/ is considered contrastive with /s/ at least in syllable-initial position but evidence for contrast in other environments is thin, probably due to syllable-final devoicing which may cause /z/ becoming /s/ in syllable codas.

The voiceless postalveolar fricative /ʃ/ occurs secondarily: (i) as an allophone of /χ/ in onsets [#\_e, i], e.g., in verb inflection *éxo* ‘have’ > *éš(i)* ‘he has’; *šéri* ‘hand’ ~ SMG *χέρι* ‘hand’, *šóni*

‘snow’ < MedGr. *χιόνι* [çóni] ‘snow’ (see also Neocleous 2020: 29; Özkan 2013: 141; Mackridge 1987: 123);<sup>23</sup> (ii) as a result of a phonological process /sk/ > /ʃk/, e.g., *škilo(s)* ‘dog’, also word-medially, e.g., *béšgo* ‘stove’, Özkan (2013: 141) mentions also a process /st/ → /ʃt/ / [ \_e, i], e.g., *ištéra* ‘later’; and (iii) in a cluster with [t \_] as the outcome of palatalized /k/ → /ʃ/ / [ \_e, i], e.g., *tšíri* ‘father’ (Section 2.3.1). Despite its historical evolution via palatalization of the velar stop, /ʃ/ can be considered at least a marginal phoneme. For example, near-minimal pairs exist with /k/ and /ʃ/ in onsets before the same vowels, e.g., *šóni* ‘snow’, *kófto* ‘cut’.

The phonemic status of the voiced postalveolar fricative [ʒ] is dubious. It occurs very rarely outside the affricate /dʒ/ and then mostly inter-vocally as a variant of a dental or postalveolar fricative which is probably idiolectal or dialectal, e.g., *hažá* vs. *hađá* ‘here’, *dréžis* vs. *tréšis* ‘you run’, but also *abaržán* ‘from there’ (07\_07042019F\_6; 15), *katuház* ‘down there’ (07\_07042019F\_6; 40). See also the paragraph on affricates below.

The voiced velar fricative /ɣ/ occurs in word-initial position before the back vowels /o/, /u/ and central-back /a/ (*yála* ‘yoghurt’, *ɣómári* ‘pannier’, *yulía* ‘cabbage dish’; cf. the complementary distribution with [j] / [ \_e, i]) and in consonant clusters before /l/ (*ɣlitši* ‘milk’, *ɣlóssa* ‘tongue’). In onsets, especially intervocally, it occurs frequently before /o/, which is in verbs in -o the citation form (*lé.ɣo* ‘say’, *á.ɣo.me* ‘go.IMP.2SG’, *e.ɣo* ‘I’, *krú.ɣo* ‘hit’, *lá.ɣo* ‘how’, *e.ɣa.bó* ‘love’). It does only in few examples of (reduced) verb inflection occur word-finally (*krú(ɣ)* ‘it hits’, 01\_02022015F\_1; 08) and practically not in codas. As well the marginal phoneme /j/ / [ \_e, i] which is a palatal correspondent of /ɣ/ before front vowels, /dʒ/ may be described as an allophone of /ɣ/ / [n\_e, i] (see Section 2.1.2.6 below).

The uvular voiceless fricative /χ/ occurs in onsets, also in word-initial position, before the back vowels /o/, /u/ and central-back /a/ (*xameléda* ‘mill’, *xóma* ‘soil’, *xuljéri* ‘spoon’, *éxo* ‘have’). Özkan (2013: 140–141) reports for RSür as spoken in Beşköy allophonic variation regarding the place of articulation of the voiceless fricatives, e.g., /x/ → /χ/, /x/. It is not fully clear whether /χ/ is the phoneme or /x/ but since the former is predominant in the present data and also accepted by Özkan (2013: 140), it is considered phonemic here. In onsets / [ \_e, i], the marginal phoneme [ʃ] could be considered an allophone of /χ/ (see above). /χ/ can occur also word-initially in a cluster [#\_r, t, l] (*xróno* ‘year’, *xtíno* ‘cow’, *xláði* (TD) ‘weed’). It occurs in syllable onsets (*bar.xári* ‘pasture’, *ar.xadía* ‘work as farmhand’) and in codas (*đáx.tila* ‘fingers’, *ex.tés* ‘yesterday’), but usually not word-finally. In Turkish loanwords, /χ/ often replaces [h] in onsets (*xalí* < Tr. *halı* ‘carpet’, *xizmét(i)* < Tr. *hizmet* ‘service’, *xastás* < Tr. *hasta* ‘sick’, *ješilúxi* < Tr. *yeşillik* ‘verdure’) but also in codas adjacent to back vowels (*baxtšé* < Tr. *bahçe* ‘garden’, *sabaxtán* < Tr. *sabahtan* ‘from the morning’). Furthermore, /χ/ often replaces /k/ / [V\_back \_] both in onsets and codas, e.g., *xujévane* < Tr. *oku-* ‘they studied’, *doxtórin* < Tr. *doktor* ‘doctor’, *jox* < Tr. *yok* ‘no’,<sup>24</sup> *kaimáxi* < Tr. *kaymak* ‘cream’, *sáluxe* < Tr. *sağlık* ‘health’, *zorlúx* < Tr. *zorluk* ‘difficulty’, although not consistently in all speakers, e.g., *kuimágluk* < Tr. *kuymaklık* ‘ingredients for Kuymak (polenta dish)’ (see also Mackridge 1987: 133–134, who mentions following Oikonomidis (1908: 99 and 106) that the frication of Turkish stops occurs more in inland rather than coastal dialects).

The phonemic status of the voiced glottal fricative [h] is unclear. It occurs only in restricted domains: (i) most frequently word-initially [#\_a] in pronouns and spatial adverbs where it functions as a sort of deictic marker, e.g., (*h*)*avúdo* ‘this’ (see Section 2.3.6, also the respective Sections on pronouns and spatial adverbs, especially 3.2.2.2); in Oikonomidis (1908: 110), [h]

<sup>23</sup> Neocleous (2020: 27) also notes the existence of palatal [ç] in ROF, but this is not attested as a variant of /ʃ/, for example before front vowels, in the present Romeyka corpus, e.g., [ˈeʃis] ‘have.2SG’, instead of [ˈeçis].

<sup>24</sup> Tr. negation marker in existentials, *yok*, colloquially used for ‘no’ (Göksel & Kerslake 2005: 276).



is treated as prosthetic consonant. Blevins (2008: 87) states about the phonemic status of such prosthetic consonants: “the occurrence of a laryngeal closing or spreading gesture at a prosodic boundary is typically non-contrastive at its point of origin, and may continue to be non-contrastive for many generations”; (ii) in onsets [a, u] in word-medial position where it may be a dialectal variant of /θ/, e.g., *káhume* vs. *káθume* ‘sit’ (Tursun 2019: 36). (iii) Özkan (2013: 141) adds to this for RSür the use of [h] in the future particle *ha* (*na* in RO<sub>f</sub>) as well as the conditional particle *he*. [h] does not occur word- or syllable-finally. Since /h/ has disappeared already in the Greek of Classical times (i.e., 4<sup>th</sup> c. BCE, M. Janse, p.c.; Horrocks 2010: 171), /h/ is likely no inherited phoneme but might have developed under Turkish contact, or even as an areal phenomenon (see also Sections 3.1.4.4.3 and 3.2.2.2).<sup>25</sup>

### 2.1.2.6 Affricates

The following consonant clusters may be considered affricates in Romeyka, although this is disputable: alveolar /ts/, /ps/, and /ks/ count as consonant clusters in Greek linguistics although minimal pairs with simple consonants are available, e.g., *psílo* ‘thin’ vs. *filo* ‘leaf’, *kséro* ‘know’ vs. *téro* ‘look’. Post-alveolar [tʃ] and [dʒ], however, are in Greek linguistics clearly considered affricates. In the present work, /tʃ/ and /dʒ/ are considered marginal phonemes as they occur secondarily as conditioned variants of palatalized /x/, /g/, and /k/ (Mackridge 1987: 123, 130–131; cf. Özkan 2013: 141). The first three consonant sequences mentioned are considered affricates here, as the release of the stop is within the fricative and not before, and both phonemes occur as a single segment. The presence and status of affricates in Greek phonology is a highly controversial issue, though, which is not touched upon here.

In the following, the affricates (including the dubious ones) are briefly characterized. For their analysis see Table B.8A/B in Appendix B.

[tʃ] goes diachronically back to an allophone of /k/ / [e, i] but can be considered to have at least marginal phonemic status synchronically. It occurs in word-initial position and in onsets [e, i] (*tšíri* ‘father’, *tšília* ‘belly’, *tšímúme* ‘sleep’, *etšíno(s)* ‘he’, *kótšíno* ‘red’, *ylitší* ‘milk’; cf. also Özkan 2013: 141; Drettas 1997: 73). It occurs also phonologically conditioned in word-final position when deletion of a front vowel, mostly /i/, took place, e.g., *utš* < AG *oúkí* ‘not’, *stétš* ‘he stays’ < *stéki*.PRS.3SG < *stéko* ‘stay’ (see also Section 2.3.1).

[dʒ] occurs only rarely and in syllable onsets: (i) historically, it is an allophone of /g/ / [n\_e, i] (historically AG /γγ/ → /ηg/), e.g., *spundžizo* ‘swipe’ or *landžévo* ‘jump’ (Tursun 2019: 28); (ii) it also appears as a voiced variant of [tʃ] (*kolondžisa* ‘pumpkins’); (iii) it occurs in word-initial position of Turkish loanwords (*džamí* ‘mosque’). Due to the restricted occurrences and the diachronic allophonic status, it is considered a marginal phoneme here.

/ts/ occurs in word-initial and syllable-initial position [V] (*tsurúla* ‘soaked’, *tsubáði* ‘corn’, *tsakóno* ‘break’, *katsímale* ‘fog’, *batsí* ‘girl’). It may also occur word-finally (*aéts* ‘like this’ *óts(e)* ‘who’; but cf. Özkan 2013: 141). It is not clear whether it should be analysed as a consonant cluster /t/ + /s/, as has been argued for SMG τσ, or /tʰ/, as has been argued by others. The fact that /ts/ can occur in codas is in fact an argument for treating /ts/ as an affricate rather than a cluster of two segments.

/ks/ occurs in some words in word-initial position [e, i] (*ksílo* ‘wood’, *kséro* ‘know’, *kseró* ‘dry’). When occurring in onsets elsewhere, though, its status as an affricate becomes very dubious, e.g., *aksínári* ‘axe’. It appears also secondarily in the aorist form of some verbs with /χ/, e.g., *alméyo* ‘milk’ > *élmeksa* ‘I milked’, *eyvéno* ‘leave’ > *ekséva* ‘I left’, *eyválo* ‘take out’ > *ekséndže* ‘he took out’ (for a complete description of the phonological rules applying to the aorist formation see Drettas 1997: 214).

<sup>25</sup> Note for example also the existence of the reduced demonstrative *ha-* in Arabic < *hada/hadi/hadol* ‘this’ (Matras 2012b: 374).

/ps/ occurs in stressed syllables in word-initial position (*psómi* ‘bread’, *psílo* ‘thin’, *pséma* ‘ly’) and in onsets before all vowels (*ápsimo* ‘fire’, *opsári* ‘fish’, *anépsi* ‘niece’, *ðipsó* ‘be thirsty’, *opsé* ‘yesterday’). It appears in the aorist form of some verbs with /f/ or /v/, e.g., *xorévo* ‘dance’ > *exórepsame* ‘we danced’, *kófto* ‘cut’ > *ékopse* ‘he cut.PST’; according to Oikonomidis (1908: 44), /ps/ develops out of the combination of a bilabial fricative with a sibilant.

## 2.2 Phonotactics

### 2.2.1 Diphthongs and vowel sequences

Romeyka exhibits (at least) the following diphthongs, i.e., a vowel sequence filling a single syllable nucleus:

/ói/        *trói* ‘eat.3SG’<sup>26</sup>  
 /ai, aí, ái/ *haídiko* ‘such’, *haís* ‘like this’, *yaíðuran* ‘donkey’, *pái* ‘he goes’, *kainís* ‘somebody’  
 /úi/        *vúidi* ‘ox’ (also *vúir* ‘ox’), *fanerúi* ‘it reveals’, *krúi* ‘s/he hits’  
 /éi/        *roméika* ‘Romeyka’, *léi* ‘s/he says’, *palé(.)i* ‘ancestors’, *ebéina* ‘I went.IPFV’  
 /au/<sup>27</sup>    *(h)a(v)údo(s)* ‘this’ (also *hao* ‘this’), *ha.u-džéga* ‘here’, lit. ‘this side’, *á.u.ros* ‘man’

All diphthongs occur as nucleus in open syllables without coda (except for /ái/ in *haís* ‘like this’ which is a contracted form of *hajés* ‘like this’). Only the dubious vowel sequence /au/ exists in word-initial position (in some adverbs of place and pronouns preceded by /h/ in the onset) while the other diphthongs appear word-medially. When a diphthong arises due to 3<sup>rd</sup> person singular present verb inflection with -i, it occurs in word-final position. /ói/ and /úi/, which are not very frequent anyway, seem to occur predominantly after the labiodental voiced fricative /v/ or the tap /ɾ/ and before a dental or alveolar fricative or tap (*vúir* ‘ox’, *ro.ízo* ‘fall’) and are likely to have appeared secondarily as the outcome of a phonological process (see below). The occurrence of /ái/ and /éi/ is less restricted with regard to the phonemic environment (with the exception of palatalization).

It is unclear whether /io/ and /ia/ qualify as diphthongs. Stressed /ío/ and /ia/ are frequently realized as bisyllabic vowel clusters (see below) but can in unstressed form – probably, among others, dependent on the dialect area – also be realized monosyllabically as a glide plus vowel after synizesis, e.g., *pjó* ‘who/what’<sup>28</sup>, *xorjá* ‘villages’, *opsárja* ‘fish.PL’. This is despite the fact that retention of the original bisyllabic structure and resistance against the tendency to become /j/ + stressed vowel generally distinguish Romeyka from SMG (Mackridge 1987: 122; Bortone 2009: 83).

Finally, there seem to be certain phonological processes that cause diphthongization in Romeyka. Further research is required on these. Potentially, shifts in word stress could play a role leading to a diphthongization of the vowel of the stressed syllable, probably to increase phonological weight. Stressed and open syllables, which tend to be phonetically longer, are likely more prone to diphthongization, e.g., *ðulévo* ‘work’ > *e’ðuilévame* ‘we worked’, *tró(γ)o* ‘eat’ > *e’tró.ede* ‘they ate’<sup>29</sup>). Apart from this, there seems to be some variation with regard to whether a monophthong is realized or a diphthong, e.g., *kainís* ~ *kanís* ‘nobody’.

<sup>26</sup> A combination of a noun with its article can also lead to the articulation of a diphthong as in *o ílo(s)* > [óilós] ‘the sun’.

<sup>27</sup> Considering the diachronic perspective, /au/ proves not to be a diphthong but rather a vowel sequence such as in the examples provided here: *auros* ‘man’ < MedGr *ἄγυρος* [‘ayuros]; *(h)aúto(s)* [(h)a’uto(s)], also *(h)avudo(s)/(h)ayudo(s)*. However, considering the pronunciation in the present corpus, the realization is sometimes close to a diphthong, so /au/ is listed here for the sake of completeness.

<sup>28</sup> For the status of /io/, consider also Section 2.1.1 about the diphthongized realization of /œ/, e.g., in Turkish loanwords like *kióli* ‘lake’.

There are not many restrictions on bisyllabic vowel sequences (hiatus) which are numerous and may include bisyllabic realizations of some of the diphthongs presented above (cf. list below). In many cases, the vowel clusters can be separated at the syllable break, for example, by the approximant /j/ (*palé.(j)o* ‘old’) which is an example of glide insertion or in the (inherited) case of the velar fricative /ɣ/ (*pá.(ɣ)o* ‘go’). According to Drettas (1997: 219, Fn.1), PG – and this applies to Romeyka as well – has no contract verbs anymore like in AG: so, word-final /ei/ and /ao/ are separated by a morphemic boundary as in *é(š).is* ‘you have’ and are likely to be parted by /ɣ/. Note that some vowel sequences separated by the velar fricative also mirror an underlying velar fricative, for example, *lé.ɣo* ‘say’, and in the suppletive verb stem, e.g., *éfa(ɣ)a* ‘I ate’ > *tró(ɣ)o* ‘eat’ (for historical intervocalic deletion of /ɣ/, cf. Holton et al. 2019: 151).

/au/	<i>a.úrini</i> ‘men’ (but cf. <i>á.u.ros</i> ‘man’)
/oi/	<i>ro.ízo</i> ‘fall’, <i>vo.ízo</i> ‘call’, <i>o.ís</i> ‘this person’
/ai, aí/	<i>fa.ízo</i> ‘I feed’, <i>fa.(j)i</i> ‘food’, <i>ma.i.rí.(j)a</i> ‘dish’, <i>ma.i.ré.vo</i> ‘cook’, <i>la.ízo</i> ‘dangle’, in loanwords: <i>xolá.(j)i</i> ‘easy’
/èi/	<i>é(š).is</i> ‘you have’, in loanwords: <i>dé.(j)i.ne</i> ‘saying’ [< Tr. <i>diyene</i> ]
/ío/	<i>đí.o</i> ‘two’, <i>pí.o(s)</i> ‘who, what’, <i>tši.me.θí.o</i> ‘sleep.NMZ’, <i>xorio</i> ‘village’, <i>kri.os</i> ‘cold’
/ia/	<i>đulía</i> ‘work’, <i>nestía</i> ‘fasting’, <i>trayođía</i> ‘folksong’, <i>tšilía</i> ‘stomage’, <i>peđía</i> ‘children’
/ie/	<i>eruvije</i> ‘he fell’, <i>epí(j)e</i> ‘s/he went’ <sup>29</sup>
/áo/	<i>pá(ɣ)o</i> ‘I go’, <i>lá(ɣ)o</i> ‘how’
/ae, áe/	<i>éfa.e</i> ‘he ate’, <i>exalá.en</i> ‘collapsed’, <i>tšebetšiná.en</i> ‘later’, <i>erlá.epsa</i> ‘I understood’, <i>(h)a.éts</i> ‘so, thus’, in loanwords: <i>jaša.évo</i> ‘live’, <i>ará.eva</i> ‘I sought’, <i>bašlá.epsa</i> ‘I started’
/éa/	<i>kréas</i> ‘meat’, <i>stéas</i> ‘tallow’, <i>avléa</i> ‘courtyard’, <i>pođéa</i> ‘apron’, <i>eksumeré.a</i> ‘outside’, in loanwords: <i>šéa(da)</i> ‘things’
/éo/	<i>paléo</i> ‘old’
/úo/	<i>akú(ɣ)o</i> ‘hear’

In addition to the list of possible vowel sequences above, the vowel sequence of /ua/ may occur in loanwords, e.g., *mu(ɣ)alími* ‘teacher’. Furthermore, sequences of /iu/ and /ou/ are nearly always separated by the fricative /ɣ/, e.g., *đíɣune* ‘they give’; *tróɣune* ‘they eat’, *foyúndane* ‘they fear’, which only in part (i.e., in some verbs) mirrors an underlying fricative in the stem and is otherwise a phonological feature (cf. Bortone 2009: 85).

The following vowel sequences do not occur (at least not without inserted fricative /ɣ/): /eu/, /iu/, /oe/, /oa/, /ue/, /ou/ (irrespective of their stress pattern). According to Mackridge (1987), the masculine and feminine definite articles *o* and *i* are dropped before vowels (a fact which cannot be fully verified by the Romeyka corpus; cf. Fn. 26); otherwise, also the following vowel sequences occur across word boundaries: /oa/, *o=árgos* ‘the bear’; /ou/, *o=uranós* (TD) ‘the sky’; /oe/, *o=eništés* ‘the brother-in-law’; /iu/ *i=úva* (TD) ‘khaki’.

Sequences of the same vowel exist (with different stress patterns; except for /uu/) although they are mostly separated by /j/ or /ɣ/:

/aa/	<i>éfa(ɣ)a</i> ‘I ate’
/ii/	<i>i.đí(j)in</i> ‘the two’
/oo/	<i>tró(ɣ)o</i> ‘eat’

Vowel sequences of /éé/ may occur in Turkish loanverbs integrated with the suffix *-évo*, e.g., *demle(j)évo* ‘brew’. Sequences of three vowels may occur in plurals, e.g., *faía* ‘breads’, in

<sup>29</sup> It is not clear whether occurrences of /ie/ in some pronouns and local adverbials like *havudiés* ‘this’, *itšiéka* ‘there’ are actual vowel clusters and not simply the result of a phonological process inserting a glide /j/ before /e/ (see Section 2.3.6).

loanwords: *šéia*(/šeja) ‘things’; and in integrated Turkish loanverbs with *-évo*, e.g., *hazirlaiévame* ‘we prepared’, *kolaiéve* ‘he watches out’.

In general, vowel sequences in loanwords are realized differently than in Turkish, e.g., Turkish disyllabic lexemes may become monosyllabic [/ajɪ/→/ai/], Tr. *çayır* > *tše(j)íri* ‘meadow’ or *tšahiræ* ‘meadows’; and Turkish monosyllabic lexemes may become disyllabic [/aj/→/aji/], Tr. *çay* > *tšáji* ‘tea’, Tr. *kolay* > *xoláji* ‘easy’.

### 2.2.2 Consonant clusters

Combinations of consonants are frequent in Romeyka. Table 3 shows the possible combinations of two consonants; examples are provided in the list below. It is distinguished between genuine consonant clusters, i.e., those within the onset or coda of the same syllable, and cross-syllabic clusters, which are almost unconstrained. The first consonant is given in the head row. Genuine consonant clusters are presented in the table with their phonemes. Examples are available for both genuine consonant clusters and cross-syllabic clusters (only indicated by an example number in Table 3). Light grey shaded combinations are possible in Turkish (or Arabic) loanwords only. *Geminates* (dark grey areas) are not attested for Romeyka although consonant length was distinctive in AG and some inherited geminates still reflect in the pronunciation, e.g., the quantifier *ólon* ‘whole’ vs. *ál(l)on* ‘other’,<sup>30</sup> or transliteration of Greek orthography, e.g., *ylóssa* ‘tongue’ (see also Neocleous 2020: 31; Mackridge 1987: 123). In general, it has to be noted that there is a striking asymmetry between syllable onsets and syllable codas with regard to the tolerance of consonant clusters. Romeyka is much more tolerant of clusters in syllable onset than in codas. Furthermore, the ability to occur in a syllable coda is a good diagnostic for recognizing affricates – a notorious endeavour – as in the case of /ts/ (see below).

<sup>30</sup> According to P. Mackridge (p.c.), also the quantifier *úlon* is, when used emphatically, in Saráchos (Uzungöl) often pronounced with a geminate consonant contrary to the general principle of PG lacking geminates.

**Table 3:** Consonant clusters in Romeyka

		↓																	
		Plosives			Nasals		Sonorants			Fricatives									
		/p/	/t/	/k/	/m/	/n/	/r/	/j/	/l/	/f/	/v/	/θ/	/ð/	/s/	/z/	/ʃ/	/ɣ/	/χ/	/h/
Plosives	/p/				mp <sup>15</sup>	17	23							sp <sup>46</sup>				59	
	/t/					nt <sup>18</sup>	24			ft <sup>40</sup>				st <sup>47</sup>		53		χt <sup>60</sup>	
	/k/	<sup>1</sup>				19	25			41				sk <sup>48</sup>	51	ʃk <sup>54</sup>			
Nasals	/m/						26		37						52		55		
	/n/			11	16		27	36										61	
Sonorants	/r/	pr <sup>2</sup>	tr <sup>7</sup>	kr <sup>12</sup>						vr <sup>42</sup>	θr <sup>44</sup>	ðr <sup>45</sup>					56	χr <sup>62</sup>	
	/j/	pj <sup>3</sup>	tj <sup>8</sup>			nj <sup>20</sup>	28		38				49						
	/l/	pl <sup>4</sup>		kl <sup>13</sup>						vl <sup>43</sup>							ɣl <sup>57</sup>	63	
Fricatives	/f/						29		39				50						
	/v/						30										58		
	/θ/					21	31												
	/ð/					22	rð <sup>32</sup>												
	/s/	ps <sup>5</sup>	ts <sup>9</sup>	ks <sup>14</sup>			33											64	
	/z/																		
	/ʃ/	pʃ <sup>6</sup>	tʃ <sup>10</sup>																
	/ɣ/						34												
	/χ/						35												
	/h/																		

<sup>1</sup> <i>liráp.ka</i> ‘Lira.DIM’	<sup>23</sup> <i>por.pá.te.ma</i> ‘walking.NMZ’	<sup>45</sup> <i>ðrómos</i> ‘road’
<sup>2</sup> <i>prin</i> ‘before’	<sup>24</sup> <i>bór.da</i> ‘door’	<sup>46</sup> <i>spíti</i> vs. <i>ospíti</i> ‘house’
<sup>3</sup> <i>pjo</i> vs. <i>pio</i> ‘who, what’	<sup>25</sup> <i>ár.go(s)</i> ‘bear’	<sup>47</sup> <i>stúði</i> ‘bone’
<sup>4</sup> <i>apatšaplán</i> ‘from over there’	<sup>26</sup> <i>or.mí</i> ‘river’	<sup>48</sup> <i>skolétši</i> ‘bug’
<sup>5</sup> <i>psómi(n)</i> ‘bread’	<sup>27</sup> <i>fúr.no(s)</i> ‘frog’	<sup>49</sup> <i>pa.títs.ja</i> ‘beans’
<sup>6</sup> <i>pši</i> (TD) ‘soul’	<sup>28</sup> <i>opsár.ja</i> ‘fish.PL’	<sup>50</sup> <i>fos.forizo</i> ‘shine’
<sup>7</sup> <i>tranó</i> ‘big’	<sup>29</sup> <i>e.mor.fó</i> ‘beautiful’	<sup>51</sup> <i>shéz.gete</i> ‘they shit’
<sup>8</sup> <i>ef.řéme</i> ‘we made’	<sup>30</sup> <i>dor.ván</i> ‘butter tub’	<sup>52</sup> <i>n’éz.ma</i> ‘cow dung’
<sup>9</sup> <i>patsí</i> ‘girl’	<sup>31</sup> <i>ér.θa</i> ‘I came’	<sup>53</sup> <i>íš.tera</i> ‘later’
<sup>10</sup> <i>inétši</i> ‘women’	<sup>32</sup> <i>rðómo</i> vs. <i>ðrómo</i> ‘road’	<sup>54</sup> <i>škílo(s)</i> ‘dog’
<sup>11</sup> <i>ðíkno</i> ‘show’	<sup>33</sup> <i>n=árse</i> ‘you will come’	<sup>55</sup> <i>ɣnoríy.ma</i> (TD) ‘knowing.NMZ’

<sup>12</sup> <i>krúyo</i> ‘hit’	<sup>34</sup> <i>yor.yorás</i> ‘Gorgoras (village)’	<sup>56</sup> <i>ey.ri.gó</i> ‘understand’
<sup>13</sup> <i>klíði</i> ‘key’	<sup>35</sup> <i>bar.xár</i> ‘pasture’	<sup>57</sup> <i>ylitši</i> ‘milk’
<sup>14</sup> <i>ksílo</i> ‘wood’	<sup>36</sup> <i>ebéj.name</i> ‘we went’	<sup>58</sup> <i>(e)ývénó</i> ‘leave’
<sup>15</sup> <i>mbéno</i> vs. <i>embéno</i> ‘enter’	<sup>37</sup> <i>al.méyo</i> ‘milk’	<sup>59</sup> <i>éx.base</i> ‘it broke off’
<sup>16</sup> <i>lim.ni</i> (TD) ‘lake’	<sup>38</sup> <i>xul.jéri</i> ‘spoon’	<sup>60</sup> <i>xtíno</i> ‘cow’
<sup>17</sup> <i>is.tan.bó.li</i> ‘Istanbul’	<sup>39</sup> <i>aðel.fó</i> ‘brother’	<sup>61</sup> <i>dóx.na</i> ‘what’
<sup>18</sup> <i>ndo</i> ‘what’	<sup>40</sup> <i>ftíra</i> ‘louse’	<sup>62</sup> <i>xróno</i> ‘year’
<sup>19</sup> <i>fén.go(s)</i> ‘moon’	<sup>41</sup> <i>kríf.ko</i> ‘hide’	<sup>63</sup> <i>ax.la.ðía</i> ‘hay’
<sup>20</sup> <i>n’ézma</i> ‘cow dung’	<sup>42</sup> <i>vroší</i> ‘rain’	<sup>64</sup> <i>ílaxsu</i> ‘bark.IMP’
<sup>21</sup> <i>án.θro.bo(s)</i> ‘man’	<sup>43</sup> <i>a.vléa</i> ‘spot outside’	
<sup>22</sup> <i>kán.ðio</i> ‘some’	<sup>44</sup> <i>án.θro.bo(s)</i> ‘man’	

All genuine consonant clusters presented in Table 3 occur in word-initial position. Romeyka allows genuine consonant clusters consisting of maximal three consonants mostly in syllable-initial position, e.g., *stroma* ‘bed’ (see also Drettas 1997: 72). Frequently occurring word-initial consonant clusters are combinations of fricatives and plosives (and reverse), and fricatives/plosives and sonorants. Mackridge (1987: 123) explains that by the fact that two voiceless fricatives do not co-occur and either of them is rendered into a stop, e.g., */\*fθ/* > */ft/*, */\*fx/* > */fk/*, */\*xθ/* > */xt/*, */\*sθ/* > */st/*, */\*sx/* > */sk/*, */\*fs/* > */ps/*, */\*sf/* > */sp/*.

The consonant clusters treated as affricates (phonemic and allophonic) in Section 2.1.2 above (i.e., */ps/*, */ks/*, */ts/*, *[tʃ]*, *[dʒ]*) appear also frequently in onsets in word-medial position, together with genuine consonant clusters like */sp/* (*as.pro* ‘white’), */tr/* (*ás.dra* ‘stars’), */θr/* (*án.θrobo* ‘man’), and clusters resulting from insertion of the semi-vowel */j/* (*ef.þés* ‘you make’).

In codas in word-final position, consonant clusters are very restricted (but cf. Drettas 1997: 67, 69, 73): While */ts/* occurs in some codas (*aéts* ‘like this’, *pa.títs.ja* ‘beans’), combinations of plosives and fricatives (*/st/*, */ts/*, */tʃ/*, */ps/*) may otherwise only occur in the coda of pronouns and nouns after apocope of their ending (*emíst(in)* ‘we’, *bats(i)* ‘girl’, *inétš(i)* ‘women’, *apops(e)* (TD) ‘tonight’). Clusters ending in a voiced stop or fricative (*/rθ/*, */rs/*, */ft/*) occur after deletion of verbal endings, i.e., often after reduction of the 3<sup>rd</sup> person suffix, e.g., *skáft* ‘s/he digs’.

### 2.2.3 Syllable structure and word shapes

The attested maximum number of segments in a single syllable is four, with a potential number of five only arising in the case of a dubious consonant cluster, which is, if occurring in a coda, likely to be an affricate, thus a single segment, e.g., *stétš* ‘s/he stays’. The minimum number of segments is one vowel or diphthong in the nucleus, e.g., *aé* ‘like this’, which is at the same time the smallest syllable forming a word. The onset can be empty, filled with one consonant or a cluster of up to three consonants. The coda can be empty as well, filled by a single consonant or a cluster of two. Although complex margins are allowed, consonant clusters are much more frequent in the onset and restricted in the coda (see Section 2.2.2).

The following canonical syllable structure rule maximally applies:

$$\sigma \rightarrow (C)(C)V(C)(C) \quad skáft \text{ ‘dig.3SG’}$$

Other possible rules include:

$\sigma \rightarrow V$	<i>i</i> ‘the.F’
$\sigma \rightarrow (C)V$	<i>na</i> .PRT
$\sigma \rightarrow V(C)$	<i>an</i> ‘if.COND’
$\sigma \rightarrow (C)V(C)$	<i>met</i> ‘with’, <i>hats</i> ‘like this’
$\sigma \rightarrow (C)(C)V$	<i>o.psé</i> ‘yesterday’
$\sigma \rightarrow V(C)(C)$	<i>ost</i> ‘until’
$\sigma \rightarrow (C)V(C)(C)$	<i>e.míst</i> ‘we’
$\sigma \rightarrow (C)(C)V(C)$	<i>prin</i> ‘before’, <i>stétš</i> ‘s/he stays’
$\sigma \rightarrow (C)(C)(C)V$	<i>stró.ma</i> ‘bed’

Despite (partial) retention of historical noun endings, Romeyka seems to have a tendency for open syllables, e.g., *pró.va.do(n)* ‘sheep’ (CCV.CV.CV(C)) (see also Holton et al. 2019: 125 for MedGr; but cf. Drettas 1997: 50, who states that closed syllables are more numerous in PG). The most frequent syllable pattern is CV, followed by CVC, and CCV in word-initial position. Notably, Romeyka has – unlike SMG – retained word-initial unstressed vowels (/a/, /e/, /i/, /o/, /u/) in nouns and verbs which form a syllable on their own (Mackridge 1987: 122; see Section 2.3.5.1).

The syllable onset can be filled by all consonant phonemes. For possible consonant cluster in the onset see Table 3 in Section 2.2.2 above. The coda can be filled by all consonants except for [h] but most occur very rarely and/or secondarily as a result of some phonological process, often involving apocope (e.g., /ð/, /θ/, [j], [ʃ], [tʃ], /z/, /x/; see also Özkan 2013: 142).

Word-finally, only /n/ and /s/ occur with some frequency (cf. Holton et al. 2019: 124 for MedGr). Consonant cluster in the coda are dis-preferred (Section 2.2.2) and appear only secondarily after apocope, e.g., *ártš(i)* ‘bears’. Even in this case, the assumed maximum of two consonants in onset or coda can still be maintained by considering /tʃ/ an affricate, hence a single segment, yielding a VCC syllable.

As for frequent word patterns in Romeyka, a tentative hierarchy of word length according to word class is presented in Table 4 below.

**Table 4:** Word patterns in Romeyka according to word class

Syllable no	Word class(es)	Examples
1	determiners; most prepositions; particles; few inflected verbs of high frequency; few pronouns; few adverbs	<i>to</i> ‘the.N’, <i>i</i> ‘the.F’; <i>met</i> ‘with’, <i>s-</i> ‘at, in’; <i>as</i> .OPT, <i>na</i> .PRT, <i>an</i> .COND; <i>léj</i> ‘s/he says’, <i>e(n)</i> ‘s/he is’; <i>ul</i> ‘all’, <i>mis</i> ‘we’; <i>kal</i> ‘again’, <i>até</i> ‘like.this’
2	most adverbs; most nouns; few prepositions; many verbs in -o; some pronouns; some adjectives	<i>bolá</i> ‘very’, <i>móno</i> ‘little’; <i>batsí</i> ‘girl’, <i>ðrómo(s)</i> ‘road’; <i>ómon</i> ‘like’; <i>éxo</i> ‘have’, <i>páyo</i> ‘come’; <i>témon</i> ‘my’, <i>kaínas</i> ‘nobody’; <i>mikró</i> ‘small’, <i>kaló</i> ‘good’
3	verbs in -me, some verbs in past form; most pronouns; most adjectives;	<i>káhume</i> ‘sit’, <i>eðóka</i> ‘I gave’; <i>eména</i> ‘me’, <i>avúdo</i> ‘this’; <i>emorfó</i> ‘beautiful’, <i>yómado</i> ‘full’;

	many nouns; some adverbs	(j)inéka ‘woman’, <i>ánthropo</i> ‘man’; <i>manaxós</i> ‘alone’, <i>andáma</i> ‘together’
4	many inflected verbs in past form; full (plural) pronouns; few nouns; few adjectives	<i>eθékane</i> ‘they put’, <i>katurésis</i> ‘you peed’; <i>teméteron</i> ‘our’, <i>emístine</i> ‘we’; <i>aksinári</i> ‘sickle’; <i>mukútiko</i> ‘tiny’
5	verbs in past plural inflection;	<i>exórepsane</i> ‘they danced’, <i>ekolíyane</i> ‘they burned’;
6	complex adverbials of time/place;	<i>tšepetšenáhen</i> ‘thereafter’; <i>apatšepérandžes</i> ‘from opposite there’
7	few verbs in (full) past plural inflection	<i>evréxumunestine</i> ‘we got wet’
8	some integrated loan verbs in plural	<i>etoplaniéfkumunestine</i> ‘we collected’
9	inflection	

## 2.3 Phonological processes and morphophonology

This section provides an overview of the most striking phonological phenomena in Romeyka; it is rather not intending to provide an exhaustive description of the many phonological processes that have shaped Romeyka throughout its diachrony.

### 2.3.1 Palatalization and affrication

Unlike in SMG, in Romeyka Hellenic /k/ is palatalized to /tʃ/ before front vowels in word-initial position, e.g., AG *κύριος* > *tšíri* ‘father’, AG *κοιμώμαι* > *tšimúme* ‘sleep’ (Tursun 2019: 33), *ke* > *tše* ‘and’, in onsets, e.g., *eki* > *etši* ‘there’, *eðóke* > *eðótše* ‘s/he gave’, and in word-final position where deletion of a front vowel took place, e.g., AG *οὐχι* > *utš* ‘not’, and especially in 3<sup>rd</sup> person verb forms, *stéki* > *stétš* ‘s/he stays’. The phonological rule is summarized:

/k/ → /tʃ/ / [ \_e, i]

This process is prevented by a preceding /s/ which becomes [ʃ], like in *škilos* ‘dog’ (Mackridge 1999, 1987). Furthermore, according to Tursun (2019: 36), this change has not taken place in Romeyka of Tonya, probably due to the stronger Christian Pontic Greek influence. Where applied, it also affects Turkish loanwords, e.g., Tr. *tüfek* > *tufédži* ‘gun’. Also note cases of affrication of plosives in integrated Turkish loanwords, e.g., Tr. *düz* > *džúzi* ‘even, plane’, although the reverse occurs as well, Tr. *çetin* > *kétin* ‘hard’.

Another striking palatalization process is /ɣ/ → /j/ / [ \_e, i] (see Sections 2.1.2.4 and 2.3.6, the latter for a variety of palatalization phenomena). Neocleous (2020: 32–33) also mentions palatalizations of the velar fricatives /g/ and /ɣ/. In the present data, /g/ is palatalized to [j] / [ \_œ], e.g., [‘jœni] ‘skin’. Other affrication processes take place in the formation of the aorist past tense: /z, ɣ/ → /ks/ / [o\_a], e.g., *voízo* ‘call’ > *evóksa* ‘I called’ (03\_30062019F\_11; 065/066), *alméyo* ‘milk’ > *élmeksa* ‘I milked’; and /f, v/ → /ps/ / [a\_i, e\_a], e.g., *skáfo* ‘dig’ > *eskapsídi* ‘you.PL digged’ (03\_07072019F\_1; 39), *xorévo* ‘dance’ > *exórepsame* ‘we danced’. This also applies to Turkish loan verbs integrated with -evo, e.g., *xujévame* ‘we study’ > *(e)xujépsame* ‘we studied’, *aráevane* ‘they seek’ > *eráepsane* ‘they sought’.



### 2.3.2 Assimilation

Conditioned assimilation takes place in the formation of the allophone [ŋ] before velar stops:  $n \rightarrow \eta / [_g, k]$ , e.g., /fěngo/ > ['fengo] 'moon' (see also Section 2.1.2). Another common nasal assimilation process occurs before bilabial stops,  $n \rightarrow m / [p, b]$ , e.g., [istam'boli] 'Istanbul'.

Furthermore, there are some distant assimilation phenomena probably related to labialization or some kind of Umlaut phenomenon, which resemble in part Turkish vowel harmony. For example, there appears interesting variation between the front and back high vowels /i/ and /u/ which is so far only attested for a single, albeit frequent, word: *tšíri* vs. *tšúru* 'father'. It appears that distant assimilation in genitive NPs after the genitive determiner *tu* triggers vowel assimilation in *tšíri* (3).<sup>31</sup>

- (3) a. *érθen*            *o*            *tšír*    =*im*  
           come.AOR.3SG the            father =POSS.1SG  
           'My father came.' (02\_21042018M\_2; 22)
- b. *du*            *tšur*    =*úm*            *to*    *merós*  
           the.N/M.GEN father =POSS.1SG the    side  
           'my father's place' (02\_02022015F\_1; 023)
- c. *du*            *pontól* =*im*            *din*            *džébi*  
           the.N/M.GEN trouser =POSS.1SG the.F.ACC pocket  
           'my trousers' pocket' (08\_04072019M\_1; 185)

This process resembles Turkish vowel harmony in that assimilation goes from left to right. However, it differs in that the grammatical element (the genitive marker) in (3b) is invariable, while the vowel of the lexical item adapts to it. In Turkish, this is reversed whereby vowels in lexical items are invariable, while the vowels of grammatical suffixes adapt. But, as in Romeyka the following possessive suffix is also affected, this process resembles indeed Turkish vowel harmony. The phenomenon is possibly related to the backing of standard Turkish /i/ in the Trabzon Turkish dialect, e.g., Tr. *fırın* > TTr. *furun* 'furnace', which also often takes place after a velar stop (Brendemoen 2002: 186).

A second phenomenon, whose analysis is not fully clear either, is the allomorphic form of the preposition *me(t)* 'with, which occurs predominantly before front vowels in the determiner as *me*, but before back vowels (but also /a/) often as *mo*. This might be related to labial environments. Incidentally, there are also examples with *mi* before high front vowels and *ma* before a low front vowel in the determiner. However, this variation cannot be considered a phonological rule as there exist quite a number of exceptions (for examples and further discussion see Section 3.2.5).

Distant assimilation also occurs in the conjunction *tše* 'and' in one speaker, whereby the final vowel seems to assimilate to the following determiner (4).<sup>32</sup>

- (4) a. *ebíyane*    *o*            *džíri*    *dži*    *i*            *mána*    *dže*    *da*            *yarðélæ*  
           go.AOR.3PL the            father and    the    mother and    the.PL    children  
           'They came, the father, the mother, and the children.' (04\_01072019F\_12; 49–50)

<sup>31</sup> Mackridge (n.d.) also notes the rare GEN.SG *tšuru* 'of the father' in Saráchos (Uzungöl) (also Asan 1998: 292 for Çoruh), whereas PG as spoken in Greece has *tšíru*.

<sup>32</sup> The conjunction *tše* 'and' seems also subject to a simplification in one speaker whereby [tʃ] is simplified to /s/: *epéjna pérena da xuljéra se do tási* 'I went and took the spoons and the cup.' (07\_04072019F\_6; 42).

- b. *i mána tš-ø o tširi tše ta yarðélæ*  
 the mother and the father and the.PL children  
 ‘the mother, the father, and the children’ (04\_01072019F\_12; 08)

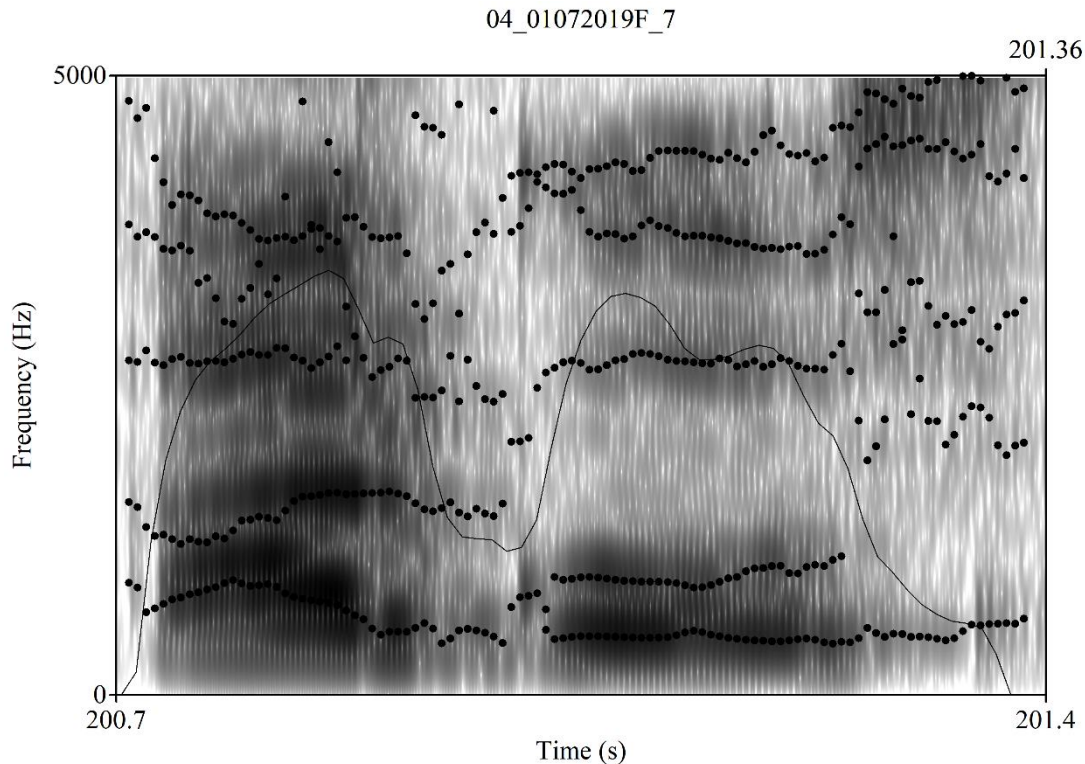
There is inconclusive evidence for the existence of final devoicing in Romeyka: (i) As outlined in Section 2.2.32.2.1, all consonant phonemes can occur in word-final position although voiceless consonants are more frequent and voiced phonemes occur mainly after apocope of verb or noun endings, e.g., *apíð* ‘pear tree’ (02\_02022015F\_1; 166), *arθób(o(s))* ‘man’ (05\_03072019M\_4; 02), *vráð* ‘it rains’ (08\_04072019M\_3; 127), especially in coastal dialects where apocope of noun endings is most advanced: *stúð* ‘bone’, *karið* ‘walnut’. (ii) Since voice contrasts with stops are undetermined in Romeyka, voice in word-final consonants is difficult to discern, e.g., [gæzle’ dev **ð**=’opsaræ] ‘he watches the fish’ (04\_01072019F\_12; 40), [’škiz **ða** ’ksila] ‘he splits the wood’ (04\_01072019F\_13; 39). (iii) Voicing of word-final consonants is determined by the environments in which they occur: adjacent to a voiceless consonant in the onset of the following word, word-final consonants tend to be voiceless as well, e.g., *okujéf so ido en* ‘she studies, it is in, say..’ (07\_04072019F\_5; 36). Before vowels, word-final consonants appear voiced as well, e.g., *so mektéb ebéjname* ‘we went to school’ (08\_904072019M\_1; 016) (see also the paragraph on intervocalic voicing below). Still, there exist some examples of final devoicing in the corpus, e.g., *do vákit ðævás* ‘time passes by’ > *ðævázo* ‘pass (by)’ (03\_30062019F\_7; 10). Also, Tursun (2019: 31) notes word-final devoicing in Romeyka for the voiceless alveolar fricative /s/.

A lenition process in intervocalic environments, intervocalic voicing, is also difficult to establish as it is not applied systematically (although common in the contact language Turkish). An example is the weakening of the alveolar stop /t/, e.g., *ospí(t(i(n)))* ‘house’ > *ospíðæ* ‘houses’, and the bilabial stop, e.g., *lébe éliyam=a* ‘we called it Lep (a children’s game)’ (04\_01072019F\_2; 017/018), but *ta za fásame* ‘we fed the animals’ > *faízo* ‘feed’ (03\_30062019F\_2; 27). It is not clear, whether the undetermined voice distinction in Romeyka stops is related to this, see ex. (5) below where the evolution of the adverbial is *ekí* > *etši* > *etš/edž* and it is unclear what triggers its voiced realization.

- (5) *ep-édž*                                      *ebíyamist*  
 from.there                                      take.AOR.1PL  
 ‘We took (them) from there.’ (03\_30062019F\_2; 29)

### 2.3.3 Final vowel lengthening

In Romeyka, vowels are lengthened in word-final position (Figure 4 exemplifies the lengthening of the second vowel in comparison to the first vowel). Final vowel lengthening takes in particular place in phrase-final position (exs. 6–8; also, in enumerations like 08\_04072019M\_1; 094/220; Özkan 2013: 139). It occurs as a tendency for all speakers and irrespective of clause types. This means at word-level that the vowel is realized longer in word-final position than word-medially and at phrase level, that the vowel is realized longer in phrase-final position than in any other position (cf. also Section 2.4.2). There is no phonemic length distinction, though. Lengthening of vowels occurs both in open and closed syllables (for the latter see exs. 7, 8). The phenomenon is probably reinforced by word/sentence stress. For the impact of different (consonantal) environments on vowel lengthening, more phonological research is required.



**Figure 4:** Vowel lengthening in a closed unstressed syllable (word *árgo:s* ‘bear’)

(6) (Examples from 04\_01072019F\_7)

- a. Word-final /a/ in *mána* ‘mother’ (no. 045), a<sub>1</sub>: 0.1686sec vs. a<sub>2</sub>: 0.2496sec
- b. Word-final /e/ in *alévre* ‘flour’ (no. 171), a: 0.0945sec, e<sub>1</sub>: 0.1367sec vs. e<sub>2</sub>: 0.2303sec
- c. Word-final /i/ in *díri* ‘cheese’ (no. 181), i<sub>1</sub>: 0.1317sec vs. i<sub>2</sub>: 0.3607sec
- d. Word-final /o/ in *xróno* ‘year’ (no. 398), o<sub>1</sub>: 0.1846sec vs. o<sub>2</sub>: 0.2051sec
- e. Word-final /u/ in *kátu* ‘below’ (no. 504), a: 0.1633sec, u: 0.1913sec, cf. *stúði* ‘bone’ (no. 023), u: 0.1596sec, i: 0.2240sec
- f. Word-final /o/ in closed syllable *árgo:s* ‘bear’ (no. 155), a: 0.1449sec, o: 0.1922sec; cf. unstressed word-initial /o/, *opsá:ri* ‘fish’ (no. 157), o: 0.0918sec, a: 0.2168sec, i: 0.1866sec; cf. unstressed word-final /o/ in *féngo* ‘moon’ (no. 273), e: 0.0664sec, o: 0.1519sec

- (7) *abohéndžeka*                      *stédži:s*  
 from.where.side                      stay.PRS.2SG  
 ‘Where are you staying?’ (01\_15022015F\_1; 04)  
 i: 0.3389sec

- (8) *to bój =im mikró e:n amá to jáš =im bolá e:n*  
 the size =POSS.1SG small be.PRS.3SG but the age =POSS.1SG much be.PRS.3SG  
 ‘My size is small, but my age is much.’ (01\_04022016F\_1; 006/007)  
 e<sub>1</sub>: 0.1228sec, e<sub>2</sub>: 0.1997sec

### 2.3.4 Variation in fricatives

In the following, some sporadic phonological variation involving fricatives is outlined whose conditions are not clear yet.

Firstly, there is variation in fricatives which is likely dialectal but also appears within individual speakers. Its conditions are yet to be determined. Variation exists, for example, between the labiodental and the velar fricative: *fovúndane* (3 tokens in the Romeyka corpus) vs. *foýúndane* (2 tokens) ‘they fear’; *đíyune* (8 tokens) vs. *đívune* (1 token) ‘they give’ (see also Özkan 2013: 140).

Furthermore, [ʒ] occurs as a variant of both [ʃ], e.g., *dréž(i)s* vs. *tréšis* ‘you run’, and /ð/, e.g., *hažá* vs. *haðá* ‘here’ (see Section 2.1.2).

Variation in other fricatives appears in comparison with examples from Tursun (2019): /θ/ vs. [h], e.g., *káθome* (TD) vs. *káhome* ‘sit’; /χ/ vs. /ɣ/, e.g., *xarízo* vs. *ɣarízo* (TD) ‘make happy’; and /v/ vs. /ð/, e.g., *véntro* (TD) vs. *đéntro* ‘tree’. In order to determine the nature of this, probably dialectal, variation, further variational research is required.

In addition, Özkan (2013: 140–141) reports variation between /θ/ and /f/ in word-initial and -final position, e.g., *galáf* vs. *galáθ* ‘basket’, as well as occasional variation between [h] and velar [x], e.g., *har* vs. *xar* ‘now’, *gáhume* vs. *gáxume* ‘sit’.

Another variation (with a single occurrence in the corpus) is found in word-final 3<sup>rd</sup> person verb inflection of *páo/epíya* ‘go/went’, where [j]+e is realized as a palatal fricative [ç], *epíje* > [e’piç] ‘he went’ (07\_04072019F\_8; 05). Further – potentially idiolectal – variation occurs in the affricates [tʃ] (or [dʒ]), which can be realized before /e/ (attested for the word *tšéka* ‘side’) as a sequence of /d/ + glide [j] or a diphthong /ie/, e.g., *avudžés* > *avudjés* ‘around here’ (08\_04072019M\_3; 133), *itšiéga* ‘there’ (08\_04072019M\_3; 125).

### 2.3.5 Elision

#### 2.3.5.1 Apheresis

Unlike in SMG and other Greek varieties, initial unstressed vowels are retained in Romeyka verbs, e.g., *epéro* ‘take’ vs. SMG *παίρω* ‘take’, and nouns, e.g., *ospíti* ‘house’ vs. SMG *σπίτι* ‘house’ (exs. from Mackridge 1987: 122; Bortone 2009: 83).

These unstressed word initial vowels are likely to be dropped in adverbs and verbs (less often in nouns) in order to avoid a hiatus, if the preceding word in the same phrase ends in a vowel, e.g., *(a)pés* ‘inside’, *(e)líyon* ‘a little’, *(e)péro* ‘take’ (see also Section 2.3.5.3). Also, in aorist verb forms, the unstressed augment /e/ may be dropped after vowels in the onset of the preceding word. When unstressed initial vowels are deleted in verbs, complex onsets may arise, e.g., *(e)ftiéyo* (TD) ‘do’, *(e)ɣvéno* (TD) ‘leave’ (Tursun 2019: 203), cf. Section 2.2.2.

#### 2.3.5.2 Apocope

In some varieties of Romeyka, post-tonic (i.e., the syllable following the stress-bearing syllable) /i/ and /u/ are retained whereas they are dropped in others (ex. 9; Mackridge 1987: 122, 131; Özkan 2013: 142). There is, however, also intradialectal variation; for example, in ROF as spoken in Çaykara all forms exist: *ospí(d(i(n)))* ‘house’. The phonological conditions governing the selection of the respective forms are still to be investigated.

- |     |                         |                                     |
|-----|-------------------------|-------------------------------------|
| (9) | a. /ospíti(n)/          | ROF as spoken in Uzungöl (Saráchos) |
|     | b. /ospít/              | RSür as spoken in Beşkøy            |
|     | c. <i>σπίτι</i> /spíti/ | SMG                                 |
|     | ‘house’                 |                                     |

There is significant sub-dialectal variation in the phonological shape of nouns that developed out of AG neuters in *-ion*: In SMG, the /on/ ending is lost, whereas RO<sub>f</sub> as spoken in Uzungöl (Saráchos) retains an older form with /in/, e.g., *omáti(n)* ‘eye’, and RSür as spoken in Beşköy drops the final /n/ (Özkan 2013), e.g., *omáti* ‘eye’ (Mackridge 1987: 132). According to Tursun (2019: 21), the feature of final vowel raising corresponds to the location of the dialect: while the /in/ ending is dropped in dialects closer to the coast, mountain dialects retain the full archaic form. The majority of nouns in the Romeyka corpus (RO<sub>f</sub> as spoken in Çaykara) end in a vowel. The corpus also contains an interesting intermediate form of a speaker originally from a higher village who married to a lower one, exhibiting a nasalized vowel instead of the final /n/: *psómĩ* ‘bread’ (04\_01072019F\_7; 169).

Apocope, i.e., the loss of word-final segments, takes also frequently place in verbs like in the following example of verb serialization, *káhundān-ø tšimúndāne* ‘they sit and sleep’ (04\_01072019F\_12; 30). More frequently, the unstressed word-final 3SG marker /i/ is deleted in some verbs (see also Neocleous 2020: 33). Mackridge (1987: 131) notes the deletion of word-final /i/ in RO<sub>f</sub> as spoken in Uzungöl in certain verbs like *éš(i)* ‘he has’ (for further forms of apocope see Drettas 1997: 101–103). Another feature which is not easy to classify is the elision of word-final /n/ in 3<sup>rd</sup> person verbal endings. There is possibly a relation to the full (or reduced) form of noun endings. Van Emde Boas et al. (2019) refer to this feature in Classical Greek as “movable n”. It is not clear yet what motivates the occurrence of /n/; its distribution in the corpus is relatively balanced:

<i>epíje</i> ‘s/he went’ (5 tokens)	vs.	<i>epíjen</i> (5 tokens)
<i>epídže</i> ‘s/he did’ (10 tokens)	vs.	<i>ebídžen</i> (5 tokens)
<i>érthe</i> ‘s/he came’ (19 tokens)	vs.	<i>érthen</i> (15 tokens)

Note that the reduced form appears even when the verb gets reduced to a single vowel:

<i>e</i> ‘s/he is’ (24 tokens)	vs.	<i>en</i> (116 tokens)
--------------------------------	-----	------------------------

### 2.3.5.3 Crasis

This section deals both with crasis, i.e., the merger of a word-final vowel with the consequent word-initial vowel, and deletion of final vowels of certain function words like prepositions and determiners before vowel-initial nouns.

Mono-syllabic words consisting of a single vowel, like determiners, are often deleted before a vowel in the onset of the following word to avoid a hiatus (ex. 10; Oikonomidis 1908: 77; Mackridge 1987: 122; cf. van Emde Boas et al. 2019: 15–16). Elision of the determiner is especially frequent before human proper names if the preceding word ends in a consonant and realizing the vowel of the determiner would lead to a hiatus (11). Especially in nouns, this phenomenon is difficult to analyse, though, as it may interfere with grammatical aspects such as definiteness or possibly contact-induced omission of determiners. Furthermore, final vowels can be also deleted in function words like the conjunction *tše* ‘and’ (12) and in verb endings with an enclitic object pronoun (13).

- (10) *eš érde ø árgo, aha, ø árgo érde*  
 have.PRS.3SG come.PRS.3SG bear aha bear come.AOR.3SG  
 ‘The bear was coming, aha, the bear came.’ (04\_01072019F\_12; 20/21)

- (11) *tše med ø eminé*  
 and with Emine  
 ‘and with Emine’ (04\_01072019F\_17; 19)

- (12) *tše ne bórum tš= érxumesine*  
 NEG PRT can.PRS.1PL and come.PRS.1PL  
 ‘We cannot come.’ (07\_04072019F\_5; 17)

- (13) *eyó n= eftéy =a*  
 I PRT= do.PRS.1SG =OPN.3SG  
 ‘I will do it.’ (C1)

Actual crasis occurs frequently and regularly for example in the combination of prepositions with determiners and nouns, e.g., *so=spídi* ‘to the house’ or *as=opsárja* ‘from the fish’; the loss of determiners after prepositions being cross-linguistically not an uncommon phenomenon. *So* itself is a merger of the preposition of locality *s* and the neuter determiner *to* whereby the resulting consonant cluster /st/ is reduced to /s/ + /o/; similarly, *asa* > *as ta* ‘from the.PL’ (Mackridge 1987: 123; see also Section 3.2.5). Also, the form of many Romeyka pronouns has historically arisen out of a merger of the determiner with the pronoun, e.g., *téso* ‘your’ > *tu*.GEN *eso*.2SG. Through these mergers, sentences in fast free oral speech can be significantly contracted (cf. Tursun 2019: 22–23). On the other hand, however, crasis seems not to be an obligatory rule as the present Romeyka corpus features some examples where the definite neuter article remains intact, e.g., *as to pislukinen eðevam eka* ‘We suffered from the pissing.’ (02\_09062019F\_1; 08).

#### 2.3.5.4 Reduction /nd/ > /d/

Complex syllable onsets with the consonant sequence /nd/ which occur in question words tend to be reduced to /d/, e.g., *nde* (10 tokens in the corpus) vs. *de* (8 tokens) ‘what’; *ndóxna* (1 token) vs. *dóxna* (4 tokens) ‘what’.

### 2.3.6 Consonant epenthesis

The glide [j] occurs adjacent to vowels in four environments resulting from insertion, synizesis, and palatalization; especially for the latter, prosodic boundaries seem to be more important than word boundaries (cf. Blevins 2008):

(i) in inter-vocal hiatus environments [V\_V], for example, between /ia/, /ie/, where a glide is perceived at the formant transitions between the two vowels, e.g., *havudijés* ‘this’, *itši(j)éga* ‘there’, which is a typologically common sound change (Blevins 2008: 84). In the plural variant *opsárja* ‘fish.PL’ (cf. *opsára*, *opsáræ*), the glide is an indicator of synizesis of the underlying plural ending /ia/. Synizesis also occurs in *xul’éri(n)* ‘spoon’, pl. *da xul’éra* ‘the spoons’ (< MedGr *χουλιάριν* /xuljárin/ ‘spoon’; 07\_04072019F\_6; 42).

(ii) word-initially [\_V] in some nouns, e.g., *(j)éma* ‘blood’, while in other nouns the historical consonant of /y/ has been either lost in total or resulted in the glide according to the allophonic rule /y/ → [j] / [\_e, i], e.g., *(j)inéka* ‘woman’. According to Tursun (2019: 26), these nouns have lost their initial consonant unsystematically in different dialects but retained it in intervocalic environments, for example, after the feminine definite article (but cf. *i inéga=nat* ‘his wife’ (04\_01072019F\_13; 13)).

The glide occurs also word-initially in a prosodic unit if the preceding words ends in a vowel, e.g., *ikí séne jen* ‘it is since two years’ (03\_30062019F\_6; 57), *amá jéla eróda=me* ‘but come and ask me’ (03\_30092019F\_7; 35), *o jálo ba* ‘the other.TOP’ (04\_01072019F\_12; 54).

(iii) in palatalized onsets before /e/, often in word-medial position, and frequently following /t/, /n/, also /l/, e.g., *eftlés* ‘you do’ (03\_07072019F\_1; 38), *en’ésteva* ‘I got tired’ (04\_01072019F\_2; 130), *n’ézma* ‘cow dung’, *n’eró* ‘water’ (04\_01072019F\_7; 250). This

process also takes place when it leads to complex syllable onsets, e.g., *flerotika* ‘wooden mixer’ (07\_04072019F\_6; 09). And it seems to correspond to prosodic units as it even occurs word-initially [e], if the previous word ends in a consonant, e.g., *ándan jérted’e* ‘when you.PL came’ (03\_07072019F\_1; 09). This phenomenon seems to be more dominant in some speakers and is likely subject to idiolectal variation.

A similar palatalization process takes places with the voiced velar plosive (which may be diachronically derived from a voiceless velar plosive): /g/ → [j] before [æ], as in the example of /gʲönksæ/ ‘breasts’. Apparently, this phenomenon extends to Turkish loanwords, such as [kʲoli] ‘lake’ (see also Section 2.1.1. on /æ/).

(iv) Finally, the glide occurs as an allophone of /y/ / [e, i] in certain verb inflections, e.g., the 2<sup>nd</sup>/3<sup>rd</sup> person aorist form of some verbs with a velar fricative in the stem like *páyo* ‘go’, i.e., *epíjes* ‘you went’, *epíje* ‘she went’ (but not *léyo* ‘say’, i.e., *lége* ‘he said’). This change can possibly lead to a susceptible, sporadically occurring sound change in codas at the end of a prosodic unit /je/ → /ç/, e.g., *epíje* vs. *epíç* ‘he went’ (07\_04072019F\_8; 05) (cf. Blevins 2008).

Laryngeal consonant epenthesis of [h] / [a] occurs in word-initial environments of some word classes, namely adverbs of place and manner, and demonstrative and personal pronouns. /h/ seems to be used as a prefix indicating demonstrative emphasis. Note that it only occurs [a]; if prefixed to demonstrative pronouns with /e/ in the onset, this vowel changes to /a/, e.g., *etšinos* > *hatšinos* ‘he.DEM’. Whether /a/ could be actually part of a “demonstrative prefix *ha*” requires further research. Note also the occurrence of word-final *-ha* in some spatial adverbials (see Section 3.1.4.3). The occurrence of word-initial [h] is probably related to prosodic boundaries as the forms with epenthetic [h] seem to occur more frequently in phrase-initial position while forms without [h] seem to dominate in phrase-medial position. Still, the feature occurs not systematically (for some examples see 14) and is probably also driven by a discourse function as it fosters a demonstrative reading (ex. 15).

- (14) a. *havúdo píos en*  
 this who be.PRS.3SG  
 ‘Who is this?’ (03\_07072019F\_1; 07)
- b. *as íxa énan bats =óbon omó haténa hájes treš*  
 OPT have.AOR.1SG a girl =DIM like OPN.3SG like.this run.PRS.3SG  
 ‘If I (only) had a little daughter like her who runs like this.’ (03\_07072019F\_1; 14)
- c. *do tsubaðálevro ø- aídiko*  
 the corn.flour like.this  
 ‘the corn flour like this’ (04\_01072019F\_2; 255)
- d. *sin bárbedi bai hájets*  
 to.the Bayburt go.PRS.3SG like.this  
 ‘It leads to Bayburt like this [gesture].’ (08\_04072019M\_2; 080)
- (15) a. *ído bakáli íxa aðá so spíti*  
 thing grocery.store have.AOR.1SG here at.the house  
 ‘I had a grocery store, here at the house.’ (08\_04072019M\_2; 050/051)
- b. [soon thereafter, after the interviewer did not understand]  
*haðádžega, haðá so spídi*  
 here.side here at.the house  
 ‘Here, here at the house.’ (08\_04072019M\_2; 060)

Word-initial [h] is a regional feature common to both Romeyka and the Trabzon Turkish dialect, where it surfaces as prefix *ha-* (Table 5; see also Brendemoen 2002: 231). According to Tursun (2019: 26), the occurrence of epenthetic [h] depends on the dialect area. Its occurrence may perhaps also be related diachronically to aspirated vocalic onsets in AG (cf. van Emde Boas et al. 2019: 4). Certainly, more research is required on its distribution.

**Table 5:** Word-initial [h] in Romeyka and Trabzon Turkish

Adverbs of place	<i>haḍatšés</i> ‘here around’	Tr. <i>ha-buraya</i> ‘to here’, <i>ha-ordan</i> ‘from there’
Adverbs of manner	<i>hajés</i> , <i>haé</i> ‘like this’	Tr. <i>ha-boyle</i> ‘like this’
Pronouns	<i>hatší</i> , <i>hadéna</i> , <i>hadín</i> ‘this’	Tr. <i>ha-bu</i> ‘this’

### 2.3.7 R-Metathesis

Another common phonological process is r-metathesis which takes place as adjacent and non-adjacent metathesis (see also Oikonomidis 1908: 86). The former type is more frequent and occurs in some nouns with high frequency, e.g., *ḍórmos* vs. *ḍrómo(s)* ‘road’, *árθob(o)/árdabo* vs. *ánθropos* ‘man, human’. It also appears frequently in Turkish loanwords indicating the occurrence of r-metathesis also in the regional Trabzon Turkish dialect (see Brendemoen 2002: 166), e.g., *truší* vs. *turší* ‘pickles’ < Tr. *turşu* ‘pickles’ (04\_01072019F\_1; 056), *krik* vs. *kirk* ‘forty’ < Tr. *kırk* (03\_30092019F\_8; 12). In the case of the latter example, already the next sentence in the recording shows the original form *kirk* ‘forty’ which suggests that the two variants are in free variation. A typical example of distant metathesis is *ormofá* vs. *omorfá* ‘nice’ (03\_07072019F\_1; 34). In the other examples of distant metathesis, either /r/ and /n/ are exchanged, e.g., *dinír* vs. *dírin* ‘cheese’ or the /r/ occurs initially in a plural form whereby it is in the singular in the final syllable, e.g., *rðénde* ‘trees’ (02\_02022015F\_1; 165) vs. *ðéndro* ‘tree’ but cf. also *ðénda* ‘trees’ in the same recording (02\_02022015F\_1; 158).

## 2.4 Prosody

### 2.4.1 Word stress

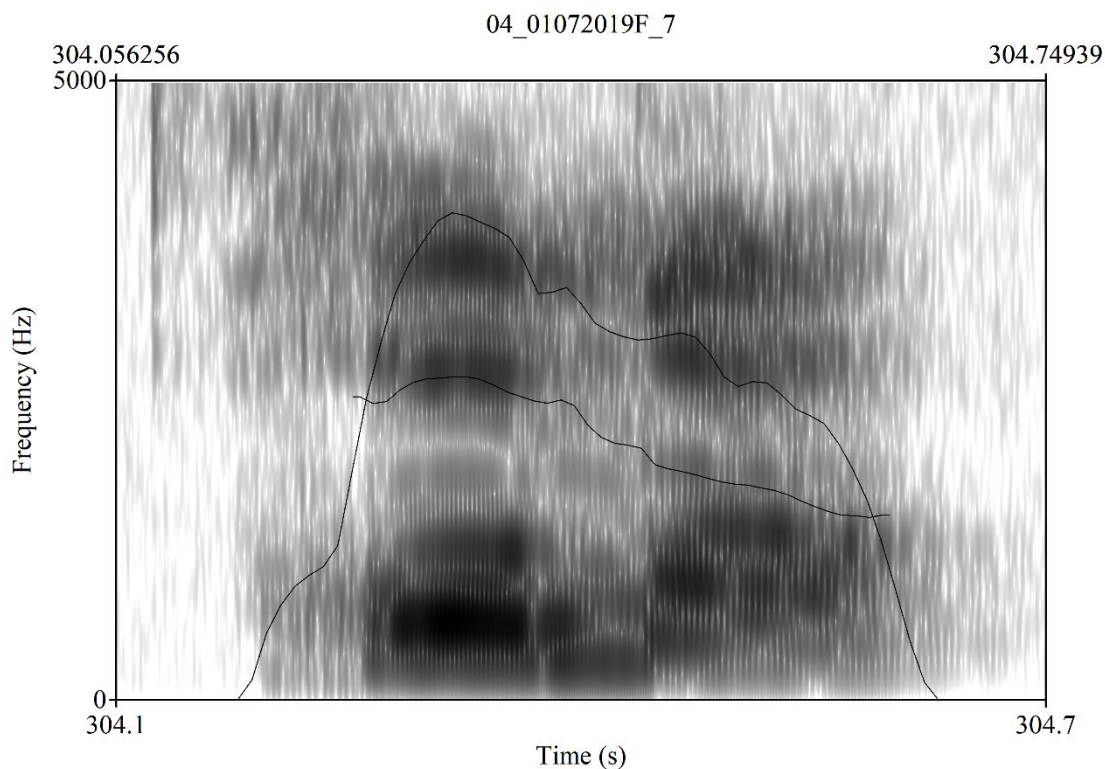
With the present restricted corpus, there is only little evidence for the fact that word stress is distinctive in Romeyka as in other Greek varieties, i.e., leading to a change in meaning, e.g., *kséro* ‘I know’ vs. *kseró* ‘dry’; *íbe*.AOR.3SG ‘she said’ vs. *ibé*.IMP.2SG ‘Say (it)!’ (see also Tursun 2019: 22). Still, Romeyka should be considered to have (dynamic) phonemic stress. Unlike SMG (MedGr, and AG), where stress can only fall at maximum on the antepenultimate syllable (Mackridge 1987: 122; Holton et al. 2019: 224), there is no positional limitation in Romeyka polysyllables and stress can fall until the seventh syllable from the end although it usually does not go farther than the fourth syllable from the end (“columnar stress” in the terminology of Mackridge 1987; Özkan 2013: 143; see Table 6).



**Table 6:** Word stress in Romeyka (poly-)syllables

Syllable no.	Stress pattern(s)	Examples
1	́x, x	<i>zó</i> ‘animal’; clitics, e.g., = <i>mas</i> ‘our.CL’
2	́xx, x́x	<i>ándra</i> ‘husband’; <i>peđí</i> ‘son’
3	́xxx, x́xx, xx́x	<i>đáxtila</i> ‘fingers’; <i>pođári</i> ‘leg’; <i>ađelfó</i> ‘brother’
4	xxxx́, x́xxx	<i>liftokári</i> ‘hazelnut’; <i>katsímale</i> ‘fog’
5	xxx́xx, xxx́xx	<i>ekoliyane</i> ‘they burned’; <i>tšebedženáhen</i> ‘later’
6	xxxx́xx	<i>apadžepérandžes</i> ‘from opposite there’
7	x́xxxxxx	<i>evréxumunestine</i> ‘we got wet’
8	xxxx́xxxx	<i>parišéfkumunestine</i> ‘we cleaned?’
9	xxxx́xxxxxx	<i>etoplańéfkumunestine</i> ‘we collected’

Romeyka is a stress-accent language and stress is expressed by a higher intensity/loudness but seems also to correspond to a higher pitch (cf. AG as pitch-accent language, van Emde Boas et al. 2019: 277) but MedGr where pitch does not play a role (Holton et al. 2019: 222); see Figure 5).

**Figure 5:** Pitch and intensity in the spectrogram of the word *xóma* ‘earth’

Detailed further research is required on the predictors of stress. Which syllable of a word receives stress is in part an unpredictable inherited lexical property (cf. Holton et al. 2019: 225) but apart from this, there are several predictors: Syllable length has probably an effect on the stress pattern (see van Emde Boas et al. 2019: 279 for AG accentuation rules). Phrasal intonation, i.e., the location of the word in the phrase, has probably an effect as well. Also, word

class seems to play a role: verbs can be subcategorized according to ultimate and penultimate stress, the latter verb class forming the major group (Drettas 1997: 205). Stress is only in part affected by morphological predictors like affixation (e.g., diminutives but not comparatives) and usually not by cliticization (cf. Mackridge 1987: 122; also Holton et al. 2019: 225 for MedGr), e.g., *ospí* ‘house’, *ospídae* ‘houses’, *t-ospíde=muna* ‘our houses’ but *ospid-ópo* ‘little house’ (08\_04072019M\_3; 146), cf. also *andraðélfó* ‘brother-in-law’ vs. *andraðélfó=m* ‘my brother-in-law’. On the other hand, stress patterns affect morpho-phonological processes like deletion of unstressed vowels (Section 2.3.5.1) or the development of the plural suffix *-/æ/* in nouns with an underlying unstressed plural in *-/ia/*, e.g., *t=aðélf-ia* > *t=aðélf-æ* ‘the siblings’ vs. *peði-a* ‘children’. According to Mackridge (1987: 122), Romeyka applies “columnar stress” to adjectives and verbs (cf. “persistent accent” in AG, van Emde Boas et al. 2019: 81). This means that affixation does not impact on stress (i) in adjectival inflection, where the stress remains for all inflections in the same position where it is in the masculine nominative singular (“base accent”, van Emde Boas et al. 2019: 281), (ii) in verbal inflection, where stress remains for all tenses in the position of the first person singular (Mackridge 1987: 122). In nouns, the base accent is defined by the masculine nominative singular (Özkan 2013: 143; see also van Emde Boas et al. 2019: 281). AG declension classes determine the base accent of nouns, e.g., *ónema* ‘name’ < AG *ōvoμα* ‘name’ as neuter 3<sup>rd</sup>-declension noun (van Emde Boas et al. 2019: 288). Özkan (2013: 143) reports for RSür that a stress shift occurs in the genitive singular of neuter nouns, which could not be confirmed in the present corpus, though (but cf. van Emde Boas et al. 2019: 282 on AG “case accent”).

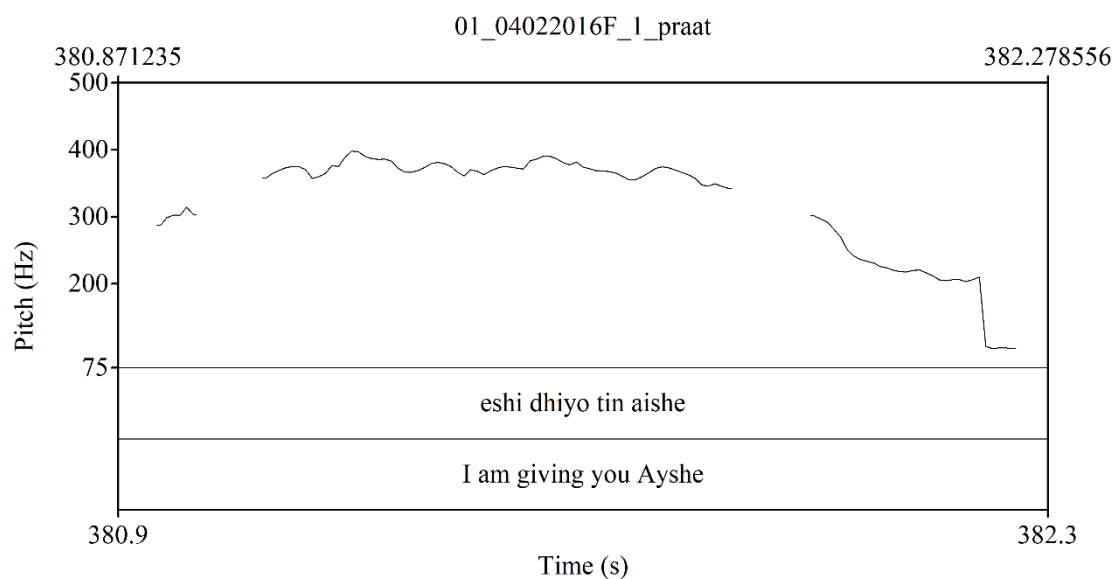
The stress patterns in Turkish loanwords require specific investigation since Turkish loan words appear to have no clear stress pattern, e.g., [*da'ne*] ‘piece’ (02\_02022015F\_1; 020; but cf. *tané* ‘piece’ in Neocleous 2020: 189). For these reasons, word accent is not indicated in Turkish loanwords throughout this section except for few very clear cases.

In compound nouns, the accent seems to fall on the first syllable of the head noun, e.g., *andraðélfó* ‘brother-in-law’ > *ándras* ‘husband’ + *aðélfó* ‘brother’; *tsubaðálevro* ‘corn flour’ > *tsubáði* ‘corn’ + *alévre* ‘flour’; *opserángišo* ‘fishhook’ > *opsári* ‘fish’ + ? (SMG *άγκιστρο* ‘hook’).

## 2.4.2 Sentence stress

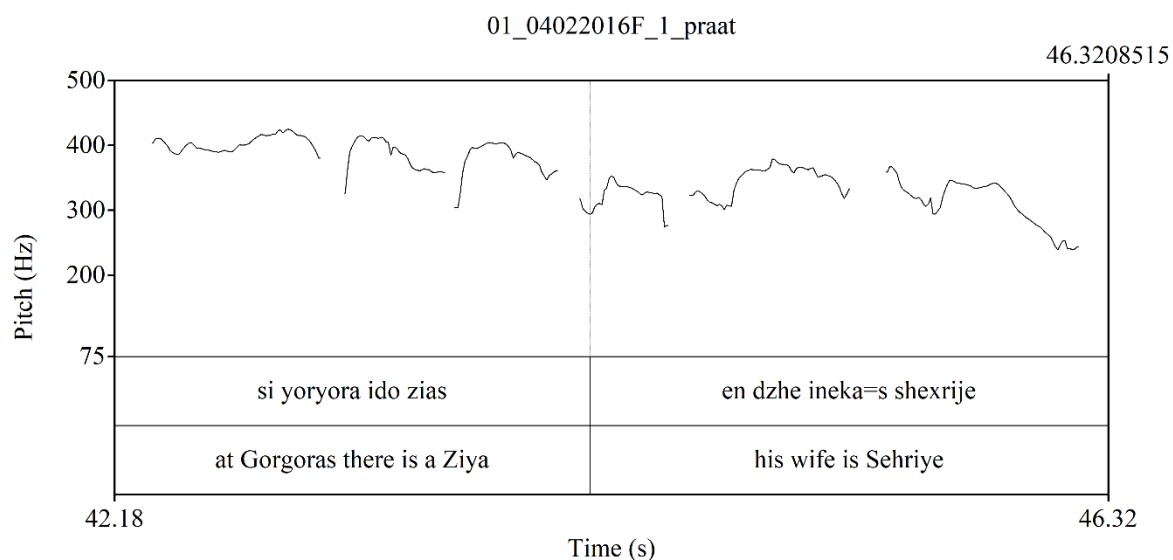
The intonation contour of phrases is linked to the communicative goal (i.e., clause type) and the position of the phonological phrase in an utterance unit (or conversational turn). The general intonation contour is flat first with a final decline, whereby the final syllable is typically lengthened at the end of a speech unit. In the following, typical intonation contours are described for different clause types.

A typical declarative sentence is featured in Figure 6. Its contour is in general relatively flat with a final short decline.



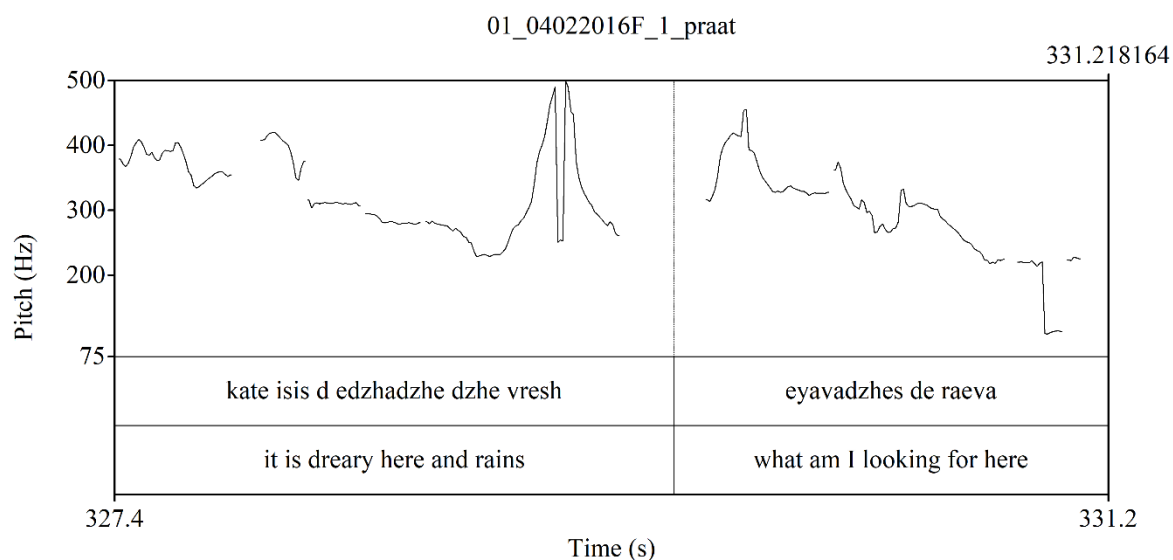
**Figure 6:** Intonation contour of a declarative clause

In an utterance unit consisting of several clauses, the intonation contour is falling at the end of each phrase, whereby the next sentence continues at the new, lower pitch level. At the end of the unit, there is a significant decline (see Figure 7).



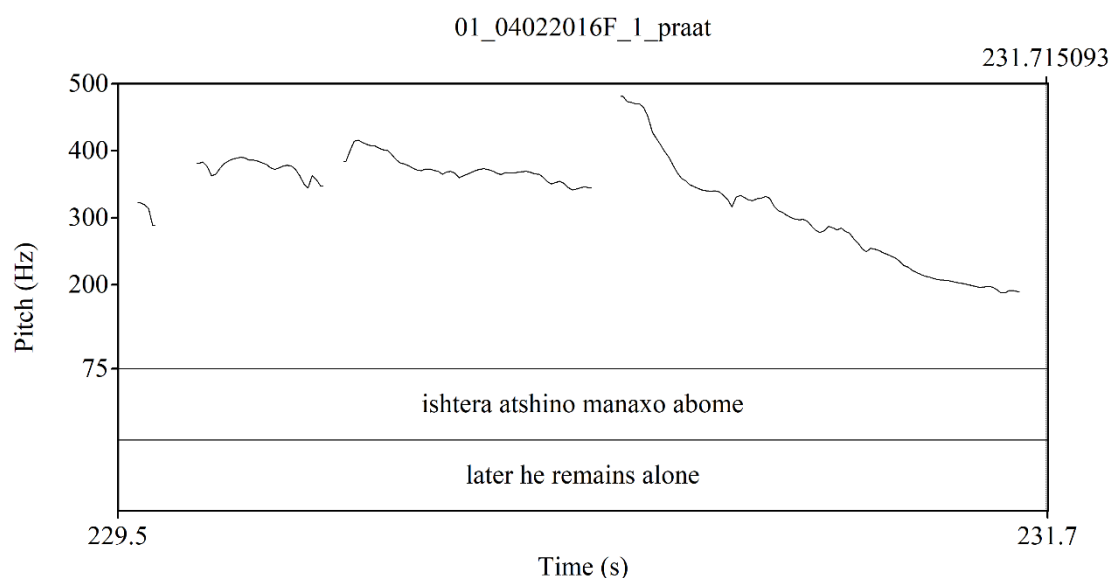
**Figure 7:** Intonation contour of coordinated phrases

The intonation contour highlighting the existence of a subsequent phrase which is part of the same utterance unit, like in enumerations, shows a final rise (Figure 8).



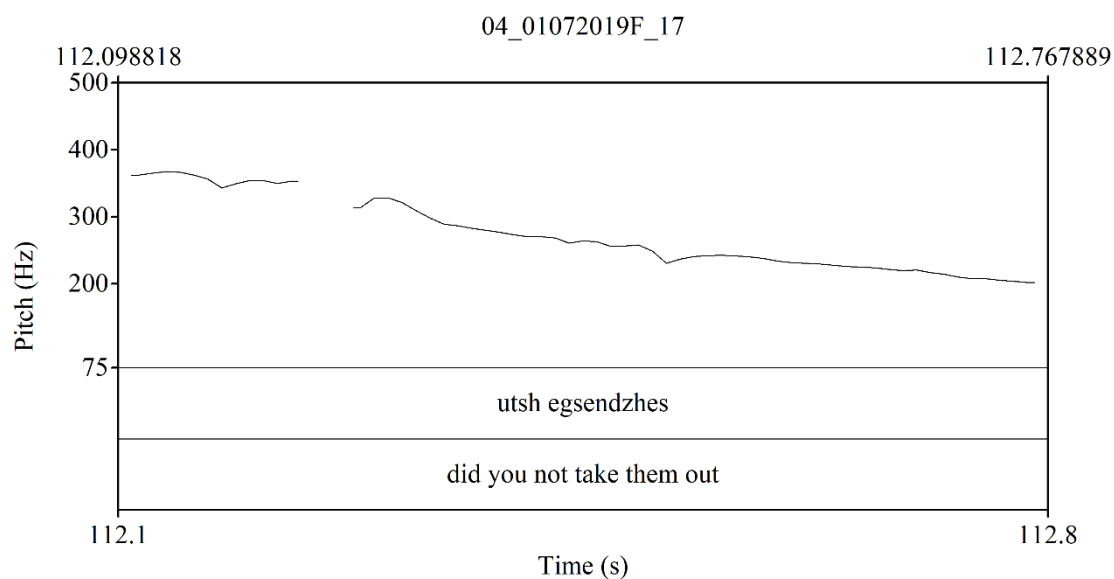
**Figure 8:** “Climactic” intonation contour

Focus, also contrastive focus, and intensification are indicated by a short rise of the highlighted word (see the word *manaxó(s)* ‘alone’ in Figure 9).



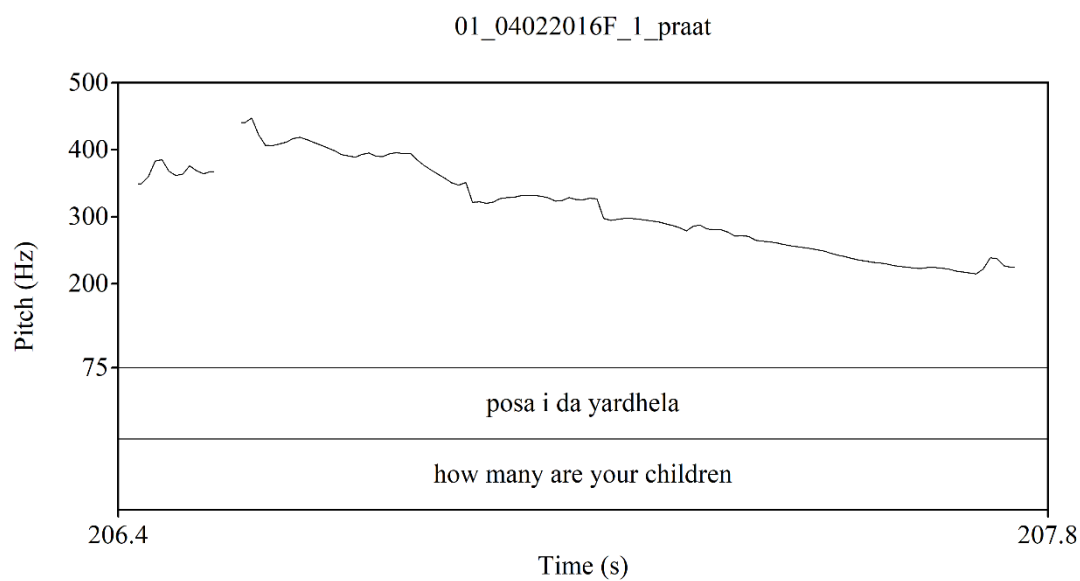
**Figure 9:** Intonation contour of focus

Interrogative clauses are usually marked by intonation although the Turkish interrogative particle *mI* is often borrowed in polarity questions (Section 5.2.3.3.1). The default contour for interrogative clauses is a rise before a decline. For the contour of a polarity question only marked by intonation see Figure 10.



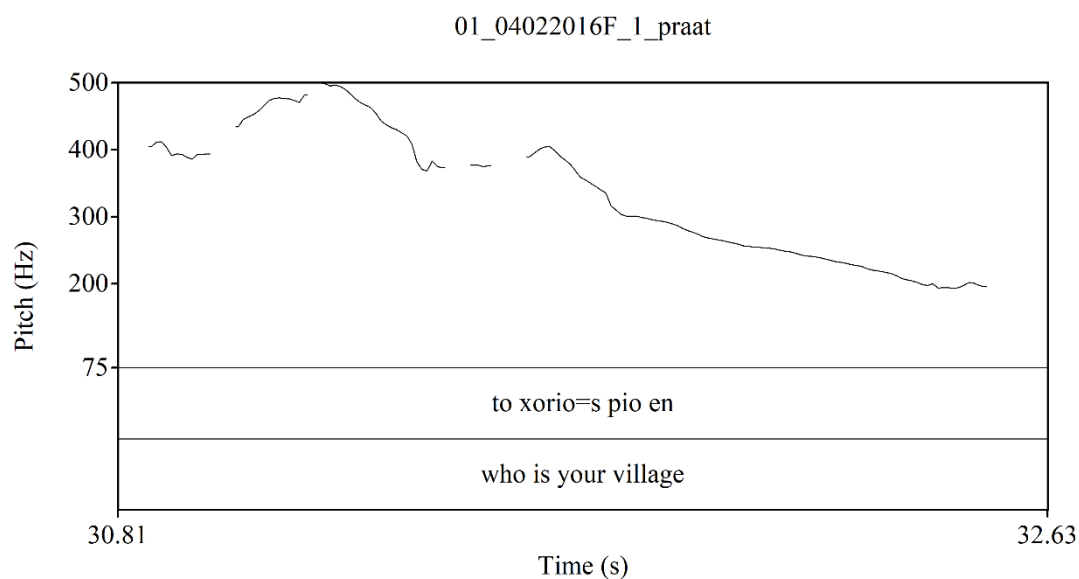
**Figure 10:** Intonation contour of a polarity question

For the intonation contour in *wh*-interrogatives see Figure 11.



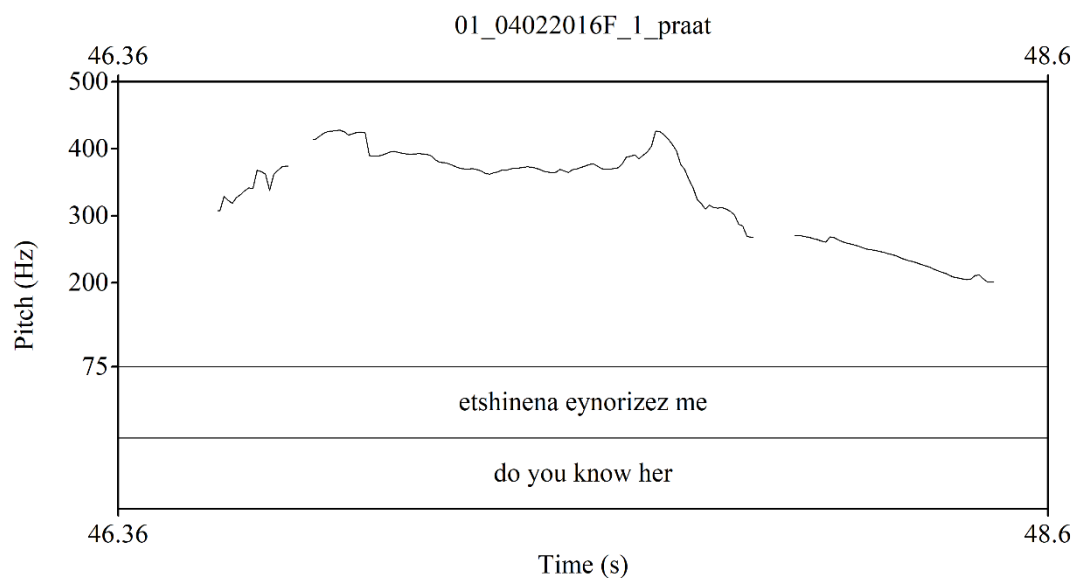
**Figure 11:** Intonation contour of *wh*-questions

If the object of the question is additionally focused, it is marked by the highest pitch (Figure 12 below).



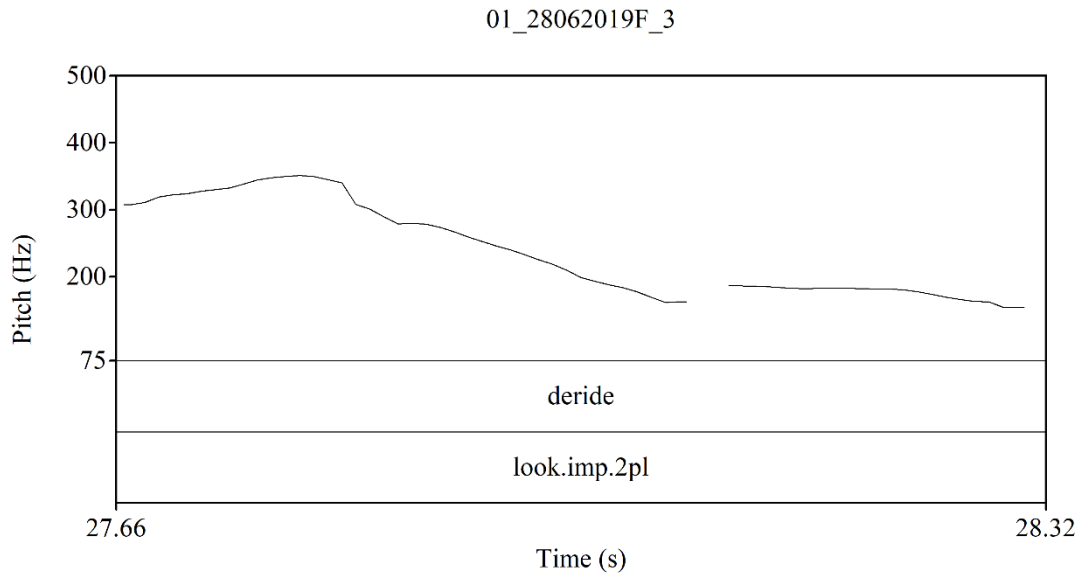
**Figure 12:** Intonation contour of a topicalized *wh*-question

A polarity question using the Turkish interrogative particle *mI* is shown in Figure 13. The rise of pitch occurs at the final syllables of the verb before the contour drops at the question particle.



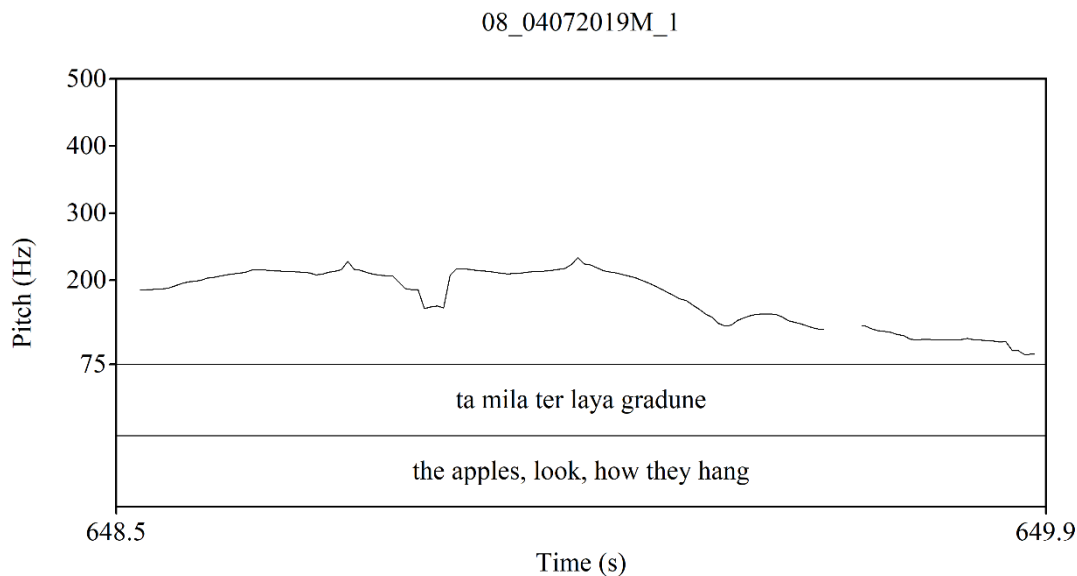
**Figure 13:** Intonation contour of a polarity question with Turkish interrogative particle

Imperative sentences are also marked by a rising and consequently falling contour (Figure 14).



**Figure 14:** Intonation contour of an imperative clause

For a complex sentence involving focus, an imperative, and an interrogative pronoun see Figure 15. The focused noun phrase of the example is highlighted by the first rise in pitch. The second rise is the imperative, followed by the third rise for the interrogative pronoun before the contour drops with the verb.



**Figure 15:** Complex intonation contour including focus, *wh*-element and imperative

Finally, forms of address like *ebáts* ‘Hey, girl!’ (03\_07072019F\_1; 01) and names follow generally the falling contour although with names also the rising contour is used to indicate that further information is following.

## 2.5 Orthography

Since Romeyka is mainly a spoken variety, no standardized orthography exists. Christian Pontic Greek has a literary tradition and is written in the Greek alphabet, which is not familiar to most Romeyka speakers, though. When Romeyka occurs in written form, for example on regional

social media groups, the Turkish alphabet is used. The Turkish graphemes are mapped on Romeyka phonemes (including marginal ones) as shown in Table 7, although there is individual variation:

**Table 7:** Some phoneme-grapheme correspondences in written Romeyka (Turkish script)

/p/	→	b <sup>33</sup> , p
/t/	→	d, t
/k/	→	k, g
/θ/	→	th
/ð/	→	d, (z)
/ɣ/	→	ğ
/χ/	→	h, (x)
[j]	→	y
[ʃ]	→	ş
[dʒ]	→	c
[tʃ]	→	ç

The Turkish script is not ideally suited to depict Romeyka phonemes. As can be seen in Table 7, there is for instance no graphemic correspondence for the phoneme /ð/. Likewise, /χ/ can sometimes not be distinguished from [h] (for the desirability of a one-to-one grapheme-phoneme correspondence involving even some non-phonemic phones cf. Lüpke 2011: 329). Still, phoneme-grapheme mapping in the Turkish script can reveal interesting insights into how the pronunciation of words and sentences is perceived but the result may be blurred when Turkish orthographic rules are applied to Romeyka, e.g., *sipiti* for (o)*spíti* ‘house’ (due to unacceptance of complex syllable onsets in Turkish). Vowel clusters are usually separated by a glide, e.g., *horiyo* ‘village’, *piyos* ‘who’. The treatment of word boundaries reveals that particles (aspectual, negation), prepositions, short interrogatives, clitic object pronouns, short verbs like the copula, and sometimes definite articles in nominative are perceived as part of the word, e.g., *kala=ise* ‘you are well’, *med=emena* ‘with me’, *na=pamen* ‘we will go’, *u=lebune=me* ‘they don’t see me’, *so=horiyo* ‘to the village’, *ekseris=mi* ‘do you know’, *b=eşi=bas* ‘where are you going’, *as=troğume* ‘let us eat’, *şoniz=ce=kahede* ‘it is snowing’, *eğö=pa esena* ‘me and you’, *iba=a* ‘I told her’. Interestingly, this mirrors the compound spelling of function words in regional informal Turkish in social media, e.g., *sen=de sağ=ol* ‘Thank you, too!’.

For the reason of familiarity of all Romeyka speakers with the Latin script (with some diacritics particular for Turkish), I apply in this thesis a Latin-based orthography with some diacritics to correctly depict Romeyka phonemes, which are partly in accordance with the IPA character, e.g., ð, θ, ɣ, æ and partly diacritics, e.g., ž, š, x, ö, ü (for details of the phonetic correspondences with the transcription system used in this thesis, see begin Chapter 2). Although the Greek script would be well-suited to display all phoneme qualities appropriately, this option is neglected here as the Greek script is not legible for most speakers.

<sup>33</sup> According to Mackridge (1987: 132, Fn. 25), Romeyka speakers of Uzungöl render Romeyka stops into voiced stops in Latin script. /θ/ is rendered as /s/.



### 3 Word classes

#### 3.1 Classes of content words

This section deals with the basic open word classes of nouns (Section 3.1.1) and verbs (Section 3.1.5) as well as open class modifiers such as adjectives (Section 3.1.2), numerals (Section 3.1.3) and adverbs (3.1.4) that modify nouns, verbs, adjectives, other adverbs, and clauses. In this chapter, the basic properties of the word class concerned are summarized with cross-referencing to the morphosyntactical chapters where their function is described in detail. In addition, this chapter contains some notes on the frequency of part-of-speech types.

##### 3.1.1 Nouns

###### 3.1.1.1 Lexical nouns

Nouns are lexical items that undergo as heads of NPs<sup>34</sup> some sort of grammatical marking (determination, inflection) to function as arguments of verbs (transitive and intransitive) or prepositions. They can be also predicates in predicative copula clauses (Section 5.2.1.5.1). Furthermore, nominalization is an increasingly productive complementation strategy, which can be likely traced back to contact with Turkish (Section 4.1.1.1; Section 5.3.2, especially Section 5.3.2.2.2 on deverbal nouns). For nominal derivation and compounding in general see Sections 4.1.1.1 and 4.1.2, respectively.

Nominal morphology is described in detail in Section 4.2. The main inflectional features that are associated with nouns are number (Section 4.2.1.2), gender (Section 4.2.1.1), and case (Section 4.2.1.3). Furthermore, nouns can undergo diminution (Section 4.2.4). The general structure of a noun word is ROOT(-DIM)-NUM/CASE. Number-distinction in Romeyka is two-fold into singular and plural. The case system distinguishes nominative, genitive and accusative/dative. Accusative and dative have merged into a single category equalling oblique case, but for traditional reasons, are glossed here as ACC. Romeyka has an inherited three-fold gender system: masculine, feminine and neuter. The patterning of the inflectional features of nouns is in general predictable according to nominal declension classes based on gender, stem endings, and word stress. However, due to diachronic and ongoing changes mainly in the gender system, but also reduction in the case system (like syncretism in plural case endings), declension classes are not always helpful to determine the forms of nominal declension (see also Drettas 1997: 129), so, for example, feminine nouns may receive neuter plural endings. Definiteness (Section 4.2.3) plays a role in the NP as well, although it is not grammatically marked on the noun and the patterning with the articles is not systematic. Division between mass and count nouns is not obviously relevant in Romeyka (Section 4.2.3), though this requires more research. Definite nouns in Romeyka are in general always preceded by the definite article which is congruent in number, gender and case (see Section 3.2.1 on articles). Most nominal modifiers precede their head noun, while anaphoric ‘weak’ possessive and object pronouns are enclitic on the noun (Sections 3.2.2.1, 3.2.2.3, also 4.2.2); for the structure of the nominal phrase see Section 5.1.

The Romeyka corpus shows 1,326 tokens of lexical nouns as opposed to about double a quantity of verbs (2,811 tokens; see also Appendix C).<sup>35</sup> Out of all nouns, 158 are proper names;

<sup>34</sup> There is no theoretical claim made here in the NP/DP debate; for the sake of necessity of a decision, nouns are considered here to be the head of an NP.

<sup>35</sup> It is important to note that the lexeme numbers listed throughout this thesis are token numbers taken from the Romeyka corpus. For reasons of the organisation of the corpus, it is at this point not easily possible to extract the number of types, although this would be undoubtedly a very important information.

in addition, there are 36 nominalizations of verbal stems. The percentage of loan nouns (for a definition of this term, see below) varies according to gender but is overall between 10–30% of the nouns (approx. 19% of feminine nouns are loans, 11% of masculine, 35% of neuter nouns; for around 250 nouns in the corpus, gender cannot be determined from context, these nouns are therefore not specified for gender, in this group 65% are loans). Overall, neuter gender is by far most frequent (643 tokens), followed by feminine (273 tokens), and masculine gender (170 tokens), suggesting that at a certain point some change has taken place in the gender system in favour of neuter gender (Section 4.2.1.1). The number of singular noun forms in the Romeyka corpus is slightly higher (826 tokens) than that of plural noun forms (500 tokens). For an overview of all part-of-speech token numbers see Table C.1 in Appendix C.

The use of the term “loan word” is in Romeyka highly disputable; it is very difficult to determine (a) when a word has entered the Romeyka lexicon, and (b) whether it forms “permanently” part of the lexicon or is used for pragmatic reasons in the communicative context, including code-switching, or as an effect of language shift whereby the Turkish vocabulary is more readily available to speakers (and in nearly any communicative setting understood by the interlocutor, anyway). In the present thesis, the term “loan word” is used in the same way as in Drettas (1997) who refers to “noms pontiques” and “emprunts”. Loan words are identified here simply based on the criterion whether their etymology is Romeyka/Greek or other, most often Turkish. With this very rough decision, it is not distinguished between loan words that have formed part of the Romeyka lexicon for centuries and those inserted more spontaneously in a particular discourse. This strict definition is applied here to shed some light on the amount of contact with (primarily) Turkish from a synchronic point of view but it is clearly considered to be a “working definition” that needs to be refined more carefully to the reality of the Romeyka lexicon under further scrutiny.

### 3.1.1.2 Proper nouns

Proper nouns<sup>36</sup> discussed in this section include human proper names/anthroponyms and toponyms, but also some ergonyms, i.e., names of artefacts including cultural concepts like child games. In general, the inflectional features of proper nouns do not diverge from that of common nouns, their position within the NP is also alike.

The (definite) article is obligatory for proper names (1), unless it is omitted for phonological reasons [<sub>V</sub>] (2; cf. Section 2.3.5.3). In some circumstances, the article is left out for unclear reasons (3, 4) - probably furthered by contact with Turkish that lacks articles altogether - but occurs in others (5). Proper nouns can also occur with the indefinite article when used as an appellative, i.e., when referring to one part of a group (6). When used as vocatives, names do not have an article, e.g., *Aiše!* (07\_04072019F\_8; 02), *ebats arzu* ‘Hey girl, Arzu!’ (03\_07072019F\_1; 01).

- (1) *o vejselis edžima dune*  
‘Veysel was asleep.’ (04\_01072019F\_5; 41)
- (2) *me don dursun tše med i glagšin tše med ø emine*  
‘with Tursun and with Glagshin and with Emine’ (04\_01072019F\_17; 17–19)
- (3) *si yoryora ido zias en dže ineka=θ šexrije*  
‘At Gorgoras there is this Ziya and his wife is Shehriye.’ (01\_04022016F\_1; 018–019)

<sup>36</sup> Proper nouns are considered here the lexical category of (proper) names following Schlücker & Ackermann (2017).

- (4) *i gülsüme - gülsüme erθe son barxari*  
 ‘Gülsüme – Gülsüme went to the pasture.’ (04\_01072019F\_17; 32)
- (5) *i aişe i tejze=m bal utş erte*  
 ‘My aunt Ayşe did not come.’ (04\_01072019F\_17; 65)
- (6) *ena şerije ksero*  
 ‘I know a Şehriye.’ (01\_04022016F\_1; 053)

Proper nouns inflect for gender and case based on the nominal declension classes (Section 4.2.1) largely determined by semantic gender, i.e., proper nouns referring to male humans are masculine, proper nouns referring to female humans are feminine, larger domestic animals such as cows may also follow this system; inanimate entities such as place names can have all three genders, further research is required here on the mechanism of gender assignment in toponyms. Anthroponyms and ergonyms are usually Turkish and are morpho-phonologically integrated into Romeyka declension classes. Case-sensitive morphological integration of the proper noun applies to masculine nominatives in *-is/- (a)s* (i.e., 1<sup>st</sup> declension class in *-a*), e.g., *Ahmet-is*, *Vahid-is*, *Veysel-is*, *Mustafa-s*, *Zija-s*, including Turkish honorific forms of address like *Alaj Bej-is* ‘mister Alay’ (but cf. *Dursun-ø*). Regarding masculine accusatives, however, case inflection is only visible on the definite article (7, 8). The same applies to accusative case governed by prepositions (9a; note, however, that the accusative ending does occur in the data of a heritage speaker in 9b). It is possible that definiteness plays a role in the distribution of nominative/accusative case endings on proper names (see paragraph (i) in Section 4.2.3). Apart from the masculine nominative singular, feminine genitive singular seems to be the only case that is morphologically marked on the (proper) noun (10, cf. the masculine genitive singular in 11).

- (7) *bolisa do veysel s=ðulia*  
 ‘I sent Veysel off to work’ (04\_01072019F\_5; 55)
- (8) *ton ali so ðormo utş iðame*  
 ‘We did not see Ali on the way.’ (C1)
- (9) a. *i aişe me don ali so bazar eşi ba* (C1)  
 b. *aişe me don alin so pazarin bai* (H2)  
 ‘Ayşe goes to the market with Ali.’
- (10) *omon di hadidžes ti mairia*  
 ‘like Hatice’s food’ (03\_30062019F\_11; 115–116)
- (11) *du veysel i tejze*  
 ‘Veysel’s aunt’ (04\_01072019F\_5; 49)

However, albeit not frequent with accusative anthroponyms (but cf. Neocleous 2020: 37), accusative case marking occurs on some feminine, masculine, and neuter toponyms governed by the local prepositions *aso* ‘from’ and *s-* ‘to’ (12–14, but cf. *sin trabezunda* (08\_04072019M\_1; 096), *sin bursa* (08\_04072019M\_2; 037), *so istanbóli* (08\_04072019M\_1; 246), *so izmíri* (08\_04072019M\_3; 111)).<sup>37</sup> It is not clear whether animacy is a factor here. Dependent upon the word class, these toponyms take word-final *-n* in the accusative, which

<sup>37</sup> Note that the *-i* ending in integrated Turkish neuter toponyms is used for the purpose of phonological integration and likely not for case marking.

occurred originally in the accusative singular of all genders of certain noun classes (Mackridge 1987: 124). However, there is significant intra-speaker variation (15; also, regarding gender assignment to integrated Turkish toponyms, see 16) which suggests that there might be also other variables at work.

- (12) *erθame sin almanjan*  
 came.1PL to.the.F Germany.F  
 ‘We came to Germany.’ (02\_21042018M\_2; 42)
- (13) *son ofin ebiyame*  
 to.the.M Of.M went.1PL  
 ‘We went to Of.’ (02\_29062019F\_2; 03)
- (14) *aso bulgaristanin din türkijan ulon eporpatesame*  
 from.the.N Bulgaria.N the.F Turkey.F all walked.1PL  
 ‘From Bulgaria we travelled whole Turkey.’ (02\_21042018M\_2; 35–36)
- (15) *si yoryora erθe ama ebebuka asi yoryoran ebebae udž en*  
 ‘She came to? Gorgoras but from the lower Gorgoras, it is not the upper.’  
 (01\_04022016F\_1; 051–052)
- (16) a. *pame si sultan murati* (04\_01072019F\_1; 033)  
 b. *so sultan murati* (04\_01072019F\_2; 110)  
 ‘(We go) to Sultan Murat.’

Finally, the names of child games are predominantly Turkish with mostly some phonological integration (17, 18 but 19), albeit no articles are assigned.

- (17) *jakartope epezame*  
 ‘We played yakar top.’ (04\_01072019F\_2; 032)
- (18) *lebe eliyam=a*  
 ‘We called it lep (=çizik taşı).’ (04\_01072019F\_2; 018)
- (19) *epesame birdurbir*  
 ‘We played birdirbir.’ (04\_01072019F\_2; 011)

### 3.1.2 Adjectives

Adjectives are defined here as a word class with descriptive semantic meaning, including words like ‘big’ or ‘heavy’ (Dryer 2013a). The following semantic properties are at least covered by Romeyka inherited adjectives: age (e.g., ‘young/old’), dimensions (e.g., ‘big/small’), values (e.g., ‘good/bad’), some colours (i.e., ‘black’, ‘white’, ‘red’), physical characteristics (e.g., ‘heavy’), human propensities (e.g., ‘happy’) (Payne 1997: 63). In terms of their grammatical behaviour, Drettas (1997: 155) designates “le nom adjective” as “véritables noms” as their properties resemble that of nouns. Adjectives can (a) directly modify a noun (attributive adjectives, Section 3.1.2.1) or (b) function as predicate in a clause (predicative adjectives, Section 3.1.2.2). Non-descriptive modifiers like demonstratives, quantifiers or numerals are treated in separate sections below. Adjectival derivation is discussed in Section 4.1.1.2. The formation of comparatives with the ancient comparative suffix *-tero(n)/(-tera.PL)* is presented in Section 5.2.1.6.

In the following, a brief overview is given on the frequency and distribution of (inherited and loan) adjectives in Romeyka. Out of 10,247 words in the Romeyka corpus, 55 are attributive adjectives and 175 are predicative adjectives (see frequencies per part of speech in Table C.1 in Appendix C). These boil down to a limited number of lexemes in the word class of adjectives, which is, thus, relatively limited in the present corpus. On the other hand, the class is very ‘open’ in the sense that it accepts loanwords which are mostly phonologically adapted and morphologically integrated, e.g., *tari* ‘narrow’ < Tr. *dar* ‘narrow’, *tehlikeli*.PL ‘dangerous’ < Tr. *tehlikeli* ‘dangerous’. Out of 55 occurrences of attributive adjectives in the corpus, 23 are Turkish loanwords and out of 175 occurrences of predicative adjectives 79 are Turkish loans. For the most frequent lexemes in the class of adjectives see Table 8 below.

**Table 8:** Absolute word form frequencies of native and loan adjectives (both attributive and predicative)

Native adjectives (examples)	Token frequency	Loan adjectives (examples)	Token frequency
<i>emorfo/omorfo</i> ‘nice’	38	<i>kalabaluxi</i> ‘crowded’ < Tr. <i>kalabalık</i>	18
<i>kalo</i> ‘good’	26	<i>rahat</i> ‘comfortable’ < Tr. <i>rahat</i>	7
<i>trano</i> ‘big’	10	<i>temizi</i> ‘clean’ < Tr. <i>temiz</i>	6
<i>mugudigo</i> ‘tiny’	4	<i>doyali</i> ‘organic’ < Tr. <i>doğal</i>	6
Total occurrences of native adjectives in the corpus	128	Total occurrences of loan adjectives in the corpus	102
Adjective occurrences in total		230	

### 3.1.2.1 Attributive adjectives

Attributive adjectives show grammatical properties similar to that of nouns with which they agree in a number of features. Adjectives agree with the head noun in number, (with some systematic variation) in gender, and in principle also in case (Neocleous 2020: 34; Drettas 1997: 167). It needs to be noted that insufficient data is available in the present corpus to establish whether declension classes still play a role in adjectival agreement. A simplified paradigm of adjectival declension is presented in Table 9.

**Table 9:** Adjectival declension in Romeyka

	SG	examples	PL	examples
M	-o(n) <sup>38</sup>	<i>emorf-o dobos</i> ‘a nice place’	-a, (-i) <sup>39</sup>	<i>mikr-a argures</i> ‘small bears’
F	-esa, (-i)	<i>i kal-esa i mana</i> <sup>40</sup> ‘the good mother’	-a	<i>emorf-a ðulias</i> ‘good works’
N	-o(n)	<i>to tran-on to peðin</i> ‘the big boy’ <i>ena tran-o xorio</i> ‘a large village’ <i>omorfo diri</i> ‘nice cheese’	-a	<i>emorf-a provada</i> ‘beautiful sheep’ <i>ta neada ta peðia</i> ‘the young girls’

This declension paradigm roughly also extends to derived deverbal adjectives in *-menos*, e.g., *pinasmen-os.M*, *pinazmen-isa.F* ‘thirsty’ (cf. Drettas 1997: 162; Mackridge 1987: 125). However, since these derived adjectives occur mainly in predicative attribution, they show forms of grammatical (and semantic) gender agreement that only remain in predicative attribution (see Section 3.1.2.2. below).

Tursun (2019: 29) notes variation in the adjectival feminine singular ending which is in the Solaklı river basin *-esa* but may be also in *-i* like in SMG, e.g., *mikr-esa* vs. *mikr-i* ‘small.F’. Note that apart from this, *-i* occurs in other cases of adjectival declension as well: (a) loan adjectives are integrated by a word-final *-i*, e.g., *o havas temiz-i* ‘the air is clean’ (08\_04072019M\_1; 239), *ena zengin-i auro* ‘a rich man’; (b) the *i*-ending occurs as a peculiarity of a small number of inherited Greek adjectives which have *-i* instead of *-o-* in comparatives with *-tero(n)*, e.g., *to=na to kat trani-tero en*, *t=alo mikro-tero en*. ‘One storey is larger, the other is smaller.’ (A1) (see also exs. 194, 197 in Section 5.2.1.6); (c) *-i* occurs as the pl.masc.nom. form of some adjectives (see Fn. 39), e.g., in the plural form of the adjective *mikro* ‘small’ used as adverbial, i.e., *mikri* ‘when we were young’ (08\_04072019M\_1; 012).

From the paradigm in Table 9, it is furthermore striking that the neuter forms of plural declension seem to have taken over masculine and feminine plural inflection. According to Drettas (1997: 167), especially feminine adjectives<sup>41</sup> tend to take the neuter plural, e.g., *ta tran-a ta ðualias* ‘the big matters’, although he also notes a feminine plural ending in *-esas* for PG, *ta kal-esas ta yariðas* ‘the good women’ (Drettas 1997: 168). Importantly, the feminine nouns fully shift to the neuter declension in the plural resulting also in the neuter plural article, e.g., *ta gentše ta batsiðes* ‘the young girls’ (H1), *t=ixtijare ta inetšes* ‘the old women’ (H1). According to Mackridge (1987: 128) also inanimate masculine nouns (and sometimes even animate and human masc. nouns) take the neuter ending in the plural, e.g., *ta pala ta kerús* ‘in olden times’. This seems to have largely generalized for the plural of all masculine and feminine nouns in the present Romeyka corpus.

Furthermore, the neuter (singular) declension of adjectives has spread to all [-hum] feminine nouns and partly also to feminine [+hum] singular nouns (Drettas 1997: 167), e.g., *to*

<sup>38</sup> The use of the accusative inflectional ending *-on* also for nominative heads (and modifiers) instead of the nominative masculine singular ending *-os* is characterized as a typical phenomenon of PG (Drettas 1997: 160–161). In the present Romeyka corpus, most masculine/neuter adjectives end in *-o*, which should equal the neuter ending, but see more clearly: *o tran-on o andras* ‘the strong man’ (Neocleous 2020: 35, ex. 21). Note furthermore, remnants of the masculine *-os* declension in predicative adjectives denoting humans (Section 3.1.2.2 below).

<sup>39</sup> Mackridge (1987: 128) notes *-i* as plural ending of adjectives modifying feminine [+hum] and all masculine nouns in the plural. However, the form occurs only once in the corpus in a rare mixed plural form, i.e., *ta mikri da aðelfa* ‘the small siblings’ (01\_06042017F\_1).

<sup>40</sup> This example is constructed based on Mackridge (1987: 125) and Drettas (1997: 160–161). Importantly, Özkan (2013: 146) notes for RSür that this feminine declension is not correct (anymore) and that the neuter declension needs to be used instead, e.g., *to kalo i mana* ‘the good mother’. Thus, the feminine inflection presented in Table 9 seems to only remain valid in predicative adjectives.

<sup>41</sup> According to Mackridge (1987: 128), all non-human feminine nouns and sometimes even some human feminine nouns.

*tranon i ðulia* ‘the big matter’, *to mavron i kosara* ‘the black chicken’, *to mikro i batsi* ‘the small girl’ (B1). Drettas (1997: 168) notes further that when for some [+hum] feminine nouns both the feminine and the neuter declension exist, the meaning differs with regard to specificity: if the feminine declension is used, the noun is specific, if the neuter declension is used, the noun is generic. In the present Romeyka data, however, the feminine singular form is neuter regardless of definiteness/specificity, e.g., *to emorfo i patsi* ‘the beautiful girl’ (H1), *kalo mualimina* ‘a good teacher.F’ (01\_04022016F\_1; 077), *ena aletero ylossa* ‘another language’ (C1). Feminine singular adjectives in *-esa* occur in the present corpus only in predicative attribution, e.g., *kalesa* ‘she is well’ (04\_01072019F\_5; 17), which is also confirmed by Özkan (2013: 145) for RSür. Note that in RSür also all [+hum] masculine nouns take neuter adjectives, e.g., *do galó o árθobos* ‘the good man’ (Özkan 2013: 146). According to Neocleous (2020: 35), all attributive adjectives in Romeyka are neuter except for masculine human NPs, e.g., *o dranon o andras* ‘the strong man[/husband]’.

These deep going changes in gender agreement that lead to a spread of neuter declension (see Karatsareas 2011, 2014) indicate that grammatical gender agreement has first shifted to semantic gender agreement (Michelioudakis & Sitaridou 2013: 366, Fn. 3), like in the case of inanimate feminine singular nouns and non-human feminine and inanimate masculine nouns in the plural, but finally, gone much further in the present Romeyka data (and in RSür, according to Özkan 2013) so that now only remnants of gender agreement are visible, for example in predicative adjectives (see below). The mix of gender declensions seems to reveal that determiner + adjective congruence and determiner + noun congruence are still intact, except for that in plural forms, all plurals follow the neuter declension, irrespective of the gender of the noun in singular. Janse (p.c., see also Janse 2004) suggests for this change the following path: The change might have departed from feminine singular nouns via syncretism of the feminine plural (e.g., *i kali i jinetšes* ‘the good women’) and the masculine plural (e.g., *i kali i arθob* ‘the good men’) to the neuter plural declension (e.g., *ta kala ta jinetšes* ‘the good women’) and then spread to the singular declension of feminines as neuters (e.g., *to kalo i jineka* ‘the good woman’). Masculine forms are targeted last and are thus most stable (for an overview of these various shifts in AMG, see Karatsareas 2014). Apart from this analysis, however, one might wonder whether the spread of neuter declension might be simply analysed as syncretism leading to invariant forms, hence decreasing (superficial) complexity in the paradigm. Especially, as neuter declension shows the least complexity in its paradigm. At least, gender distinctions are cross-linguistically most likely to be lost in the plural rather than in the singular, and gender-distinction only in the singular is a common pattern in some Indo-European languages.

Finally, adjectives have shown case agreement in AG, which is still reported for PG in Drettas (1997: 165), e.g., *iða ton tranon turkon* ‘I have seen the grand/important Turk’. But since case is not discernible in the neuter declension of adjectives, information on case is often not visible (20). Therefore, case agreement is – if at all, given the change of many adjectives, especially in the plural, to neuter gender – only visible on the article of the (singular) head noun, e.g., *to tranon.NOM/ACC ti ðulian.ACC* ‘the big matter’ (Drettas 1997: 167), see also (21), (22).

(20) *omorfo duvari eftes*  
‘You build a nice wall.’ (08\_04072019M\_3; 072)

(21) *hatšino d=omorfo don dobo etšiega krati*  
‘It has<sup>?</sup> a nice place there.’ (08\_04072019M\_3; 162)

(22) *to kendi ti ðulias to pos efðed edži*  
‘How do you do your own work there?’ (04\_01072019F\_2; 212)

Regarding the syntax of attributive adjectives, they are pre-nominal modifiers, e.g., *ena zengini auro* ‘a rich man’ (AdjN, see also Drettas 1997: 165), whose position in the NP is in immediate prenominal position (except for the definite article of the definite head noun). An infinite number of attributive adjectives can co-occur, e.g., *íxa tráno varí yomári si ráša=m* [...] ‘I have a big, heavy load on my back’ (Tursun 2019: 496). Adjectives (both attributive and predicative) can be modified by adverbs of manner like *bola* ‘very’. Unlike in PG, in Romeyka adjectives cannot take possessive suffixes (cf. Drettas 1997: 166). Their position in the NP is the following: POSS – (ART) – NUM – ADV – (ART) – ADJ – (ART) – N (see also Section 5.1). Crucially, in definite NPs the phenomenon of determiner spreading (Karatsareas & Lekakou 2016, i.a.) applies, i.e., the definite article of the head noun occurs also obligatorily before attributive adjectives and numerals (Mackridge 1987: 128, Fn. 18). If a noun is modified by two adjectives, the definite article occurs three times, e.g., *to.N emorfo.N to.N gentšete i.F batsi.F* ‘the beautiful young girl’. In Romeyka, these definite articles are at least congruent in number (and in principle also in case) but, as pointed out above, feminine and masculine head nouns may be modified by a neuter adjective, which takes the neuter article, e.g., *to emorfo i patsi* ‘the beautiful girl’ (H1).

In the present questionnaire data, there is considerable variation with regard to the occurrence of the determiners in both singular and plural. Incidentally, definite forms occur with only the article of the head noun (23a) or even without articles (23b), especially in the plural (23c). It seems definiteness has been not correctly identified in the translation task from Turkish or that there is interference from Turkish, i.e., the lack of determiners in Turkish and the deviant concept of expressing definiteness/specificity by nominal case marking.

In indefinite singular contexts, the indefinite article mainly occurs when the head noun is specific (24) and is otherwise left out (25, but cf. 26). Indefinite plural NPs often lack a determiner (27), although incidentally the definite article appears before the head (28, also 29 in a singular NP).

- (23) a. *drana da beðia*  
‘the old girls’  
b. *mikro batsi so kuli*  
‘The little girl is at school.’ (C1)  
c. *tema.PL tesera mavra.PL kosares.PL*  
‘my four black chicken’ (H2)
- (24) *ena omorfo ospid-obo en etšis*  
‘There is a beautiful little house.’ (08\_04072019M\_3; 146)
- (25) *geniše ðormo utš en*  
‘There is no broad way.’ (08\_04072019M\_1; 077)
- (26) *emorfon bal guvetin yarðelin emune*  
‘I was a handsome and strong child.’ (02\_21042018M\_2; 328)
- (27) *trana ba obsarae eplezame bola ba*  
‘We caught as well big as many fish.’ (08\_04072019M\_1; 159)
- (28) *layamis da insane*  
‘happy people’ (04\_01072019F\_1; 187)
- (29) *temis to gemlegin eforenes*  
‘You wore a clean blouse.’ (03\_07072019F\_1; 28)



In forms of address, no articles appear (30).

- (30) *e mugudigo batsi*  
 ‘Hey, little girl!’ [part of a saying] (04\_01072019F\_1; 082)

### 3.1.2.2 Predicative adjectives

Predicative adjectives share largely the grammatical properties of attributive adjectives, i.e., number and gender agreement (but no case agreement). Importantly, remnants of grammatical gender agreement are still prevailing in the singular of predicative adjectives (but not in the plural!) where attributive adjectives have shifted to semantic agreement or even lost gender distinction. This applies notably to feminine singular adjectives (of some declension classes, see Drettas 1997: 160–162) in *-esa/-isa*, e.g., *kalesa* ‘she is well’ (04\_01072019F\_5; 17), see also (31, 32), but also to some masculine singular forms in *-os* denoting humans, e.g., *layos e* ‘How is he?’ *kalos e* ‘He is well.’ (04\_01072019F\_5; 26/27). This also includes the forms of derived adjectives in *-menos* (Section 4.1.1.2; Mackridge 1987: 125). Neocleous (2020: 61) highlights the use of the passive perfect participles in *-menos* as predicative adjectives and notes that they agree with their head in gender, number and case. According to Özkan (2013: 145), predicatively used adjectives exhibit grammatical and semantic agreement, if the noun denotes a human referent, e.g., *ime omorfos* ‘I.M am beautiful’, *ise omorfi* ‘you.F are beautiful’ (Tursun 2019: 27). Animate and inanimate masculine nouns, however, take the neuter declension (33, 34, respectively). In general, like with attributive adjectives, many adjectives (of all gender and even those denoting human and animate nouns) may have the neuter nominative singular in *-o(n)* (35).

- (31) *olon dranisa eyo mune*  
 ‘I was the oldest.’ (04\_01072019F\_2; 07)
- (32) *ixtijar layosa edune – kalesa kalesa*  
 ‘How is<sup>2</sup> your mother? - Well, well.’, lit. ‘the elderly (female) person’  
 (04\_01072019F\_5; 16–17)
- (33) *teso o škilo mavro en*  
 POSS.2SG the.M dog.M black.N/M COP.3SG  
 ‘Your dog is black.’ (C1)
- (34) *to jaš=im bola trano en*  
 ‘My age is advanced.’ (01\_04022016F\_1; 005)
- (35) *temo i patsi emorfo en*  
 ‘My daughter is beautiful.’ (A1)

Based on remnants of inherited declension classes, other adjectival declensions occur as well. According to Mackridge (1987: 125, also 128), some feminine animate nouns that had been modified by masculine adjectives in AG continue to do so, e.g., *temeteron i nife emorfos en* ‘our daughter-in-law is beautiful’. Furthermore, there are some (probably feminine) predicatively used adjectives in *-a* whose singular form coincides with the plural in *-a*, e.g., *makra* ‘far’ (36a/b). These might go back to an AG declension group of adjectives in *-os* that have *-a* as feminine form (Holton et al. 2019: 704, 715).

- (36) a. *makra ise*  
 ‘You.F are far away.’ (01\_14012015F\_1; 3)  
 b. *ula mona makra*  
 ‘Everything is a bit far.’ (07\_04072019F\_6; 31)

The plural ending of predicative adjectives consistently follows the neuter declension with the plural in *-a* (cf. Mackridge 1987: 126), irrespective of gender (37a/b).

- (37) a. *ta peḍia =m abora utš ine*  
 the.PL son.N.PL=POSS.1SG bad.PL NEG COP.3PL  
*ta peḍia =m ba kala ine*  
 the.PL son.N.PL=POSS.1SG TOP good.PL cop.3PL  
 ‘My sons are not bad, my sons are also good.’ (03\_30062019F\_7; 34)  
 b. *ta nifaḍe =m ba kala ine*  
 the.N.PL daughter-in-law.F.PL=POSS.1SG TOP good.F.PL COP.3PL  
 ‘My daughters-in-law are also good.’ (03\_30062019F\_7; 46)

The syntax of predicatively used adjectives is outlined in more detail in Section 5.2.1.5.2. Predicative adjectives follow the noun as they adhere to the general order of predicates and subjects in Romeyka. The order of predicative adjective and verb is variable, though. It depends in part upon the clause type (see Section 5.2.1.5) but also other variation (e.g., idiolectal, dialectal) occurs. In copula clauses, the predicative adjective often immediately follows the noun since copula clauses tend to be verb-final in the present corpus (38; but cf. Mackridge 1987: 128 for examples of both verb-second and clause-final copulas).<sup>42</sup> In 3<sup>rd</sup> person singular, the copula is often omitted (39). Inchoative predicates like ‘become’ appear more frequently before the predicative adjective, although there is variation as well (40). Both orders occur also with other predicative adjectives such as *manaxesa* ‘alone’ (41a/b), *tomara* ‘together’ or *endama(n)/andama* ‘together’, whereby information structure, namely focalization, can also play a role (see 42a/b).

- (38) *teseteron to zon monon aškemon edune*  
 ‘Your cow was a little black.’ (H1)
- (39) *etšinos xaremenos*  
 ‘He is happy.’ (02\_29062019F\_2; 24)
- (40) *ejendune xastas*  
 ‘It (=the child) became sick.’ (01\_06042017F\_4; 124)
- (41) a. *ondan sora ebomena manaxesa*  
 ‘Thereafter, I.F stayed alone.’ (01\_06042017F\_4; 141)  
 b. *don avle manaxesa erxumun*  
 ‘Midday, I went on my own.’ (04\_01072019F\_2; 068)
- (42) a. *makris is esi* (H1)  
 long.F COP.2SG you  
 b. *bola magri ise* (H2)  
 very long.F COP.2SG  
 ‘You are tall.’

<sup>42</sup> Cf. Neocleous (2020: 55) for the syntax of participles in *-menos*: “Note that the auxiliary verb always proceeds [sic!] the participle”.

Finally, two additional phenomena that occur with predicative adjectives should be mentioned here. Firstly, the suffix *-dena* is used in copula clauses to indicate origin (mostly of a person) based on a toponym, e.g., *patanu-dena* ‘from Bacan’, *γoryoran-dena* ‘from Gorgoras’. It is possible, that *-dena* creates an adjectival form of origin comparable to the Turkish suffix of origin *-li*, e.g., *Çaykara + li > Çaykaralı* ‘somebody from Çaykara’. Thereby, *-dena* might be possibly related to a certain feminine adjectival declension in *-ena* that Drettas (1997: 163–164) presents for the integration of Turkish nouns and adjectives into Greek (see 43, 44), for example based on the Turkish suffix of origin *-li*. However, the expression of places of origin with *-dena* seems not to be limited to female agents. As a different analysis, the suffix *-dena* could be a reanalysis of the Turkish ablative suffix *-DAn* fused with a Greek ending like *-ena*.

- (43) *ama i mana=θe mono merak-li-ena dune*  
 ‘But her mother was a bit anxious.’ (01\_06042017F\_4; 062–063)

- (44) *ap esas pola memnun-ena en*  
 ‘She is very content with you.’ (01\_04022016F\_1; 068)

Secondly, there appears to be a rare form of expressing intensity in adjectives with *(e)na* (45, 46). The analysis of this phenomenon is unclear but it seems not to be related to the Turkish intensification strategy with partial reduplication of the first syllable of adjectives, e.g., *mas-mavi* ‘very blue’, *ıp-ıslak* ‘soakingly wet’ which reportedly occurs in Cappadocian (Bağrıaçık & Janse 2016). However, Janse (p.c.) also notes in Cappadocian the use of the indefinite article *ena* to mark intensification together with *polá* ‘many’, probably as a calque of Tr. *birçok* ‘several’, e.g., Capp. *éna polá ascérja* ‘very many soldiers’ (example M. Janse, p.c.). As a different hypothesis, *(e)na* might be also related to the Turkish superlative marker *en* (M. Janse, p.c.).

- (45) *ena-morfa indane, na-morfa mirizune*  
 ‘They are very nice, they smell very nice.’ (03\_30062019F\_11; 108)

- (46) *omon d=eroise s=ormi ejendune na-tsurula*  
 ‘When she fell into the water, she got soakingly wet.’ (04\_01072019F\_13; 45)

### 3.1.3 Numerals

The lower cardinal numerals in Romeyka (used to) share some grammatical properties with adjectives, namely gender- and case-agreement. Drettas (1997: 172) reports for PG variability for the numerals until ‘four’ with gender-sensitive forms for masculine and neuter and case-agreement. Feminine forms are reported to be neuter or, if [+hum] masculine. Drettas (1997: 173) notes further that in indefinite masculine NPs, the neuter form can be used. It seems the use of the neuter form is in Romeyka extended to all contexts (irrespective of definiteness), so that the numerals are de facto invariant with regard to gender. This brings along that case-agreement could be only visible in the genitive, for which, however, no examples are attested. Thus, it is not clear whether the lower numerals, except for ‘one’, still show any case-agreement. The numeral *ena* ‘one’ (*ena.NOM.N.SG*, Özkan, n.d.), which functions also as indefinite article (Section 3.2.1.2), seems to show case-agreement (47). Although in comparison with (48), it is not clear, whether simply the neuter form is used which has the same form in nominative and accusative case (the masculine nominative form *enas* does not exist in the present corpus; for unclear variation between *ena* and *enan* ‘one’, see also Section 3.2.1.2).

- (47) *ena ðio tria yarðele exo enan batsi ðio peða enan inega exo*  
 ‘I have one two three children, a girl and two boys, I have one wife.’  
 (06\_03072019M\_2; 14-16)

- (48) *argo enan edun, erðane ala ðio erðane*  
 ‘The bear was one, others came, two came.’ (05\_03072019M\_4; 31–32)

Cardinal numerals can be treated as a noun and receive a determiner that is sensitive to number, although the numeral itself is invariant according to number (49).

- (49) *ta za to enan etšilizane veja ta ðio tšilizane*  
 ‘One of the cows tumbled, or two tumbled.’ (08\_04072019M\_1; 110–111)

As for the syntactic properties, numerals in Romeyka precede the noun. In the NP, numerals precede other modifiers like adjectives (Section 5.1). While indefinite NPs do not require a determiner, it is obligatory in definite NPs. However, determiner spreading seems not to be obligatory; if the definite article occurs before the numeral, no definite article is required before the head noun (exs. 50, 51; Drettas 1997: 171), although still often present (exs. 52, 53). Note that Karatsareas & Lekakou (2016: 195) state for Cappadocian, that the determiner of the head noun can be in certain environments phonologically null-realized; this seems not straightforward in the case of exs. 52, 53, i.e., [\_C], though.

- (50) *erðen me ta ðio patsið=atjes*  
 ‘She came with her two daughters.’ (07\_04072019F\_5; 04)
- (51) *esi ešis i ðio mayula, mi ges=ada don andras*  
 ‘You have two cheeks, don’t give them to your husband.’ (01\_28062019F\_4; 85–86)
- (52) *mo ta dört tane ta za*  
 ‘with four cows’ (04\_01072019F\_2; 087)
- (53) *ts=aišes ta tria t=aðelfae s=okuli bane*  
 ‘Ayşe’s three siblings go to school.’ (04\_01072019F\_11)

Note, however, that in heritage speakers under influence of Turkish, definite articles may be omitted altogether (54).

- (54) *tema tesera mavra kosares efepsan*  
 ‘My four black chicken escaped.’ (H2)

In Romeyka, native cardinal numerals exist up to value five (Mackridge 1987); a common trait of contact languages, which could be potentially linked to finger-counting. For all other cardinal numbers Turkish loans are used, often phonologically integrated, e.g., *juz* ‘hundred’ < Tr. *yüz*, *uš* ‘three’ < Tr. *üç*. As of value ten, the Turkish arithmetic base is copied along with the lexicon, e.g., *on* ‘ten’ + *bir* ‘one’ = *on bir* ‘eleven’. In ex. (55), the compound number is inserted as part of the Turkish adpositional phrase.<sup>43</sup> Heritage speakers might know the native numbers only up to value 4 and even competent speakers may slip out Turkish loans for the lower cardinal numbers (see Table 10 for the token frequencies of native numerals and loans).

<sup>43</sup> The function of the *-i* marker on the first number compound is unclear but it might be an attempt to integrate the number (for phonological integration of loans by *-i*, see Section 4.1.4).

- (55) *on biri on iki jaşında etuturevane=masine nestija*  
 ‘Eleven, twelve years old, they made us fast.’ (04\_01072019F\_2; 126–127)

**Table 10:** Cardinal numerals in the Romeyka corpus till value 10

Inherited	Token no.	Turkish	Token no.	Value
<i>ena</i>	3	<i>bir/pir</i>	20	‘one’
<i>đio</i>	17	<i>iki</i>	13	‘two’
<i>tria</i>	14	<i>uř/utř</i>	9	‘three’
<i>tesera</i>	6	<i>dört</i>	12	‘four’
<i>pende</i>	1	<i>beř</i>	17	‘five’
		<i>alti</i>	2	‘six’
		<i>jedi</i>	2	‘seven’
		<i>sekiz</i>	1	‘eight’
		<i>dokuz</i>	0	‘nine’
		<i>on</i>	7	‘ten’
Total no. of all numerals:			146	

Alongside Turkish numbers, often the Turkish numeral classifier *tane* ‘piece’ is borrowed (33 tokens in the corpus); like in Turkish, it is optional and also used for nouns of high countability like humans or animals (cf. Gil 2013a). However, unlike in Turkish, *tane* triggers in Romeyka the plural form of the head noun (only in highly attrited speech, a singular noun may follow, e.g., cf. (08\_04072019M\_1)). *Tane* only occurs together with Turkish numerals (56).

- (56) *dört dane yarđele exo, đio beđia đio batsiđæ – sekiz dane eyo=xo*  
 ‘I have four children, two boys and two girls. – [other speaker:] I have eight.’  
 (04\_01072019F\_1; 050–052)

Frequently, also the Turkish adverb *tek* ‘single, only’ is borrowed for emphasis and follows *ena* ‘one’ (57).

- (57) *ođande de udž ebidžes sade ena dek de n=eftas=a*  
 ‘Why did you not make another, only one, what will you do with it?’  
 (01\_04022016F\_1; 083)

Finally, all measure words are borrowed from Turkish: *kiři* ‘person(s)’, *kere* ‘times’, *sefer* ‘times’ as well as units, e.g., *kilo* ‘kilo’, *lira* ‘lira’ *saat/sahat* ‘hour’, *sene* ‘year’, *gün* ‘day’, *ai* ‘month’. Temporal adverbials as in (58) are often completely borrowed from Turkish.

- (58) *iki gün sora aso spid=nades eři ba*  
 ‘Two days later, she will be going from her house.’ (01\_06042017F\_4; 023)

On ordinal numbers in Romeyka no information is available so far; this domain requires further research. Compound Turkish loans including an ordinal number are borrowed as such, e.g., *ilk okuli* ‘primary school’.

### 3.1.4 Adverbs

Adverbs are a poorly defined remnant category of items that mainly modifies adjectives, verbs and other adverbs. While principally in SMG, “[a]dverbs never agree in number or gender with an NP, unlike adjectives” (Alexiadou 1997: 7), some adverbs in Romeyka do not show any agreement but many frequent adverbs agree with the subject of the clause in number. These adverbial expressions are derived from adjectives<sup>44</sup> or NPs of adjectives and a noun (e.g., temporal adverbs like *mijan/mies* ‘once’ (< sg.acc. *mí(j)an iméra* ‘one day’), *embro/embra* ‘before’, *ifetos/ifeti* ‘this year’ or many frequent adverb of degree *mono/mona* ‘little’.

Adverbs can be divided regarding their syntactic scope which leads to different syntactic behaviour: sentence adverbs have a wider scope than VP-adverbs which in turn show more syntactic variability than adverbs modifying other adverbs or adjectives. While most adverbs are modifying adjuncts, some can be considered complement-like, these are selected by a verb (i.e., “predicate adverbs”, Alexiadou 1997: 6). Generally, complement-like adverbs occur in pre-verbal position but can also occupy a postverbal position or appear clause-initially if focalized. Modifying adverbs precede the head they modify, which is often a property of OV languages (Neocleous 2020: 237). This applies certainly to those adverbs modifying adjectives and other adverbs, but also to those modifying VPs, although, depending on the scope of the adverb, more variation is possible. The syntactic position of an adverb may indicate a different meaning of the same adverb (59 vs. 60).

With 1,193 tokens in the Romeyka corpus, adverbs appear much more frequent than adjectives (230 tokens). Adverbs can be classified semantically into four subcategories, which are discussed in detail below.

- (59) *ti mana=s bola eyapesa*  
 ‘I loved your mother much.’ (03\_07072019F\_1; 10)

- (60) *s=opsarema bejname bola*  
 ‘We went fishing a lot.’ (08\_04072019M\_1; 139)

#### 3.1.4.1 Adverbs of degree

Adverbs of degree (or quantification, e.g., ‘a lot’, ‘very’, ‘little’, ‘a bit’) are usually invariable, except for those – very frequent ones – that are derived from adjectives. While the invariable adverbs cannot undergo comparison, those derived from adjectives did at least diachronically (Holton et al. 2019: 841). *Bola* ‘very, a lot’ (< MedGr *πολλά*.N.(NOM.)ACC.PL ‘many’) and *mono* ‘little, a bit’ (< AG *monon*.ACC.SG ‘by itself, alone, only’) are the most frequent adverbs of degree in Romeyka and (at least the latter is) derived from an adjective.<sup>45</sup> Furthermore, the adverb *fazla* ‘much (more)’ is borrowed from Turkish. Occasionally, even for adverbs where Romeyka lexemes exist, Turkish forms may slip out first and are then corrected by the speakers, e.g., *tšox/tšok* ‘very’.

Adverbs of degree modify adjectives (61–63), other adverbs of degree (62) or manner (63), and verbs.

- (61) *pola kalabaluk ejendune*  
 ‘It became very crowded.’ (01\_06042017F\_4; 017)

<sup>44</sup> Nearly all derived adverbs in Romeyka are diachronically case-forms of adjectives (van Emde Boas et al. 2019: 83).

<sup>45</sup> *Monos* is classified in AG as quantifier (van Emde Boas et al. 2019: 358).

- (62) *mono fazla konuŝevume*  
 little more speak.PRS.1SG  
 ‘I speak a bit too much.’ (01\_28062019F\_3; 17)
- (63) *mono džidin ebsindane*  
 ‘They cook a bit difficult.’ (03\_30062019F\_11; 090)

Note that *bola* is also used as quantifier modifying count and mass nouns (Section 3.2.3) and as predicative adjective. As quantifier (64, 65) and predicative adjective (66), *bola* agrees with the head in number (*poli(n)*.SG, *pola*.PL < AG *πολύ-ς* ‘many’, *pola*.N.ACC.PL, Holton et al. 2019: 833).<sup>46</sup> The same applies probably to *eliyo* ‘a little’ (< AG *(ó)liyo(n)*.N.SG ‘a little’ < *óliyo*s ‘few, small’, van Emde Boas et al. 2019: 4) in (66).

- (64) *plesame bola opsara*  
 ‘We caught many fish.’ (08\_04072019M\_1; 162)
- (65) *bola đromos*  
 ‘much way’ (04072019F\_10; 12)
- (66) *ama do vutero poli do vutero bolin adone eliyo utš en do diri*  
 ‘But [there was] plenty of butter, the butter was plenty, the cheese was not little.’  
 (04\_01072019F\_1; 190–192)

When modifying adjectives or adverbs, the adverb of degree strictly precedes the head (exs. 61–63 above; also cf. Drettas 1997: 155). When modifying a verb, most adverbs of degree are pre-verbal (67–70; also cf. Drettas 1997: 155, *pola kalatševs* ‘you talk a lot’), although post-posed adverbs are not rare (71–74, also see 75 for both positions). There seem to be several variables affecting the syntactic position of an adverb of degree, potentially among them semantic meaning, syntactic scope, information structure, and general word order-directionality (see Section 5.2.1.1). Further research is required here.

- (67) *ton džiri=m bal bola eyabo*  
 ‘I love my father a lot.’ (02\_21042018M\_2; 30)
- (68) *ama džine ba bola jašlanefđane*  
 ‘But they have aged a lot.’ (04\_01072019F\_5; 35)
- (69) *t=aleyō mono tšilien*  
 ‘The horse stumbled a bit.’ (08\_04072019M\_2; 108)
- (70) *mono katsa*  
 ‘I sat a while.’ (03\_30062019F\_11; 029)
- (71) *evreksen bola*  
 ‘It rained hard.’ (04\_01072019F\_5; 61)

<sup>46</sup> Note that the form *poli* in ex. (66) is homonymous with the Turkish adverb *bol* ‘abundant’, so the first clause translates into Turkish as *ama yağı bol* ‘there is plenty of butter’. This form (following loan word integration into Romeyka *bol-i*) may interfere with the inherited quantifier *boli/bola*. In general, it is not clear to which degree homophony of Tr. *bol* and R. *bola* impacts upon the frequency of the adverb *bola* in Romeyka. Moreover, it might be possible that Tr. *bol* goes etymologically back to a Greek word.

- (72) *boro na steko bola*  
 ‘I cannot stay much.’ (03\_30062019F\_6; 02)
- (73) *kavurevune mono*  
 ‘We stir a bit.’ (03\_30062019F\_11; 086)
- (74) *s=ena sakul espringsam=a mono*  
 ‘We squeezed it a bit in a bag.’ (03\_30062019F\_1; 21)
- (75) *habadžega monon faneruite mono..*  
 ‘A bit is visible from here.’ (07\_04072019F\_6; 24)

#### 3.1.4.2 Adverbs of manner

Adverbs of manner (e.g., ‘well, ‘quickly’) are partly derived from adjectives. They usually modify verbs or other adverbs. Manner adverbs differ from predicative adjectives in that they cannot occur as predicative complements. While those manner adverbs that are “true” members of the class of adverbs are invariable, manner adverbs that are derived from adjectives (e.g., *kala* ‘well’ < *kalos* ‘good’) share largely the properties of predicative adjectives (Section 3.1.2.2). They are expected to agree in principle in gender and number with the subject and can potentially undergo comparison, although no data are attested for this (but cf. Holton et al. 2019: 841 for the diachronic situation).

Manner adverbs based on adjectives in *-os* (or *-is*) take as adverbs the accusative plural ending in *-a* (e.g., *kal-os* ‘good’ > *kal-a* ‘well’, see Holton et al. 2019: 833).<sup>47</sup> These adverbs are invariable (see 76–78). But note that the hapax legomenon in ex. (79) seems to suggest that potentially there might also occur feminine declension forms in *-i*.<sup>48</sup>

- (76) *teke u boro na yonuševum emorfa*  
 ‘On my own I cannot speak nicely.’ (03\_07072019F\_1; 35)
- (77) *kala epitšen adena*  
 ‘She did [it] well to her.’ (02\_29062019F\_1; 20)
- (78) *kala maθison=a*  
 ‘Teach her well!’ (01\_02022015F\_1; 04)
- (79) *kali eftas*  
 ‘You.F do it well.’ (03\_30062019F\_7; 08)

Adverbs of manner occur both pre- and post-verbally (80 vs. 77/78).<sup>49</sup> While adverbs of degree are predominantly specifier-type adverbs (or adjuncts), manner adverbs are often more complement-like, which allows more syntactic flexibility (Alexiadou 1997: 80–83). However, manner adverbs can be assumed to occur in default position post-verbally (80, 81; cf. also Bağrıaçık 2018: 205–207 for Phrasiot), whereby pre-position arises due to information structure (82–84): in general, the preverbal position is reserved for focused elements, whereby

<sup>47</sup> But cf. Alexiadou (1997: 3), who notes for SMG a different semantic meaning of adverbs in *-os* and in *-a*.

<sup>48</sup> Note, however, that this form was produced by a heritage speaker whose Romeyka competence is mainly limited to comprehension, so the form is dubious; probably there is analogy to *poli* ‘many’ or the form occurs as a result of vowel raising before the consequent high vowel.

<sup>49</sup> This is in contrast to SMG, where “bare manner adverbs like *kala* ‘well’ are ungrammatical in final position” (Alexiadou 1997: 11).



postverbal position is non-focal. The adverbials in exs. (82–84) are focused, while those in (81) are non-focal suggesting that the information has been mentioned before and is known to the hearer. According to Neocleous (2020: 102), manner adverbs can even appear between auxiliary and main verb.

- (80) *u=ksero kala*  
‘I am not sure.’ (01\_04022016F\_1; 035)
- (81) *jirevume erxumestine kal hajes endama*  
‘We cast [them] up and we come again together like this.’ (04\_01072019F\_1; 017–018)
- (82) *kal eyo kruy=atona*  
‘Again I hit him.’ (02\_2906019F\_1; 16)
- (83) *haets efoveθame*  
‘Like this we were afraid.’ (02\_29062019F\_2; 41)
- (84) *emistine haets eðevasam da imeras*  
‘Thus we spend our days.’ (08\_04072019M\_1; 045–46)

As for their lexical distribution, according to Payne (1997: 69), manner adverbs form the largest subcategory of all adverbs in a language. This seems not necessarily to hold true for Romeyka, at least not for invariant adverbs; manner adverbs derived from adjectives are more productive. The number of total occurrences of manner adverbs (including adverbs of degree) is in the corpus with 354 tokens slightly lower than that of temporal (415 tokens) and spatial adverbials (424 tokens). The class of (invariant) adverbs appears to be relatively “closed” as it were with a limited number of lexemes, many of which occur with high frequencies and only limited borrowings. The most frequent invariable adverb of manner (64 tokens) has the meaning of ‘like this, such’ and exists with different variants: *(h)a(j)e(ts)*, *(h)ai(s)*, *hats*, *(h)aidi(ko)*; the latter variant (probably derived from an adjective) inflecting for number.

The following adverbs that belong more or less to the group of adverbs of manner are borrowed from Turkish: *fazla* ‘much (more)’, *as(l)inda* ‘actually’, *pek* ‘relatively’ (together with the Tr. clausal negator *jok*), *sade* ‘only’, *zate(n)* ‘after all’, *tek(e)* ‘alone, only, single’, *airija* ‘separately’, *hizlija* ‘quickly’, *andžak* ‘solely’, *baştan* ‘initially’, *jaja* ‘by foot’ (cf. *porpatix-da* ‘walkingly’ participles in Section 4.3.7.3), *mesela* ‘for example’, *tabi* ‘of course, naturally’. Adverbs borrowed from Turkish follow largely the syntactic properties of inherited adverbs (85). No adverbs associated with epistemic meaning have been found.

- (85) *hitš utš enumunesten andalo ta peðia airija portanenane*  
‘We never were together, the boys walked on their own.’ (02\_02022015F\_1; 075–076)

Finally, the reduplication of manner adverbs well-known from Turkish (Kornfilt 1997: 466) is copied in Romeyka: either by calquing with inherited words (86–89) or by copying the Turkish matter and pattern altogether (90, 91; see also intense code-switching in 92).

- (86) *eθekane tria xorja xorja*  
Tr. *bıraktılar üç tane ayrı ayrı*  
‘They put three [buckets], separately.’ (05\_03072019M\_4; 11–12)
- (87) *daxtira daxtira epije tše sade jelane eseve so spidi*  
‘He ran quickly and headed home laughing.’ (C1)

- (88) *yále yále tsíma tsíma eporpátesa tše ekséva son parxári [...]*  
 ‘Slowly, slowly, on foot, I could head towards the pasture [...].’ (Tursun 2019: 496)
- (89) *eh.. eliyo eliyo apomaθáno stéko*  
 ‘Slowly, slowly, I am weaning.’ (Tursun 2019: 142)
- (90) *argo javaše javaše erθe*  
 ‘The bear came slowly.’ (04\_01072019F\_12; 43)
- (91) *ebetši ebetši jada tšabuxa tšabuxa erθam*  
 ‘We returned from there very quickly.’ (08\_04072019M\_1; 182)
- (92) *jane herkes seve seve buranin fasuljesini ebere*  
 ‘Everybody buys gladly beans from here.’ (08\_04072019M\_1; 136)

### 3.1.4.3 Adverbs of time

Many temporal adverbs are part of aspectual/quantificational adverbs that provide information about definite time (*extes* ‘yesterday’), indefinite time (‘often’, ‘yet’) or frequency (definite (/cardinal count) or indefinite(/durative); *her gūn* ‘every day’, *panda* ‘always’, *mian* ‘once’). Most of temporal adverbs like *osimero* ‘today’ and *extes* ‘yesterday’ are non-derived and invariable. Others, however, like *embro(n)/embra* ‘before’ are derived from adjectives and seem to show plural agreement with their head. Alternatively, the difference between the two forms of *embro* and *embra* may be rather a semantic one, with *embra* referring to a larger time frame in the past. These derived adverbs can undergo comparison, e.g., *embra-dera* ‘more before’, i.e., ‘in earlier times’.

In terms of their syntactic behaviour, temporal adverbs are at the fringe between being sentence adverbs or VP-adverbs (Alexiadou 1997: 7). In first instance, they modify verbs whereby they are in very free distribution. According to Neocleous (2020), temporal adverbs separate the subject from the verb (in V2 order), but the adverb can also follow the subject in OV orders. According to Alexiadou (1997: 11), in SMG temporal adverbs that denote definite time occupy mostly the clause-final position but can also appear in postverbal position (with V2 word order). If they occur in clause-final position, they are stressed. Temporal adverbs used as sentence adverbs can also occur in clause-initial position, then they are focalized (Alexiadou 1997: 9–14). In the Romeyka corpus, temporal adverbials occur predominantly in immediate pre-verbal position and often following the object NP (93–97), although in the data of Neocleous (2020), most object NPs are either immediately pre- or post-verbal with adverbs occurring after the subject or in clause-initial position. Strong Turkish OV order in Turkish-dominant speakers may have an impact here. Apposition of temporal adverbs is possible as well (98). More detailed research on the syntactic distribution of adverbs in Romeyka is required.

- (93) *emena ifeti jardimi bisun*  
 ‘Help me this year!’ (08\_04072019M\_3; 100)
- (94) a. *o džiri=m mian iða* (02\_21042018M\_2; 11)  
 b. *mian iða tonan* (02\_21042018M\_2; 12)  
 ‘I saw my father(/him) once.’
- (95) *apopse elate s=emena*  
 ‘Come to me tonight!’ (04\_01072019F\_2; 186)

- (96) *extes erθa*  
‘I came yesterday.’ (04\_01072019F\_5; 07)
- (97) *sabale n=arxume*  
‘I will come tomorrow.’ (07\_04072019F\_5; 31)
- (98) *esi ndona bidžes opse tš=osimero tše*  
‘What are you doing, today and tomorrow?’ (04\_01072019F\_5; 06)

If different adverbs co-occur, based on their specific scope restrictions certain hierarchies apply. According to Alexiadou (1997: 9), sentence adverbs are higher than manner adverbs (99, 100), and speaker-oriented sentence adverbs are higher than subject-oriented. Basically, only one temporal adverbial of a kind can occur in a clause, otherwise their syntactic scope must be different (101, but cf. 102).

- (99) *sabaxtan kal esku*  
‘In the morning get up again.’ (04\_01072019F\_1; 158)
- (100) *ebedžinaen kale ebije*  
‘Later, he went again.’ (04\_01072019F\_13; 25)
- (101) *har sabale ibe n=arxume*  
‘Now she said, I will come tomorrow.’ (07\_04072019F\_5; 30)
- (102) *osimero sabaxtan ekativen*  
‘She came up today in the morning.’ (07\_04072019F\_5; 06)

As for the lexical frequencies, quite a variety of inherited temporal adverbs exist in Romeyka, e.g., *osimero* ‘today’, *extes/opse* ‘yesterday’<sup>50</sup>, *sabale* ‘tomorrow’, *t=alo=mera* ‘the day after tomorrow’, *ab=obse* ‘tonight’, *ifetos/ifeti* ‘this year’, *har(e)/harin* ‘now’, *embro(n)/embra* ‘before’, *elijora* ‘early’, *panda* ‘always’, *mia(n)/mijes* ‘once’, *ako(me)* ‘still’, *etote(s)* ‘back then’, *(ep=)ištera* ‘later’. A variety of forms exist for a compound adverb meaning ‘thereafter’: *eb=edžede*, *d=edže=nahen*, *eb=edž=ae*, *tš=eb=edži (=nahen)*.

The times of the day are mostly (based on) Turkish, e.g., *sabaxdan* ‘in the morning’, *(d=)akšem(is)* ‘(in the) evening’, *do kindi* ‘(in the) afternoon’, *gedžejle* ‘in the night’ (but cf. 103). Also, the following adverbs: *bazi/bazen* ‘sometimes’, *eskiden beri* ‘since old times’. *Ondan sora* ‘then’ is frequently used in Turkish-dominant speakers as a filler in descriptions of activities. *Hidž* ‘never’ is borrowed from Turkish and needs to be combined with a clausal negator (like in Turkish; ex. 104).

- (103) *nixta ekusa ti lalian*  
‘I heard the sound in the night.’ (03\_30062019F\_6; 15–16)
- (104) *aets hidž emorfa ðulias udž ebidže*  
‘Like this he never did any good things.’ (01\_06042017F\_4; 137)

<sup>50</sup> According to Oikonomidis (1908: 45), *opse* goes back to *extes* via an intermediate state of *ephθe*, whereby labiodental fricative + dental fricative result in /ps/. *Extes* (< AG χθές ‘yesterday’) itself is probably a fossilized form of an earlier adjective or noun that was lost in AG (van Emde Boas et al. 2019: 86).

Complex temporal adverbial clauses (Section 5.3.3.1.1) are either formed with *anda/onda* ‘when’, *omon* ‘like/when’, or *(o)bote* ‘when(ever)’. Conditional expressions which are somehow formulaic in Turkish are borrowed as well, e.g., *osada šindže* < Tr. *olsa da şimdi* ‘if we (only) had now’.

#### 3.1.4.4 Adverbs of place and spatial orientation

The domain of spatial deixis in Romeyka is complex, which may be related to the geographical particularities of the area; at the same time, this is one of the features that sets Romeyka apart from other Greek dialects. Adverbs of place establish situative definiteness in Romeyka drawing on a complex set of localizers, which may be accompanied by pointing gestures (see also Section 4.2.3). Simple adverbs of place are largely a non-derived category and as such morphologically invariant (but cf. *ebugo/ebuga* ‘under’). Manifold derivation strategies exist to form complex spatial adverbs. For a tentative attempt at providing an overview of the complex spatial expressions attested in the present Romeyka corpus, see Table D.1 in Appendix D.<sup>51</sup>

In total, adverbs of place appear in the Romeyka corpus approximately as frequently as adverbs of time (424 tokens of spatial adverbs, 415 tokens of temporal adverbs); however, the relational occurrence of each lexeme is much lower, indicating the high number of different lexemes. This high number is due to the fact, as Drettas (1997: 449) states, that adverbs of place and spatial deixis in PG are located at the interface of lexicon and morphology with many spatial expressions going back to derived forms. The encoding of topographic information by means of these complex forms is partly grammaticalized (see Urban’s 2020 “mountain linguistics”). The individual spatial elements belong to different word classes like nominal or verbal lexemes, spatial adverbs, and prepositions (Drettas 1997: 450); especially the distinction between adverbs and prepositions is often blurry and shall not be discussed here (but consider for example the terminology of “primary prepositions” and “secondary postpositions” in Karatsareas (2016: 62) for AMG). Here, these classes are subsumed together under the category of adverbs of place according to their function as spatial modifiers. In the following, first the syntactic behaviour of spatial adverbs is outlined. Secondly, some general building principles including simple adverbs and their combinations together with some special localizers are sketched (cf. Drettas 1997: 449–508).

In terms of their syntactic behaviour, location adverbs modify VPs, being like temporal adverbs located at the fringe between sentential and VP-adverbs. Spatial adverbs appear both in immediate post-verbal position (105–109), but also pre-verbally (110). Following Mackridge (1987: 133) spatial adverbs are post-verbal. However, according to the WOVA dataset (Schreiber 2021), only 42% of the locations are post-verbal; it seems word order varies according to speaker. In general, the distribution of location adverbs is similar in flexibility to that of PPs, which are very mobile (Section 5.2.1.1). In the WOVA dataset, 78% of goals are post-predicate. Note that in some cases, spatial adverbs occur together with the locative preposition *s*, e.g., *s=avudo ađaha* ‘here’ (05\_03072019M\_3; 20); (111). Location adverbs may even occur between modal and main verb (112).

Spatial adverbs can combine with (object) personal pronouns (113, 114, also ex. 126 below).

(105) *eksangame eksumer*

‘We took them [=the animals] outside.’ (03\_30062019F\_2; 28)

<sup>51</sup> Table D.1 in Appendix D is to be read from left to right, tentatively displaying additional grammatical information attached to a spatial element to the right. Morpheme boundaries are mostly indicated, but unclear segmentations are left out.

- (106) *evalam abes*  
‘We took them [=the animals] inside.’ (03\_30062019F\_1; 13)
- (107) *eyo an bao eksu ruizo so nero*  
‘When I go outside, I fall into the water.’ (04\_01072019F\_13; 22–23)
- (108) *do havas en hadžega*  
‘How is the weather there?’ (03\_07072019F\_1; 40)
- (109) *o zaman na bame eban so xorio*  
‘At that time, we will go up to the village.’ (04\_01072019F\_17; 60–61)
- (110) *emis edžaha epejenane*  
‘We used to go there.’ (09\_04072019\_7; 18)
- (111) *erθa s=ađaha katsa ga*  
‘I came here and sat down.’ (03\_30062019F\_6; 34)
- (112) *utš eθelene buθen ebejene*  
‘She didn’t want to go somewhere.’ (01\_04022016F\_1; 105)
- (113) *ondan sora ejnedžikse epan=am*  
‘Later, he married over my head.’, lit. ‘over me’ (01\_06042017F\_4; 133)
- (114) *ama jela erota=me, bal kad exo apes=am*  
‘But come, ask me: I have something inside me.’ (03\_30092019F\_7; 35)

Several types of adverbs may occur together, even several location adverbs, as complex adverbs consisting of different spatial elements (from different word classes) prove, e.g., *ađa opis* ‘behind here’ (08\_04072019M\_2; 020); (115). Adverbs of definite location (e.g., ‘here’) modify adverbs of spatial relations going back to prepositions (e.g., ‘below’). Spatial adverbs appear higher than adverbs of manner (116).

- (115) *habadži beran asa rašia jaja ebeinane me ta za*  
‘From opposite there to the mountains they went with the animals by foot.’ (08\_04072019M\_1; 103–104)
- (116) *abadutšam hajes epejnane*  
‘From up there they used to come like this [gesture].’ (07\_04072019F\_10; 11)

Simple adverbial spatial expressions occur frequently together with the prepositions *s-* ‘at, to’, *as-* ‘from’, *os* ‘until’ (Drettas 1997: 454). The localizers *ađa* and *etši* often precede a prepositional phrase that specifies them (117). They can, however, also be directly combined with a preposition, depending on the valency of the verb (118), or they occur without a preposition (119).

- (117) *ađa so mehele*  
‘here in the neighbourhood’ (08\_04072019M\_1; 060)
- (118) *erđa s=ađaha katsa ga*  
‘I came here and sat down.’ (03\_30062019F\_6; 34)

- (119) *alon d=ajarevume hatsi tše hađa*  
 ‘What else are we looking for here and there?’ (08\_04072019M\_1; 245)

The simple spatial adverbial expressing a direction of movement (*e*)*ka* ‘down’ is not easily distinguished from a preposition and occurs in compound verbs of motion like ‘sit’, ‘put’, ‘pass’, ‘lay’ (120).

- (120) *ta xorafae ula ha eθedž ega*  
 ‘The fields were put down.’ (03\_30062019F\_6; 54)

Simple (and compound) spatial adverbials denoting primarily a location like *embro* ‘before’, *ebuga* ‘under’ can be used without prepositions (121) or they can also be further specified by a prepositional phrase (122, see also 123). In this case, the sequence of prepositional phrase and the adverbial is variable (124 vs. 125), probably with a difference in focus. Furthermore, simple spatial adverbials can be used in a nominalized form, e.g., *i ađadžega* ‘here’ (08\_04072019M\_1; 282); (126).<sup>52</sup>

- (121) *erθame eban*  
 ‘We came up.’ (03\_30062019F\_1; 14)
- (122) *so spidi embro enan ido ormi*  
 ‘In front of the house [is] a river.’ (04\_01072019F\_13; 04)
- (123) *son baxari tšan ebiyam adutšam bola drana opsara ine*  
 ‘We went up to the pasture; up there are many beautiful fish.’ (08\_04072019M\_1; 156)
- (124) *eđekam=ada s=apsomo ban*  
 ‘We put them on the stove.’ (03\_30062019F\_1; 16)
- (125) *o zaman na bame eban so xorio*  
 ‘At that time, we will go up to the village.’ (04\_01072019F\_17; 60–61)
- (126) *eksendže d=eban=adjes*  
 ‘She took off her above [= her jacket].’ (04\_01072019F\_5; 43–44)

More complex adverbs like *etšiega* ‘there’ occur without a preposition (127) frequently in pre-verbal position (128) although they also appear post-verbally (129, 130) (see also Section 5.2.1.1).

- (127) *etšiega tu spidi do duvari xala en*  
 ‘The wall of the house there has collapsed.’ (08\_04072019M\_3; 056–057)
- (128) *etšiega esdegame*  
 ‘We stayed there.’ (08\_04072019M\_3; 084)
- (129) *ebiyanem etšiega*  
 ‘They went there.’ (08\_04072019M\_3; 183)

<sup>52</sup> Note also the strange predicative use in ex. (15) of this Chapter, (partly) repeated here for convenience (i):

(i) *ama ebebuka asi yoryoran ebebae udž en*  
 ‘But from upper Gorgoras, it is not the lower.’ (01\_04022016F\_1; 052)

- (130) *eksangame eksumer*  
 ‘We took them outside.’ (03\_30062019F\_2; 28)

Still, complex adverbials with special localizers can be followed (131) or pre-ceded (132) by a specifying prepositional phrase.

- (131) *habađadžega as=emetero din džami buga gaheves ine tši*  
 ‘At this side from our mosque below there is a café.’ (08\_04072019M\_3; 157–159)

- (132) *s=avudo ađahadžega*  
 ‘here at this (very) side’ (05\_03072019M\_3; 51)

In sum, spatial expressions in Romeyka are inherited. Turkish adverbs of place like *arasına* ‘in between’, *buranin* ‘from here’, *üzerinden* ‘above’, *dışardan* ‘from outside’ are occasionally used in code-switching when single Turkish phrases (often nominal or prepositional phrases) are inserted; this depends on the speaker, though. For forms of the adverb ‘here’/‘there’, there seems to be a continuum reaching from the Turkish forms *bur-da* ‘here’/*bura-lar-da* ‘around here’ and *or-da* ‘there’, via the regional variants *ha-bur-da* ‘here’, *ha-ore-ja* ‘to there’ to Romeyka forms like *ha-vu-da* ‘here’, *ha-vu-džega/-džijega* ‘this side’.

#### 3.1.4.4.1 Simple adverbs of place

Simple adverbs of place are *ađa* ‘here’ and *etši* ‘there’. Drettas (1997: 462–463) additionally describes *atu* as a third option in expressing distance for a more specific location. In the Romeyka corpus, *atu* appears less frequently and only together with suffixed spatial elements, e.g., *adu-tš-an/ado-tš-an* ‘up there’ and *(h)adu-džeka* ‘there (side)’. It is not clear whether this form contrasts with forms like *(h)avu-džega* ‘here (side)’, lit. ‘this side’; the localizer *havu* (probably derived from the Greek demonstrative pronoun *avudos*) is in Trabzon Turkish of the area used interchangeably with regional Turkish *habu* ‘this’. The existence of three simple locational adverbs which indicate three degrees of relative distance resembles the three-fold Turkish system with *burada* ‘here’, *orada* ‘there’ and more specific and less frequent *şurada* ‘~yonder’ (Kornfilt 1997: 311).

On the vertical axe, Romeyka distinguishes between adverbials that can be used both as locations and directions, *(e)ban/apan* ‘on’ and *(e)buga/(e)bugo* ‘under’, and the processual adverbs *an* ‘up’ (< AG *άνω* ‘above, upwards’, van Emde Boas et al. 2019: 86) and *(e)ka* ‘down’. Drettas (1997: 466–474) also mentions *afka* ‘below’ which does not figure in the corpus, though. Instead, another form appears in the corpus, *katu* ‘below’ (< AG *κάτω* ‘below, downwards’, van Emde Boas et al. 2019: 86; 04\_01072019F\_2; 134). *(E)ka* appears frequently after verbs, e.g., *kaθ eka* ‘sit down’,<sup>53</sup> while *an* only appears in the form *tš-an* (< AG *και άνω* ‘and up’, Mackridge 187: 133) in compound adverbials (133). Drettas (1997: 496, see also 493, 460) discusses a potential etymology for this form going back to *an* ‘up’ with the prefixed conjunction *tše* ‘and’. While the use of *an* is restricted in the Romeyka corpus, *eban* is also used as a direction rather than a static location (ex. 134).

- (133) *ab-adžan ab-adžan ab-adu-džan hajes epejnane*  
 ‘They went from up, from up there like this.’ (04072019F\_10; 10–11)  
 vs. *(h)ab-aržan* ‘from up there’ (07\_04072019F\_6; 15, 07\_04072019F\_6; 18)

- (134) *erθen tomara ksendžen ata eban dže bič* (07\_04072019F\_8; 17)  
 ‘He came, pulled them out all together and left.’

<sup>53</sup> But see Section 4.1.1.4 for derived verbs including spatial specifiers.

On the horizontal axe, *embro* refers to things in front of something (also temporal ‘before’) while *opis* means ‘behind’ (< AG *opíσω* ‘backward’, van Emde Boas et al. 2019: 86). In terms of intern/extern, *apes* refers to the direction of inside, often used together with verbs of movement, e.g., *eseven apes* ‘he entered inside’. *Eksu* (< AG *εξω* ‘(to) outside, away’, van Emde Boas et al. 2019: 87) refers to the direction of outside, e.g., *bayo eksu* ‘I go outside’ (04\_01072019F\_13; 22–23). When it comes to expressing proximity/distance, no form for “close” is attested in the Romeyka corpus (Drettas (1997: 482–483) mentions *suma(n)* for PG) and long distance is expressed by the adjective *makra* ‘far’. A particularity of PG is the use of the adverb *(e)pera(n)* referring to a distant location which is at sight but separated from the point of reference by for example a valley. Laterality is expressed by the Turkish loan *jan* ‘side’, which appears together with a possessive pronoun, e.g., *so jan-is* ‘to/at your side’ (01\_04022016F\_1; 038–039; cf. Drettas (1997: 480)). The adverb of *(a)plan* refers to a relative lateral proximity (Drettas 1997: 481–482) but figures only once in the Romeyka corpus in the compound form *ab-adž-ablan* ‘from that side’ (04\_01072019F\_5; 52).

Simple spatial adverbs can be combined with each other, e.g., *edž-ega/adž-ega* ‘down there’ (e.g., 07\_04072019F\_6; 04, also *etši-ka*, A1), *edž-eperan* (04\_01072019F\_17; 35), or with some prefixed prepositions indicating a direction of movement, e.g., *ap(o)/ep* ‘from’ > *ap-opis* ‘from behind’, *ep-eksu* ‘from outside’ (see also 135).

- (135) *ap-aða-džan ap-atši-peran-džes išlajevanje so kušmer*  
 ‘From up here from opposite there they pass’ to Kushmer.’ (07\_04072019F\_5; 46–47)

#### 3.1.4.4.2 Post-posed localizers *-tšega*, *-tšes*, *-merea*

Simple or composed spatial adverbs can be specified by postposed localizers. Since the term to the right specifies the term to the left, stacking is possible (Drettas 1997: 495) and frequently results in complex longer spatial expressions, e.g., *ap-adž-eperan-džes* ‘from opposite there’.

The special postposed localizers which refer to an undefined area are *-tšega*, *-tšes*, *-merea*. Drettas (1997: 493) also considers *-tšijan* (PG *-kjan*) to be part of this category which refers to an undetermined space above the horizontal line (136).<sup>54</sup> Furthermore, he discusses *-ka* as a special localizer semantically referring to a space below the horizontal line and combining mainly with the simple adverbs of ‘here’, ‘yonder’ and ‘there’ but which comes to denote a certain point as a referential anchor point, e.g., *aða-ka* ‘down here, just here’ (Drettas 1997: 493, 503–505). In this form, *-ka* does not occur in the Romeyka corpus, but for a potentially related form *-ha* in Romeyka see below.

- (136) *adžijan adžijan so boyazi*  
 ‘up there at the canyon’ (08\_04072019M\_2; 087)

*-Merea* refers to a zone without clear limitation (Drettas 1997: 488). It is always suffixed to other spatial adverbials, e.g., *eksu-mer(ea)* ‘to an outside (area)’ (03\_30062019F\_2; 28; ex. 137 and 138). But also note the existence of a noun *to meros* ‘side’ (ex. 22 in Chapter 4). *-Tšes* refers to a terrain on a horizontal line (Drettas 1997: 496) with an approximate meaning of ‘side’ (139). The same applies to *-tšega* albeit it refers to an area at closer proximity to the point of reference (ex. 140; cf. Drettas 1997: 506). Interestingly, *-tšes* can apparently also specify temporal adverbials (141). As for the etymology of *-tšes* and *-tšeka*, Papadopoulos (cited in Drettas 1997: 496) suggests for *-tšes* a basis of *ke eso* ‘and inside’. *-Tšeka* could possibly go back to *ke ka(t(u))* in analogy to *tš-an* ‘and up’. However, in another analysis, there might be a

<sup>54</sup> It is at this point not fully clear whether *-tšijan* is distinct from *tš-an*. It might be that *tš-an* as a simple localizer can be combined with other simple localizers, while *atšijan* is a localizer in its own right.



relation between the two forms proposing that *-tšes* is the plural form of a noun *tšeka* ‘side’ > *tšekes* > *tšes* ‘sides’. Both *-tšes* and *-tšeka* can not only be attached to simple spatial adverbs but also directly to the NP, e.g., to nouns, proper names, pronouns or possessive suffixes (142 and 143). On top of that, they seem to combine also with definite articles like neuter singular nominative *to* (144; probably also including the homophonous interrogative pronoun (*n*)*do* ‘what’, see Section 3.1.4.4.4 below) and the feminine singular accusative determiner *ti* following a preposition (ex. 145), e.g., *di-džega* (08\_04072019M\_3; 180), *di-džes* (03\_07072019F\_1; 42).<sup>55</sup> The form *tš-ab-adriž-ga* ‘as of here’ in (145) is to some extent intransparent although it also shows the localizer *-ka* in its use as an “anchor point”.

- (137) *so rašin epan-merea*  
‘towards the mountain top area’ (01\_28062019F\_2; 19)
- (138) *s=oros apes-merea ime*  
‘I am somewhere inside the forest.’ (01\_28062019F\_3; 42–43)
- (139) *eskiden beri h-aða-tšes imestine*  
‘We have been here since long.’ (08\_04072019M\_1; 005–006)
- (140) *ula adu-džega ixa*  
‘I had them all here.’ (08\_04072019M\_2; 057)
- (141) *jani gözüz na en tomara na başlaeps so kseraθinimo adode-džes*  
‘When it will be autumn, we start the drying all together around that time.’ (03\_07072019F\_1; 44)
- (142) *i resmiye-džes ebejnede*  
‘You were about to go to Resmiye (’s side).’ (03\_07072019F\_1; 13)
- (143) *erθesana s=emetera-džeka so.. ađio ma su tšur=um-tšeka*  
‘They came to our side, without? [coming?] to the side of my father.’ (07\_04072019F\_6; 07)
- (144) *do-ika ixa tikanin*  
‘I had a shop here.’ (08\_04072019M\_2; 062)
- (145) *osa di-ka me din araba farina=da tšabadrižga me t=aliyo*  
‘Till there I brought them by car, as of there? with the horse.’ (08\_04072019M\_2; 097–098)

### 3.1.4.4.3 Pre- and suffixal demonstrative *ha*

For the environments for which Drettas (1997: 503–505) describes the function of *-ka*, the Romeyka corpus shows composed forms with the suffix *-ha*, e.g., *aða-ha/eða-ha* ‘(just) here’, *adža-ha/(e)dže-ha* ‘(just) there’, and other combined forms of simple spatial adverbs like *eb-edža-ha* ‘from (just) there’ or *ða-ha-buko* ‘down here’. While *-ha* is not at first sight Greek, Maria Petrou (p.c.) considers a possible relation to Cypriot Greek *eđo* ‘here’ or *ðaha* < *ða-hame* (< *hamo/hame* in AG). Areally, similar forms exist in Northern Kurdish with postposed *-ha* following deictic elements (G. Haig, p.c.). Based on ex. (146), it seems that *-ha* as a specifier

<sup>55</sup> Consider also the unclear forms in (ii).

(ii) *kate isis d=edžadže tše vreš eravadžes de raeva*  
‘It is somehow dreary there and rains, what do I seek there?’ (01\_04022016F\_1; 128–129)

of a certain place can be stacked, which would be reasonable if we assume its function to express local specificity and a demonstrative meaning.

- (146) *h-ab-adžega monon .. monon faneruite mono h-ap-ađa h-ap-ađa đa-ha-ha*  
 ‘From here a bit... a bit is apparent, from here, from here.’ (07\_04072019F\_6; 23–24)

Apart from the postposed suffix *-ha*, there is also a prefix *h(a)-* which occurs not only for spatial adverbials but also for demonstrative pronouns both in Romeyka and the Trabzon Turkish vernacular, e.g., Tr. *ha-bu* ‘this’ (R. *h-avudo* ‘this’ < MedGr *οὗτος* ‘this here’), *ha-orda* ‘there’ (see also Section 2.1.2.5). According to its use, *h(a)-* seems to function as a specifying prefix with deictic/demonstrative meaning in more than one context.<sup>56</sup> In spatial deixis, it combines with simple adverbs of space or with special localizers, e.g., *h-ađa-pisa* ‘behind’ (04\_01072019F\_7), *ha-džega* ‘this side’ (04\_01072019F\_13; 20), cf. *hao-džega* ‘this around here’ (08\_04072019M\_1; 288).<sup>57</sup> However, when combining with simple spatial adverbs, it never combines with the variant of the adverb in /e/ like *eb-edži* ‘from there’ but rather the variant in /a/, *h-ab-adži* ‘from there’, which proves phonological sensitivity. When comparing the forms in *h(a)-* and *-ha*, the semantic difference becomes not quite clear, e.g., *ađa-ha* ‘here’ vs. *h-ađa* ‘here’ (see also ex. 147). Also, the difference between forms with and without *h(a)-* is not clearly discernible, phonological conditioning like /h/ following a vowel in the coda of the preceding phonological word seem not to apply, though (ex. 148).<sup>58</sup>

- (147) *h-ađa so boyazi tšan en*  
 ‘It is above the valley here.’ (08\_04072019M\_1; 208)
- (148) *hadudžega bejname da tikana adudžega sane*  
 ‘We went over there, the shops have been over there.’ (08\_04072019M\_2; 047)

#### 3.1.4.4.4 Other complex spatial expressions

Another ending that occurs in some spatial but also some temporal adverbials is *-(n)ahen/-θen*, e.g., *ka-hen* ‘down’ (TD), *eb-edži-nahen* ‘thereafter’ (04\_01072019F\_17; 38). Drettas (1997: 508) describes it as an old derivation suffix which is not productive anymore but whose remnants still occur in some adverbials. Probably, it goes back to the AG ablative suffix *-θεν* [-tʰen], which was also used with nouns but seems to survive in Romeyka only in combination with adverbs: *kapo-θen* ‘from somewhere’ (< MedGr *κάπο-θεν*), *pu-θen* ‘from where’ (< MedGr *πόθεν*, M. Janse, p.c.). It seems to have a semantic meaning like ‘approximate’. (149) lists its occurrences in the corpus (note its occurrence in the quantifier).

- (149) *kapo-θen* ‘from somewhere’ (03\_30062019F\_11; 083)  
*e-bahen ebuka-hen* ‘up and down’ (04\_01072019F\_2; 147)  
*h-atši-buka-hen* ‘from down there’ (07\_04072019F\_6; 01)  
*(h-)ađa-bahen(-djes)* ‘from here up’ (08\_04072019M\_3; 011; 024)  
*ebuka-hen-džes* ‘from (around) below’ (04\_01072019F\_1; 182)  
*hado-bahen* ‘from here?’ (08\_04072019M\_3; 046)  
*olon-ahe* ‘everything?’ (09\_04072019\_7; 45)

<sup>56</sup> It is not clear whether the informal regional answer particle *he* ‘yes’ is related to *ha*. If so, the regional Turkish question tag *he-mi* (Section 5.2.3.3.2) could be potentially also related.

<sup>57</sup> M. Janse (p.c.) reports the use of the prefix *ha-* in Cappadocian together with the manner adverb *atsá* ‘like this’ > *ha-tša* ‘like this, thus’, also *ha-útša* (< MedGr *atsá*).

<sup>58</sup> Tursun (2019: 68) notes that the adverbials with initial *h(a)-* occur preferably in clause-initial position. For the form *ađa-ha*, he suggests an etymology from AG *eđó đa* > *eđó θe* > *ađa-ha*.

Finally, in complex spatial elements also a combination with interrogative/relative *pu* ‘where’ occurs (150).<sup>59</sup> Drettas (1997: 495) describes the meaning of the following complex adverbials as *pu-džega* ‘where exactly’, *pu-tšes* ‘where, to which direction’. The forms in (151) occur in the corpus.

- (150) *os pu-džega pijede do-tšes eperan joksa*  
 ‘Until where did you go? Did you go over and back, or?’ (04\_01072019F\_17; 33)
- (151) *budžedža* ‘everywhere’ (01\_28062019F\_3; 51)  
*buθen* ‘somewhere’ (01\_04022016F\_1; 105)  
*buθendžega* ‘whereelse’ (08\_04072019M\_1; 254)

### 3.1.5 Verbs

Verbs have the following distributional properties: they mark events in a text, function as head of the verb phrase (VP), and predicate of the clause (Payne 1997: 47). Verbs generally govern arguments bearing certain semantic roles like agent, instrument, experiencer, recipient, patient. Morphologically, verbs in Romeyka show subject agreement in person and number, express some tense-aspect-mood (TAM), and may also inflect for valency changing operations and nominalization. Several categories, such as clausal negation, and some kinds of modality, are expressed periphrastically in the VP. Verb inflectional classes affect agreement and TAM marking (Section 4.3.1). Several strong verbs exist and cause a relatively high amount of (inherited) suppletion in the verbal tense paradigms. On the infinitive, a notorious topic in Greek linguistics, see Section 4.3.7. Verbal inflection is described in Section 4.3. For verbal derivation see Section 4.1.1.4. The general structure of the verb word is (TENSE-)ROOT-AGR/TENSE. The citation form of verbs is the first person singular present tense. A VP consists minimally of a finite verb. Auxiliaries precede the main verb, which is usually finite, although in complementation, an (increasing) number of non-finite strategies exist. In general, the existence and function of auxiliaries in Romeyka, like *exo* ‘have’, is awaiting further research; in the present corpus, auxiliaries are not very prominent (but see Neocleous & Sitaridou 2022). The order of verb and its arguments is variable between SVO and SOV (Section 5.2.1.1). The copula *ime* ‘be’ seems to play a special role in word order variation; the theoretical discussion about a possible enclitic status of *ime* ‘be’ is not taken up here. For VP operations and verbal syntax, see Sections 5.2 and 5.3.

As for the frequency and distribution of lexemes (token frequency), the Romeyka corpus lists 2,811 verbal forms: 1,221 in present tense and 1,424 in past tense (aorist or imperfective), as well as about two dozen non-finite forms (incl. participles). This distribution of tense forms is likely caused by the elicitation methods applied (i.e., free narratives). Frequent lexemes are the copula *ime* ‘be’, *éxo* ‘have’, *ínume* ‘become’, *érxome* ‘come’, *pá(y)o* ‘go’, *eftáyo* ‘make’, *thélo* ‘want’, *tróyo* ‘eat’, *léyo* ‘say’, *téro* ‘look/watch’, *đíyo* ‘give’, *eyválo* ‘go up’, *sevéno* ‘enter’, *féro* ‘put’, *kséro* ‘know’, *bóro* ‘can’, *stéko* ‘stay’, *káhume* ‘sit’. For the most frequent inflected verb forms in the Romeyka corpus see (152).

- (152) (i) 3<sup>rd</sup> person singular present and aorist tense  
 (ii) 1<sup>st</sup> person singular present tense and 1<sup>st</sup> person plural aorist tense followed by their respective counterparts in number (V.PRES.1PL, V.AOR.1SG)  
 (iii) 3<sup>rd</sup> person plural in aorist and present  
 (iv) 2<sup>nd</sup> person singular present and 2<sup>nd</sup> person singular imperative.

<sup>59</sup> Note that the combination of interrogative elements with other spatial elements may also extend to the interrogative (*n*)*do* ‘what’, e.g., *do-tšes eperan* ‘what.side opposite’ (04\_01072019F\_17; 33).

The amount of loan verbs in the Romeyka corpus is around 3,5% (for a statement on the definition of “loans” in the present thesis, see Section 3.1.1.1). Turkish loan verbs can be roughly separated into four functional groups: (a) Some Turkish verbs are borrowed where no Romeyka verb is known or where the concept is more readily associated with Turkish, e.g., *okujevo* ‘study’, *uraševo* ‘deal’, *jašajevo* ‘live’, *kætševo* ‘migrate’, *izlejevo* ‘watch (TV)’, *kavurevo* ‘mix’, *kulanjevo* ‘use’, *inanevo* ‘believe’, *kizevo* ‘be angry’, *arajevo* ‘seek’; (b) For frequent verbs where an inherited Romeyka verb exists, integrated Turkish loans provide a double set of verbs: *konuševo* ‘speak’ (*lalo* ‘speak’), *tšališevo* ‘work’ (vs. *đulevo* ‘work’), *düšünevo* ‘think’ (vs. *nunizo* ‘think’ vs. *θaro* ‘believe’), *gæsterevo* ‘show’ (vs. *đikno* ‘show’), *sevinevo* ‘rejoice’ (vs. *xareno* ‘rejoice’). The semantic coverage of the loan verbs seems in most cases not to differ considerably from that of the inherited words, so the choice for an inherited or borrowed verb might be related to the dominant language of the individual speaker; the pragmatic circumstances favouring the selection of a Turkish loan against an inherited verb need more detailed research; (c) The borrowing of complex Turkish verbs seem to be a means to express some morphological functions like causatives or passives (e.g., *gez-dur-evo* ‘walk sb. around’) or at least, complex Turkish verbs seem to be readily borrowed for their expression of semantic verbal categories for which the Romeyka equivalent seems to be less easily available (e.g., passive/reflexive *ješlanevo* ‘grow old’), see also Section 4.1.4; (d) Turkish predicative expressions may be for pragmatical reasons borrowed as fixed forms – often in the regional Trabzon Turkish dialect – including Turkish verbal inflection, e.g., *ol-sun* ‘so be it’, *bošli-jum* < Tr. *borçlu-yum* ‘I am indebted’, *arasinda-juk* < Tr. *arasinda-yız* ‘we are in between’ (see also Mackridge 1987: 134). For the morphological (and phonological) integration of Turkish loan verbs see Section 4.1.4.

## 3.2 Function word categories

### 3.2.1 Articles

This section covers definite and indefinite articles in Romeyka. Articles in Romeyka express information status of NPs (see also Section 4.2.3. on definiteness and related categories). They have both an anaphoric and a non-anaphoric function and may draw on general and inferred knowledge (see Dryer 2013b). Articles in Romeyka are free words preceding the head (and some modifiers); they occur before nouns in all grammatical roles, before proper names, and before modifiers of the NP, that is, adjectives and numerals but usually not quantifiers.<sup>60</sup> Borrowed Turkish nouns are generally also assigned an article, e.g., *t=akšam* ‘(in) the evening’, *do konušema* ‘the speaking’ (but cf. Neocleous 2020: 38, especially ex. (27); for gender assigned to loan words see the subsequent Section 3.2.1.1, also Section 4.1.4). The grammatical and syntactical properties of definite and indefinite articles are described in detail below.

The Romeyka corpus contains 1,180 occurrences of articles. The number of 1,396 occurrences of noun forms reveals that in practice not every noun appears with an article. Definiteness is one major reasons for this (see just below and Section 4.2.3). 1,090 articles in the corpus are definite articles, only 90 are indefinite articles. Only about one third of definite articles in the corpus is a plural form. The distribution of gender forms in singular definite articles is relatively balanced: 200 occurrences of the singular feminine article, 195 occurrences of the singular masculine article, and 297 occurrences of the singular neuter article. Regarding the ongoing changes in the gender system, this may be a surprising finding (cf. Section 4.2.1.1).

<sup>60</sup> Diachronically, they also became part of full possessive pronouns, e.g., *to + emon* > *temon* ‘my’.

### 3.2.1.1 Definite articles

Romeyka exhibits three definite articles: the masculine singular article *o*, the feminine singular article *i*, and the neuter singular article *to*. This three-way gender distinction only applies to singular, in plural, there is a two-way distinction between feminine/masculine *i* and neuter *ta*. While gender assignment used to be predominantly grammatical, with a weak association to the biological category of gender, a change took place, presumably starting with the plural, to more semantic gender agreement, based on animacy rather than gender, leading to a two-way distinction: *i* for the feminine/masculine plural of human and animate heads; neuter *ta* for the plural of inanimate heads. This tendency has partially spread into the singular, though in principle a three-way distinction is maintained in the singular.

Definite articles agree with the head in number, gender, and case (see Table 11; for selected examples, see also 153–157 below). ROF as spoken in Uzungöl (and surrounding villages) retains an ancient shape of the masculine/neuter genitive singular article, *tu*, which is *ti/tsi* in other Pontic varieties (Mackridge 1987). The masculine accusative plural article in ROF as spoken in Uzungöl is *tus* which is also a diachronically inherited form and which appears elsewhere in the form *tsi/tsu/ti/tin/tis* (Mackridge 1987: 132). Note that the Romeyka corpus features *tu* in the masculine/neuter singular genitive, but the masculine accusative plural article is *tsi*. In the masc./fem. accusative singular, it is not quite clear what determines the occurrence of final /n/ in the articles *ti(n)* and *to(n)* – both diachronically inherited forms: while the form with word-final /n/ occurs frequently before vowels (but cf. *ti emsale* ‘the peers’), *tin/ton* are also used before plosives, e.g., *sin borda* ‘at the door’, *son barxar* ‘to the pasture’, *ton dursun* ‘Dursun’. According to Özkan (2013), the articles in Romeyka present an array of different forms, especially in the masculine and feminine plural. However, it seems that in the plural, case agreement has collapsed to a single oblique case for acc./gen., namely *tsi* for masc./fem., which is a rather innovative form, and *ta* for neuter. (Note that the case distinction nominative vs. oblique case only applies to plural definite articles and does not reflect in the nominal endings anymore). Furthermore, in the feminine nominative plural, the neuter plural article *ta* has spread via the plural of inanimates > animates/non-humans > human nouns and is now predominant also for human fem.pl. nouns, e.g., *ta patsiðaes* ‘the girls’, *ta nifaðe* ‘the daughters-in-law’, although it may exist alongside the inherited plural in *i*, e.g., *i inetš* ‘the women’ (but also *ta inetšes*). In fact, thinking ahead based on the present situation in the Romeyka corpus, it is likely that there will be in the plural only a two-way system between a nominative plural article for all nouns, *t(a)*, and a syncretic form *tsi/ti* (or other, subject to strong micro-dialectal variation) for ACC/GEN case, though at present – and probably subject to considerable diatopic variation, also related to the degree of language shift/attrition in the respective variety – some complications remain in the nominative plural of masculines/feminines through the use of *i* and possibly also in the accusative plural of masculines/feminines through the distinction between *ti* and *tsi*.

**Table 11:** Definite articles in Romeyka

		SG	PL
M	NOM	<i>o</i>	<i>i, (t(a))</i>
	ACC	<i>to(n)</i>	<i>tsi, (ti)</i> <sup>61</sup>
	GEN	<i>tu</i> <sup>62</sup>	<i>tsi</i>
F	NOM	<i>i</i>	<i>t(a), i</i>
	ACC	<i>ti(n)</i>	<i>tī, (tsi)</i>
	GEN	<i>tsi</i>	<i>tsi</i>
N	NOM	<i>t(o)</i>	<i>t(a)</i>
	ACC	<i>t(o)</i>	<i>t(a)</i>
	GEN	<i>tu</i> <sup>63</sup>	<i>t(a)</i>

- (153) *du barxari da tšeire*  
 ‘the meadows of the pasture’ (08\_04072019M\_3; 092)
- (154) *mi tsi aulus udž epikan=a*  
 ‘We did not do it with the men.’ (01\_06042017F\_4; 068)
- (155) *ta za to enan etšilizane*  
 ‘One of the animals stumbled.’ (08\_04072019M\_1; 110)
- (156) *si nife-s=muna andraðelfo*  
 ‘the brother-in-law of our bride’ (07\_04072019F\_5; 12)
- (157) *dema ti emsale ulinus endyona*  
 ‘I beat all my peers.’ (02\_21042018M\_2; 16)

Turkish loan nouns usually receive an article which may be deleted for phonological reasons before a vowel (Section 2.3.5.3). According to Neocleous (2020: 38), the definite article can also be deleted before loan nouns beginning with a consonant, which is, however, not observed in the present corpus.<sup>64</sup> Integrated Turkish loan nouns are predominantly assigned neuter gender, e.g., *do dolab* ‘wardrobe’, *do džop* ‘waste’, *do gemlegin* ‘blouse’, *do düyun* ‘wedding’. Inanimate nouns ending in a vowel usually receive feminine gender, e.g., *i pendžere* ‘window’, *i tšikolata* ‘chocolate’, *(i) araba* ‘car’ (Mackridge 1987: 127). Masculine gender is only assigned to semantically clear masculine entities, e.g., *o avdži* ‘hunter’; see also Sections 4.1.4 and 4.2.1.1.

<sup>61</sup> Also *tus* retaining an AG form in ROF as spoken in Saráchos (Uzungöl), but which is not attested in the Romeyka corpus (other masc.acc.pl. forms of the definite article are *tsi/tsu/ti/tin/tis*, Mackridge 1987: 132).

<sup>62</sup> *Tu* form for genitive masculine and neuter only in ROF as spoken in Saráchos (Uzungöl) (Mackridge 1987) and RSür (Dawkins 1937), other varieties use *tsi*.

<sup>63</sup> See Fn. 62 just above (also cf. Neocleous 2020: 38).

<sup>64</sup> Considering ex. (27) in Neocleous (2020: 39), i.e., *ø fotodžis* ‘cameraman.NOM’, the optional (and originally phonologically determined) omission of definite articles may be limited to the nominative of human (and animate?) nouns. However, this is countered in the present corpus by several nouns denoting humans in the nominative (that even start in a vowel and receive the definite article anyway), e.g., *o avdži* ‘hunter’, *o öretmenis* ‘teacher’.

Note that in few nouns of the corpus, the original determiner is re-analysed as part of the noun and an additional neuter article is used, e.g., *do ofengos* ‘the moon’ (04\_01072019F\_12; 39).

As stated above, the definite determiners in Romeyka, (especially if singular), generally agree with their head noun in number, gender, and case. However, the spread of the neuter plural declension to [–HUM] feminine and inanimate masculine nouns (see also Section 4.2.1.1), and the spread of neuter adjective inflection to adjectives qualifying a [–HUM] feminine noun (Section 3.1.2.1), contribute towards a mixed system (Mackridge 1987: 128). Michelioudakis & Sitaridou (2013: 366, Fn. 3) note that in Romeyka, we find both an inherited grammatical gender agreement system, and a semantic agreement system, the distribution of the two being conditioned by properties of the head such as animacy and gender. Following Karatsareas (2011), they note that position on the Animacy Hierarchy is relevant, with humans behaving distinctly from non-human nouns (see also Karatsareas 2014). In example (158), we witness the initial article *to* indicating semantic agreement (if neuter is considered to reflect non-humanness), while the article immediately adjacent to the head exhibits syntactic (feminine) agreement. This would also be in line with Corbett’s (2006: 234) prediction that with stacked agreement targets, as in (158), the target more distant from the controller will show semantic agreement, if any target does (see also Karatsareas 2014). Singular definite articles immediately preceding their heads always exhibit syntactic agreement in Romeyka (Michelioudakis & Sitaridou 2013: 366, Fn. 3). But in the present Romeyka corpus, the shift of masc./fem. singular nouns to neuter declension occasionally also affects human heads, e.g., apparently even human proper names (159),<sup>65</sup> especially in translation data. Note also that in South Cappadocian, the shift to the neuter definite article in the singular of all genders is completed (Janse 2004: 12).

- (158) *to kokinon i kosara*  
 DET.N red.N DET.F hen.F  
 ‘the red hen’ (Mackridge 1987: 128; glosses added)

- (159) *t=ali dio yarðelæ iše*  
 ‘Ali has two children.’ (C1)

As for the syntactic behaviour, the definite article occurs before nouns in all grammatical roles, including proper names, and modifiers like adjectives and numerals.

In general, the masculine and feminine nominative singular definite articles may be dropped for phonological reasons before word-initial vowels to prevent a hiatus (160),<sup>66</sup> this is particularly striking in proper names (161; cf. Section 2.3.5.3): out of 163 proper names in the corpus, 105 have an article. But note that the article occurs in some examples even when a hiatus is the result, e.g., *i aiše*, *i almanja*, but tends to be frequently omitted before *emine* (see also 162). This could be fostered by the otherwise resulting hiatus of two high vowels. While the definite article is predominantly dropped in clause-initial position, it may also be left out elsewhere (163, 164). When the noun is introduced with an article in one clause, the article tends to be omitted in the following clause (165, 166). Note that the observation that masc./fem. definite articles are frequently deleted in Cappadocian (i.e., animate masculine and all feminine nouns lack the definite article in the nominative, Janse 2004: 12), not only [ \_V ] but even in definite contexts (Karatsareas 2013: 193), leads Karatsareas (2013) to argue for a contact-explanation being only the last step of a row of inherited changes in AMG, while Janse (2004)

<sup>65</sup> A different analysis of the possessive construction in ex. (159) would be a calque of the Turkish *var* existential with the noun in the genitive + verb-final ‘have’. In this case, *t* is the reduced masc./neut.gen. article *tu*.

<sup>66</sup> This probably extends to the masc./fem. plural definite article, although no examples are attested for plural.

assumes a relation to indefiniteness marking. At least in Turkish loan nouns, the occasional omission of definite articles (in definite contexts) might be attributed to a contact-explanation (167; also cf. Neocleous 2020: 38).

- (160) *andras=im sarija malia eš*  
‘My husband has blond hair.’ (A1)
- (161) *ahmetis n=arde*  
‘Ahmet came.’ (A1)
- (162) [...] *emigas bola xastas edune o vahides*  
‘[...] [T]he uncle was very sick. – Vahid.’ (04\_01072019F\_5; 25)
- (163) *patanu-dena en alai bejis*  
‘He is from Patanu. – Alay Bey.’ (01\_04022016F\_1; 032)
- (164) *tin aiše ahmetis epori tše elep*  
‘Ahmet can see Ayşe.’ (A1)
- (165) *o öretmenis pudžekah en öretmenis so kul en*  
‘Where is the teacher? The teacher is at school.’ (A1)
- (166) *i gülsüme gülsüme erθe son barxari*  
‘Gülsüme? She went to the pasture.’ (04\_01072019F\_17; 32)
- (167) *dolabi taxemeno dune*  
‘The cupboard was closed.’ (04\_01072019F\_5; 62)

Furthermore, the occurrence of the definite article is related to definiteness (Section 4.2.3, see also subsequent Section 3.2.1.2). While definite articles are primordially used with genuine definite reference, they do not occur in most other reference types (similar to SMG), such as generics and indefinite specifics (see also 168–171 for some examples).

- (168) *lebe eliyam=a*  
‘We called it lep.’ (04\_01072019F\_2; 018)
- (169) *atlama epikane saya epigane*  
‘They did the Atlama [dance], they did the Saya [dance].’ (01\_06042017F\_4; 011)
- (170) *etšine šexrije ekuje*  
‘She was called Şehriye.’ (01\_04022016F\_1; 049)
- (171) *kadaxor ibe*  
‘She said Kataxor [Çaykara].’ (06\_03072019M\_2; 60)

Concerning the use of definite articles with nominal modifiers, determiner spreading (cf. Alexiadou 2014; Karatsareas & Lekakou 2016) is an important particularity of Romeyka: Attributively used adjectives preceding the noun require their own definite article, so the NP contains two articles: before the adjective and the noun (172, 173; Mackridge 1987). The article of the nominal head agrees fully with the noun, while the modifier and its article are with feminine nouns often in neuter gender (more or less sensitive to animacy, see above and Section



3.1.21). For (many dialects of) Cappadocian, Karatsareas & Lekakou (2016: 197) suggest that determiner spreading is morphosyntactic agreement in terms of definiteness: definite masculine singular nouns lack determiner spreading, e.g., *o kalos anθropos* ‘the good man’ (see also Janse 2004, 2019: 100–101; cf. the Romeyka neuter form in 174).

(172) *hatšino d=omorfo don dobo etšiega krati do kliði=θe*  
 ‘The key of the nice place is left there.’ lit. ‘This nice place has left its key there.’  
 (08\_04072019M\_3; 161)

(173) *olon da tehlikeliĵa da dobe ta rðome*  
 ‘the most dangerous places, roads’ (08\_04072019M\_2; 091)

(174) *ja de d=omorfo ospidi*  
 ‘Look, what a nice house!’ (04\_01072019F\_12; 05)

If used as modifier, numerals also get an article (175, 176). This does not extend to quantifiers, though (see Section 3.2.3).

(175) *mo ta dðrt tane ta za*  
 ‘with four animals’ (04\_01072019F\_2; 087)

(176) *erθen me ta ðio patsið=atjes*  
 ‘She came with her two daughters.’ (07\_04072019F\_5; 04)

The possessive pronouns, which can be considered possessive adjectives, diachronically contain the neut.sg.nom./acc. definite article plus the pronoun (Drettas 1997: 148), e.g., *t=emo* ‘my’, *t=emetero* ‘our’ (177), the retention of which is an archaism. The merger of definite articles with modifying pronouns happened at an earlier stage of the language; therefore, the article is simply treated as part of the pronoun here. Note also that the article is not visible anymore when merged with a preposition (178; see Section 3.2.5).

The third person feminine possessive pronoun *t=etšine* is based on a demonstrative pronoun. Definite articles are thus different from demonstratives (see Dryer 2013b) with whom they co-occur when used as modifier (179, but cf. 180). This extends also to the borrowed Turkish reflexive pronoun *kendi* (181).

(177) *tadine t=ito serinluxi haðaðžega bere*  
 ‘Their cooling reaches till here.’ (lit. takes till here’) (08\_04072019M\_3; 140)

(178) *as=emetero din džami buga*  
 ‘below our mosque’ (08\_04072019M\_3; 158)

(179) *i edžin epije.. tetšines d=anepse n=epije*  
 ‘She went; her niece would have gone.’ (03\_30062019F\_6; 05)

(180) *edžin do yarðeli xasdas edun*  
 ‘Her child was sick.’ (08\_04072019M\_3; 042)

(181) *to kendi ti ðulias to pos efded edži*  
 ‘Why did you not do your own work there?’ (04\_01072019F\_2; 211)

## 3.2.1.2 Indefinite articles

The indefinite article has the form of the numeral ‘one’ *ena* and appears in the same position of the NP as the definite article. It exists in the corpus in two variants *ena* and *enan*. It is unclear, though, what determines the occurrence of both forms. While *enan* is the neuter nominative or singular accusative, *ena* may go back to the same form (as is noted by Özkan, n.d.). The variation might be phonologically conditioned [\_V/C], but the examples are contradictory (182–188). It is also unclear, whether the masculine nominative form *enas* (also *inas*, *is*, see also Section 3.1.3) is still used in Romeyka – at least as indefinite article: Tursun (2019: 249) notes it in few examples, e.g., *o enas entonen ton alon* ‘one hits the other’, while in the Romeyka corpus also the neuter<sup>67</sup> form is used, e.g., *d=enan d=alo* ‘each other, lit. ‘the one the other’ (01\_15022015F\_1; 22).<sup>67</sup> Note, however, that in these cases *enas/enan* is not the indefinite article but the numeral used pronominally. In any case, the feminine form *mia* is in Romeyka only used as temporal adverbial in the meaning of ‘once’, as Tursun (2019: 362) rightly states, e.g., *alo mia* ‘once more’. Feminine and masculine nouns also take the indefinite article *ena(n)*, e.g., *ena inega* ‘a wife’ (04\_01072019F\_12; 04), *ena(n) imera* ‘one day’, *ena arko* ‘a bear’ (B1), *ena auro* ‘a man’ (C1). It can thus be assumed that the indefinite article shows neither gender nor case agreement anymore, at least as far as nominative and accusative are concerned; for other case forms, further data are required.<sup>68</sup>

- (182) *as ixa nan bats=obon omo hatena hajes treš*  
 ‘If I (only) had a little girl like her who runs like this.’ (03\_07072019F\_1; 14)
- (183) *ađaha elebo ena ospit*  
 ‘Here, I see a house.’ (05\_03072019M\_3; 03)
- (184) *ešiega en enan denizi ađadžega ba exo eyo botamin*  
 ‘There is a sea, here I have a valley.’ (08\_04072019M\_1; 275)
- (185) *ton džiri=m eyrapsan enan xardin*  
 ‘She’ wrote my father a letter.’ (02\_21042018M\_2; 18)
- (186) *efidže so đormo ban ena ksilo*  
 ‘He lost a piece of wood on the path.’ (04\_01072019F\_13; 41)
- (187) *ebejn estađen s=enan apiđi*  
 ‘It [= a cow] went and stopped at a pear tree.’ (08\_04072019M\_2; 109)
- (188) *evalam=a s=ena betra bugo*  
 ‘We put it under a stone.’ (03\_30062019F\_1; 21)

In modified indefinite NPs, determiner spreading does not apply, the indefinite article occurs only once before the modifier and marks the whole phrase as indefinite (189–192). The adjective is in principle likely to agree in gender and case with both the head and the indefinite article, which is, however, practically always the neuter case.

<sup>67</sup> Mackridge (1987: 125) states that the indefinite article *ena* retains in Ophitic the AG masculine declension of feminine nouns, i.e., the feminine form takes the masculine endings.

<sup>68</sup> No genitive form is attested in the available data.

- (189) *eyo ba exo ena mikro pats-obo*  
 I TOP have.1SG INDET small girl-DIM  
 ‘I also have a little daughter.’ (07\_04072019F\_5; 33)

- (190) *ena omorfo ospid-obo en etšis*  
 ‘There is a beautiful little house.’ (08\_04072019M\_3; 146)

- (191) *o džobanis ta ketšia ena zengin arđobo bulisen*  
 ‘The shepherd sold the goats to a rich man.’ (B1)

- (192) *as ena alo pendžere*  
 ‘from another window’ (05\_03072019M\_3; 09)

Unmodified indefinite NPs usually lack a determiner (193–195). In this case, the noun is likely to be non-specific. Specific and referential singular indefinite NPs are usually marked by the indefinite article (cf. 182–186 above). For a discussion of indefiniteness see Section 4.2.3.

- (193) *eš ađelfo*  
 ‘Does he have a brother?’ (A1)

- (194) *eš zo*  
 ‘Does he have a cow?’ (04\_01072019F\_5; 29)

- (195) *ots erde o musafiris eš kliđi*  
 ‘Whoever comes as your guest has a key.’ (08\_04072019M\_3; 155–156)

The indefinite article functions also as indefinite pronoun (196). It can occur with proper nouns, if referring to an identifiable referent (197), although such identifiable and referential referents can apparently also appear without indefinite article (198), in this case, pronominal *ido* substitutes for that (for the indefinite pronoun *itos*, see Section 3.2.2.7).

- (196) *enan ebire efaen*  
 ‘He took one and ate it.’ (05\_03072019M\_4; 26)

- (197) *ena šerije ksero tše do so katoxor kahuton edži*  
 ‘I know a Şehriye, she lives in Çaykara.’ (01\_04022016F\_1; 053)

- (198) *si yoryora ido zias en*  
 ‘At Gorgoras, there is a certain Ziya.’ (01\_04022016F\_1; 018)

### 3.2.2 Pronouns

Pronouns can fill the position of an NP (e.g., as pronominal subject/object) and full pronouns have mostly the same distributional properties as the NP (Payne 1997: 42) while anaphoric clitics like “weak” possessive and object pronouns differ in their distribution. Phonologically, full pronouns can carry primary stress. This Section covers all “pronominal” forms including full pronouns, clitic pronouns, as well as other pronominal-type items, such as (borrowed) reflexives, relativizers, and interrogatives. Many of the latter have an uncertain status as pronouns or may be only derived from pronouns. In this Section, thus, pronouns are defined in their broadest sense as elements that can fill the position of a NP and/or are referential to an individual semantic entity. Namely, only part of the Romeyka relativizers and interrogatives

are pronouns but for the sake of clarity all elements in the respective function are described together in this Section.

### 3.2.2.1 Personal pronouns

Romeyka is a pro-drop language; verbal subject person marking is obligatory (see also Section 5.2.1.1), while (subject) person marking using personal pronouns is driven by discourse patterns including emphasis and contrast (ex. 199). Full personal pronouns vary according to person, number and gender. Gender distinction exists only for third person singular (see Greenberg’s 1963 Universal 45 cited in Siewierska 2013). This threefold gender distinction is the same as in SMG and differs strongly from Turkish, which has no gender. Personal pronouns exist for the grammatical relations subject and object. (In)direct object pronouns have full and clitic forms. As indicated in Section 1.2.4, the (en)clitic object and possessive pronouns are the only elements considered enclitic in this thesis although this decision is certainly to some degree arbitrary. As full (or strong) object pronouns are in the present thesis those pronouns considered that (i) show the long form considered archaic in comparison to pronominal forms in other modern Greek dialects, that (ii) are prosodically to a certain degree independent, and that (iii) may occur freely in the clause, i.e., in both pre- and post-verbal position. As clitic (or weak) object pronouns are those pronominal elements defined that occur solely in immediate post-verbal position and are prosodically difficult to differentiate from the verb.<sup>69</sup> The discussion of whether the weak object pronouns are clitics or suffixes shall not be dealt with in detail here since a clitic/affix distinction is often notoriously difficult to maintain, and the forms in question may exhibit properties of both affixes, and of clitics. Michelioudakis & Sitaridou (2012) do not consider the weak pronouns as clitics except for the 3<sup>rd</sup> person weak object pronoun *-æ* which is according to them the only true object clitic (also Neocleous 2020: 120). In the present thesis, the weak object pronouns are as a preliminary decision preceding further scrutiny of the particular properties of suffixes and clitics considered clitics; their properties are the following: (a) they do not affect the word accent of the verb (see “columnar stress” in PG which does not depend on the penultimate or tree-syllable rule otherwise effective in Greek, Mackridge 1987), (b) they are not mobile and no other element can occur between the verb and the pronominal element, except for other object clitics, (c) they are at the end of a long grammaticalization line which would require a separate analysis. The Romeyka corpus shows 600 tokens for subject and object personal pronouns (full and enclitic; with around 2800 finite verb forms), see Table 12. The distribution of full and clitic pronouns in Romeyka is an interesting topic that is likely to add an important aspect to the word order properties of the language. In addition, there are not few full object pronouns that occur post-verbally indicating an ongoing change – a detailed analysis of the (object) pronominal system remains a desideratum.

**Table 12:** Pronouns in subject and object function in the corpus

	Subject	Object	Totals pronouns
Full pronoun	269	126	395
Clitic pronoun	0	217	217
Totals	269	343	612

- (199) *ipa jox esi hado katuresis*  
 ‘I said, “No, you have peed this!”.’ (02\_9062019F\_2; 29)

<sup>69</sup> This definition of full (strong) and clitic (weak) object pronouns applies also to full and clitic possessive pronouns, with the only difference that clitic possessives can also attach to adverbs (see Section 3.2.2.3).

Romeyka personal pronouns are presented in Table 13. Note that there is a considerable amount of micro- (and other) variation in the form of the pronouns; the ones presented in Table 13 stem from ROF. Some variation occurring in the corpus data is depicted in the examples below. Furthermore, it remains a topic for future investigation, what determines the high amount of variation within the individual forms of the pronominal system, especially regarding the selection of full (=long) pronominal forms versus the shorter forms of the same pronoun. This variation seems to be partly dependent upon the diatopic variety; furthermore, it is in part phonologically conditioned by preceding or following phonemes (vowel vs. consonant); other factors might be sentence stress and position in the clause (see also discussion of the individual forms below).

**Table 13:** Subject and object pronouns in Romeyka

	Subject pronoun	Full object pronoun	Clitic object pronoun
1SG	<i>eŷo</i>	<i>emen(a(n))</i>	<i>=me</i>
2SG	<i>esi</i>	<i>esena(n)</i>	<i>=se</i>
3SG	<i>atos.M</i> <sup>70</sup> ; <i>ate.F</i> ; <i>ato.N</i>	<i>(a)to(n(a(n)).M/N</i> ; <sup>71</sup> <i>ate(n(a)).F</i> <sup>72</sup>	<i>=((a)n)a (/=ae)</i> <sup>73</sup>
1PL	<i>(e)mis(t(in(e)))</i> <sup>74</sup>	<i>emas(una)</i>	<i>=mas</i> <sup>75</sup>
2PL	<i>esis(t(in(e)))</i>	<i>esas(un(a))</i>	<i>=sas(en)</i>
3PL	<i>atin(e/i/a)</i> <sup>76</sup>	<i>atinus(a), atonusine, etunesine</i> <sup>77</sup>	<i>=(a)ta</i>

In the following, firstly some particularities of the subject pronouns are described; secondly, some of the morphological and syntactic particularities of the object pronouns.

<sup>70</sup> The demonstrative pronoun *etšinos* is also used as gender-sensitive third person pronoun in the singular: *etšinos.M*, *etšine(na).F*, *etšino.N*, and as invariant *etšine* in the plural. Note that it is not very likely that the form *etšine* is used for both 3<sup>rd</sup> person (feminine) singular and plural, rather the form *etšin(i)* would be likely for 3<sup>rd</sup> person plural; possibly, the forms simply came in the present corpus to coincide due to a similar pronunciation – this needs to be checked with further data.

<sup>71</sup> Like with the subject personal pronouns, a second set of third person object pronouns based on the demonstrative *etšinos* is used with the forms *etšino.M/N*, *etšinena.F*.

<sup>72</sup> *Ati(n(a))* according to Michelioudakis & Sitaridou (2012: 219).

<sup>73</sup> It is unclear whether the 3SG object clitic can be also used with 3PL referents. This seems to be suggested by Neocleous (2020: 120; see also Section 5.2.1.3) and some data in the present corpus seem to support this (i).

(i) *fasulijas bote en erxume bero=na* [...]

‘When there are beans, I come and take them [...].’ (03\_30062019F\_11; 041)

<sup>74</sup> Also, *imist(ine)*, e.g., (03\_07072019F\_1; 09).

<sup>75</sup> While the short form of the object clitic is certainly *=mas*, the (probable) full pronominal form *masin(e)* occurs only in immediate post-verbal position so that it is not clear, whether this is still the full pronoun (although not emphatic) or whether it is cliticizing with the verb (see also the description of individual forms below; Özkan (2013: 143) reports the form *masuna* of the postposed object pronoun).

<sup>76</sup> Table 9 in Neocleous (2020: 40) seems to indicate that the third person plural pronoun is still sensitive to gender with the forms *atini.M.PL*, *atine.F.PL*, and *atinæ.N.PL* (but note the different accent here). Although in the present Romeyka corpus the form *adine* appropriates a masculine noun and *adina* neuter referents, the far most frequent form is *adin*, which does not show any information on gender. Forms like *hadini(n)* that occur in translation data cannot be clearly attributed to a gender either, since by means of the translation task from Turkish (which has no gender distinction) the gender cannot be clearly identified. But note that Drettas (1997: 180) only lists for PG masc./fem. *atin* and neuter *ata*. In addition, invariable *etšine* is used.

<sup>77</sup> For third person plural, again a double set of object pronouns exist with the following demonstrative forms *etšin(a/i)*, *etšinus*, it is not clear whether in the endings still gender-sensitivity is reflected. In general, a plethora of forms of both sets exist, see description of individual forms in Section 3.2.2.1.2.

#### 3.2.2.1.1 Subject personal pronouns

While first and second person singular subject pronouns only have a single form (but cf. 200, 201), a double set of third person singular subject pronouns exists. In addition to the personal pronouns *ate* for feminine, *atos* for masculine, and *ato* for neuter, a second set of third person pronouns exist for all three genders: *etšinos* ‘this one’ is the demonstrative pronoun with the feminine form *etšine.F/etšinena.F* (202, 203), *etšinos/atšinos* for masculine (204), and *etšino/adžino.N* (205) for neuter. While the use of demonstrative pronouns for third person personal pronouns is typologically widespread (Bhat 2013), it is not fully clear why there is a double set of third person pronouns and what conditions their selection. Tentatively, *ato* is frequently used as object pronoun, also with a demonstrative reading, and referring to inanimate objects (cf. ex. 206); possibly, the form *etšinos* fills a gap here. Apart from that, demonstrative *etšinos* is more emphatic than the personal pronoun *atos*. In comparison with clearly demonstrative uses of both pronouns (Section 3.2.2.2), Drettas (1997: 179) considers *atos* to display greater proximity (in discourse) than *etšinos*. Thus, deixis might be the decisive factor in the distribution of 3<sup>rd</sup> person subject personal pronouns: *atós* (proximal) ~ *(h)avútos* (distal) ~ *etšinos* (distal).

- (200) *resmije tš=esi bola arkadaši estini*  
'You and Resmiye were close friends.' (03\_07072019F\_1; 16)
- (201) *eritšes es*  
understand.PRS.2SG you  
'Did you understand?' (06\_03072019M\_2; 43)
- (202) *xameleda en etšine har u=fanerui ki*  
'It is a mill, it cannot be seen now.' (07\_04072019F\_6; 02)
- (203) *etšinina ba utš erte*  
'She (also) did not come.' (04\_01072019F\_17; 66)
- (204) *atšinos ba le ah erte argos opsari do fanimo*  
'He said, "ah, the bear came to eat the fish".' (04\_01072019F\_12; 44)
- (205) *ištera atšino manaxo abome*  
'Later, it [= the child] remains alone.' (01\_04022016F\_1; 089)
- (206) *ado utš eksero*  
'I don't know it.' (04\_01072019F\_1; 078)

Both sets of third person pronouns can occur with the prefix *h-* that fosters a clearly demonstrative reading (cf. the use of *h-* in special adverbials in Section 3.1.4.4), e.g., *h-ade.F* (207), *h-atšino.N* (see ex. 172 this chapter above). In (208), *hado* occurs as a generalized gender-neutral form likely modelled on the (gender-neutral) Turkish demonstrative *bu*, TT *habu* ‘this’.

- (207) *hade n=eftej=eme jadimi*  
 ‘This one (=she) will help me.’ (02\_2906019F\_1; 19)
- (208) *hado en jabandži u=boro na kroy=a*  
 ‘She is a foreigner, I cannot blame her.’, lit. ‘I cannot hit her.’ (01\_28062019F\_2; 30)

The form of the first person plural personal pronoun varies between the short form *mis* (209) and the full form *emistine* (210). The use of the full or short form of the pronoun is subject to dialectal variation whereby the villages farther from the sea tend to use the full form (e.g., Ogene). Whether the syllable margins are reduced is phonologically conditioned: *mis* [V\_; \_V], *emis* [\_V], *emist* [\_C]; *emistin* [\_V], *emistine* [\_C]. The second person plural pronoun has also a short (*esis(t)*) and a long form (*esistine*) (211), whose distribution likely follows the same principles. The third person plural pronouns show again a plethora of variants, e.g., *atin(e/a)* (212) vs. *etšin(i/e/a)/atšin* (213); a variant with prefixed *h-* occurs as well: *h-atin(i)* (214). Note that the endings of the third person plural pronoun in /a, e, i/ might evidence gender-sensitivity (as suggested from Table 9 in Neocleous 2020: 40), which could, however, not be finally proved by the existent data set (see also Fn. 76 above).

- (209) *mis aiše leyum=ana*  
‘We call her Ayşe.’ (01\_04022016F\_1; 011)
- (210) *emistine dormo utš ixame*  
‘We did not have a road.’ (08\_04072019M\_1; 061–062)
- (211) *esisdine mualimi istine*  
‘You are teachers.’ (C1)
- (212) *adina so xorio ine so xorio biyane*  
‘They are at the village, they went to the village.’ (03\_30062019F\_6; 18)
- (213) *etšin dona binane*  
‘What did they do?’ (04\_01072019F\_17; 20)
- (214) *hatini tšimundan*  
‘They slept.’ (04\_01072019F\_12; 25)

### 3.2.2.1.2 Object personal pronouns

Romeyka has both full (“strong”) and clitic (“weak” = unaccented) object personal pronouns, the latter being enclitically attached to the verb (for a general statement on the term “clitic”, see Section 3.2.2.1). Enclitic pronouns only occur in non-nominative functions. Although a detailed analysis of cliticness of Romeyka pronouns is still pending, the word order of full object pronouns is (like with full object NPs, Section 5.2.1) variable, they can follow the verb and other elements (Michelioudakis & Sitaridou 2012: 219). The linguistic status of full object pronouns in post-verbal position is not entirely clear, namely whether they should be still considered full pronouns or whether they are on a way to cliticization.

Unlike in other Greek varieties, in Romeyka both full and clitic object pronouns seem to be an unmarked option, i.e., full pronouns seem not to be linked to emphasis or focus (Michelioudakis & Sitaridou 2012: 219). As for their distribution in the Romeyka corpus, 32% of the object pronouns are full pronouns;<sup>78</sup> the majority of pronouns is thus clitic. Still, if one compares this to the Multi-CAST corpus of Cypriot Greek (Hadjidas & Vollmer 2015) where a vast majority of object pronouns is clitic, the amount of full pronominal forms is higher in Romeyka. It could be considered whether this variation in frequency might be related to contact with full pronouns in Turkish, although Cypriot Greek is also impacted by Turkish influence.

<sup>78</sup> Separated according to person and number: 37,5% of full 1SG pronouns, 35,7% of full 2SG pronouns, 41,9% of full 3SG pronouns, 34,6% of full 1PL pronouns, 30% of full 2PL pronouns, and 13,3% of full 3PL pronouns. Importantly, third person singular object pronouns tend to be more often full pronouns than other persons, and third person plural object pronouns are least likely to be a full pronoun.

But nevertheless, Cypriot Greek has remained in contact with SMG, which is completely lacking in Romeyka. Generally, a reduction in the use of clitic object pronouns appears to be a contact phenomenon of languages under Turkish influence. Still, it is not clear, what determines the distribution of full vs. clitic object pronouns in the present corpus although the enclitic object pronominal form could be considered the default form. More detailed research on the pronominal system of Romeyka is needed.

Object clitics occur not only on verbs but also on spatial adverbs, e.g., *apes=am* ‘inside me’ (03\_30092019F\_7; 35). Clitic clusters and the order of object clitics is discussed in Section 5.2.1.4. Like in Christian PG, (object) clitic/pronoun doubling by means of a resumptive pronoun occurs in Romeyka (cf. Janse 2008 on Cappadocian), although not obligatorily and, as it seems, merely for pragmatic purposes. For a more detailed description of this phenomenon see also Section 5.2.1.4.

In the following, variation in the Romeyka object pronouns is presented in more detail.

The first person singular full object pronoun is *emen(a(n))* (< AG *emenane*, as still reported by Neocleous 2020: 40, Table 9); the clitic is *=me*. In a few cases, speaker confuse the first person object pronoun with the possessive pronoun (215, also 06\_03072019M\_2; 54). The second person singular object pronoun is shown in (216, 217). It also occurs with deletion of the unstressed initial vowel, i.e., *sena* (01\_04022016F\_1; 098). Michelioudakis & Sitaridou (2012: 219) note that the syntactic status especially of first- and second-person object pronouns is unclear as they cannot easily occur in a nonadverbial position.

For the third person singular free object pronoun again a double set exists: (i) there is considerable variation in the personal object pronoun *(a)to(n(a(n)))* (218) used for both masculine and neuter objects, but the shortest form *to* (219) is only attested for neuter; the feminine form is *atena(n)/(a)dina* (220); (ii) the demonstrative-based object pronoun *etši(n(o))* occurs for masculine and neuter (221) and *etšinina* for feminine. Third person object pronouns with emphatic *h-* (also as demonstrative form *h(a)(vu)-*, ex. 222) exist, e.g., *h-atenan* (02\_2906019F\_1; 20). It seems that the personal pronoun *ato(s)/ate* covers all genders while demonstrative-based *etšino* is only used for inanimate neuter referents. Note that the reduced form of the masculine/neuter object pronoun *to* appears only post-verbally and tends to cliticize with the verb (other than found by Michelioudakis & Sitaridou 2012: 219). This suggests a split arising for the neuter object pronoun whereby *etšino* is the full pronoun and *(a)do* tends to become a clitic (see above).

- (215) *emena to xoma exoriksen emuna met etšino*  
 ‘My<sup>2</sup> land spared me.’ (translation unclear; 01\_04022016F\_1; 002–003)
- (216) *esena u=dušünefkume sena dibu ne baθane [...]*  
 ‘I don’t think of you, if something happens to you [...].’ (01\_04022016F\_1; 098)
- (217) *eši kalo mualimina omon esenan*  
 ‘She has a good teacher like you.’ (01\_04022016F\_1; 077)
- (218) *eθekam atonan so varelin dže gremasam adona*  
 ‘We put it (=a baby bear) in a tub and hung it.’ (02\_29062019F\_2; 11–13)
- (219) *temizlejepsa do*  
 ‘I cleaned it.’ (04\_01072019F\_17; 09)
- (220) *hadenan bal eraksen ama kala epitšen adena*  
 ‘It bit her, and it bit her hard’, lit. ‘but it made her well’ (02\_2906019F\_1; 20)



- (221) *pan med edžino glošgeze obize*  
 ‘You go with it (=money), you return [with it].’ (08\_04072019M\_1; 318)

- (222) *havu-dina erodas d=edžer emen anda utš inanevis*  
 ‘You (can) ask her, if you don’t believe me.’ (03\_30062019F\_8; 09)

The third person singular object clitic =(a)n(a) seems to be the only third-person form with neutralized gender (ex. 223, 224, 225). In the data of Michelioudakis & Sitaridou (2012: 219), the form of this clitic is -æ, gender neutral and possibly also number-neutral. Özkan (2013: 146) distinguishes in RSür between =(a)n)a.M/N and =(a)n)æ.F. A singular occurrence features the clitic =ta for a feminine referent (226) and the possessive =at is used in (227) in a locational adverbial. The motivation of the distribution of the long (=ana) vs. the short clitic form (=a) is unclear and requires further research (228; cf. the explanation by Özkan 2013: 146, i.e., /a/ / [an\_]).

- (223) *xandilaza=na*  
 ‘I tickle her.’ (02\_02022015F\_1; 144)

- (224) *anne as eberum hado ton arkon as eberum, ebiram=an*  
 ‘Mother, let us take this bear, let us take it! We took it.’ (02\_2906019F\_1; 05)

- (225) *efaan=ada džirothane=ana dibu tš=ebomene sin gofa*  
 ‘They ate them, they finished it, nothing remained at the bucket.’ (04\_01072019F\_12; 55)

- (226) *na karibefke=ta*  
 ‘She will miss her.’ (01\_06042017F\_4; 022)

- (227) *embros=at s=ena sergi eban opsaræ*  
 ‘In front of him on a blanket he has fish.’ (05\_03072019M\_3; 29)

- (228) *beker eθelesan=ana edera θelesan=a*  
 ‘Unmarried (men) wanted her, others wanted her.’ (01\_04022016F\_1; 106)

The first person plural full object pronoun is *emas(un(a))* (229); the clitic =*mas(e/u)/mis*. The longer form *masin(e)* (230) only occurs in immediate post-verbal position, so that its syntactic status as free pronoun is not certain anymore. The full pronoun can combine with prepositions, e.g., *ap=emas* ‘from us’, *s=ema(suna)* ‘to us’ (231). Like it has been noted for other persons, local prepositions occur in some instances with the possessive pronoun instead of the object pronoun (232); this might be a matter of language competence possibly indicating a change. Note that a less competent speaker also provides the possessive pronoun without preposition (233). Furthermore, subject personal pronouns are occasionally confused with object pronouns, e.g., *har esena tš=eyo na sindišenum* ‘now you and I will talk’ (04\_01072019F\_5; 06). While the confusion of personal pronouns with possessive pronouns in ‘have’-clauses is likely contact-induced by Turkish (Section 3.2.2.3), the uncertainty about the choice of subject versus object pronouns might be an instance of attrition.

- (229) *adin bal emasun eyabune*  
 ‘They also liked us.’ (09\_04072019\_7; 17)

- (230) *eliyane=masine apopse elate s=emena*  
 ‘They told us, “come to me tonight”.’ (04\_01072019F\_2; 186)
- (231) *da insane bal foyundan n=axundane s=emasuna*  
 ‘The people are afraid of coming to us.’ (02\_29062019F\_2; 37–38)
- (232) *adžin s=emetera erxusane*  
 ‘They came to us.’ (09\_04072019\_7; 19)
- (233) *pudžega na bas demedera na bas*  
 ‘Wherever you will go, you will go to us.’ (06\_03072019M\_2; 04)

The free form of the second person plural object pronoun is *esas(un(a))* (234); the form of the clitic =*sas(en)* (235); if combined with a preposition, e.g., *s=esa* ‘to you.PL’, *ap=esas* ‘of you.PL’ (236).

- (234) *ena hedije pal esasuna na điyo*  
 ‘I will also give a present to you.’ (A1)
- (235) *an θelide ebero=sas tše bao adutšam*  
 ‘If you like, I take you up there.’ (08\_04072019M\_4; 03–04)
- (236) *ena vradī erxumune s=esa*  
 ‘One evening, we (will) come to you.’ (04\_01072019F\_2; 310)

For third person plural full object pronouns, again a double set of pronouns exists with considerable variation: (i) *adin(us(e))/adonusine/etunesine* (237), also *atini/u*, *adindana* (238); (ii) *etši(ne)na* (239, 240, 241), *etšinin*, *etšinus/etšinós* (242). Although the variation in vowel endings resembles gender-sensitivity, no certain connection could be established with the existing examples although gender-sensitive forms would be not unlikely, at least for the demonstrative set; further research is required here. Emphatic forms with *h-* are *h-atin(in)*, *h-atšin*. As hapax legomena, *ada* occurs as free form (243); and together with a preposition probably demonstrative *s=avunes* (244). The clitic pronoun is =(a)*ta* (245), also =*nada* (246). Again, it is not clear what determines the distribution of the longer or shorter forms =(n(a)(ta))).

- (237) *ama utš eboresane dosin adonusine as derume na krui etunesine mi*  
 ‘But they could not shoot them. Let us see whether he will hit them.’  
 (04\_01072019F\_12; 15–16)
- (238) *adindana do kliđi ti fadime eđokane*  
 ‘They gave the key to Fatma.’ (B1)
- (239) *etšinena eynorizez me*  
 ‘Do you know them?’ (01\_04022016F\_1; 020)
- (240) *ebejenam tše erxumunestine š etšina yo ferina so xorio*  
 ‘We went and returned and I brought them to the village.’ (08\_04072019M\_2; 159)
- (241) *butiün jazun etšina ula ixa=ta eyo*  
 ‘The whole summer, I had them (=the cows) all.’ (04\_01072019F\_2; 139–140)

- (242) *etšies exume akrabađe na bam edžinus elebome mono tše borbadim tš=erxumis*  
 ‘We have relatives there, we go, see them, walk a little and return.’ (08\_04072019M\_3; 114–117)
- (243) *ada ba gelen sene na diyo*  
 ‘I will give them (away) next year.’ (03\_30062019F\_6; 59)
- (244) *ena vrađi pejnam s=avunes*  
 ‘One evening, we went to them.’ (04\_01072019F\_2; 311)
- (245) *evalam=ada apes*  
 ‘We took them inside.’ (03\_30062019F\_1; 14)
- (246) *epejnane eksu evraxustane erxusane eyvalina=nada dže*  
 ‘They went outside, got wet, came, and I undressed them.’ (04\_01072019F\_2; 116)

Further research is required on politeness in Romeyka. A single example in the corpus that could be analysed as showing politeness uses the second person plural pronoun to address a single person (247).

- (247) *ap esas pola memnun-ena en*  
 ‘She is very content with you.’ (01\_04022016F\_1; 068)

### 3.2.2.2 Demonstratives

Demonstrative pronouns in Romeyka show a plethora of forms. Basically, they go back to three pronouns: (i) the demonstrative pronoun *etšino(s)/etšine*, (ii) the third person personal pronoun *ato(s)/ate*; (iii) a third pronominal element carrying demonstrative meaning (*h*)*avu* + article, i.e., *avudos.M*, *avude.F*, *avudo.N* (Drettas 1997: 179); (iv) there is a pronoun *it(en)os*, whose analysis is not fully clear, but which is likely also to express a demonstrative meaning.

All pronominal forms can occur with the prefix *h(a)-* (in Neocleous 2020: 43 cited as *χατός* (> *h-atós*) and *χατšós* (probably > *h-atšinos*). In Turkish, *ha* is a deictic particle that exists in Turkish of the larger Anatolian area and that expresses emphatic stress, for example, for raising attention, e.g., *ha-bu* ‘this.EMPH’ (C. Bulut, p.c.). In Romeyka *ha-* is used in the same way as in Turkish, namely with personal and demonstrative pronouns as well as spatial adverbs, and possibly also with the manner adverb *hajes*, cf. TTr. *ha-bu* ‘this.EMPH’, R. *havudos* ‘this.EMPH’; TTr. *ha-oreya* ‘there.DAT.EMPH’, R. *hađatšes* ‘here.(side).EMPH’; TTr. *ha-boyle* ‘like this.EMPH’, R. *hajes* ‘like this’<sup>(?)</sup>. However, its grammatical status as prefix or particle is not clear. As it usually occurs at the beginning of pronouns and adverbs, it is considered a prefix here but note that it seems to occur also as a free element in (248). However, considering the suffixal element *-ha*, which also has a deictic emphatic function, e.g., *ađa-ha* ‘here.EMPH’, an analysis as affix or even as particle might be more convincing. Furthermore, it is not clear whether the vowel /a/ is part of the prefix since it only occurs [a]. Note that *ha(/he)* is also used as emphatic answer particle for ‘yes’ in the larger Anatolian area (including Western Armenian, A. Donabedian, p.c.).

- (248) *i mana=s ha elej*  
 ‘Your mother says this?’ (07\_04072019F\_5; 43)

Demonstrative pronouns seem not to express distance in first instance.<sup>79</sup> Instead, according to Drettas (1997: 179), the forms *atos/avudos* differ from *etšinos* in terms of proximity in discourse, i.e., *atos/avudos* precede *etšinos* in discourse, *avudos* is more proximal than *atos* (Drettas 1997: 180). However, this might often equal a distinction according to distance with *atos/avudos* referring to something more proximal than *etšinos* (see also van Emde Boas et al. 2019: 353 for AG; but cf. exs. 11, 12). Demonstrative pronouns occur both as modifier of a nominal head and as anaphoric pronouns. As anaphoric pronouns, they agree with their head in number, gender, and case, i.e., they occur as subject and object pronouns. If used as modifiers, however, it seems the neuter forms are selected as modifiers of nouns of other genders (possibly sensitive to animacy, see the discussion in Section 3.1.2.1); for examples see Neocleous (2020: 42–43, ex. 30, 34). If a demonstrative determines a noun phrase, the definite article is obligatory before the noun; in this case, the article shows full nominal agreement for number, gender, and case whereby the demonstrative inflects for gender and number (252, but cf. ex. 263 below), as do other modifiers (see Section 3.1.2). Due to this feature, modifying pronouns like demonstratives and possessives may be technically better called adjectives rather than pronouns. The three options to express demonstrative meaning in Romeyka are separately characterized below. For an overview of the inflectional forms of the three pronouns see Table 14 (cf. Drettas 1997: 180).

**Table 14:** Demonstrative pronouns in Romeyka

Sg.	Nom.	Masc.	<i>etšinos/atšinos</i>	<i>avudos</i>	<i>atos</i>
		Fem.	<i>esthine</i>	<i>avude</i>	<i>ate</i>
		Neut.	<i>etšino</i>	<i>avudo</i>	<i>ato</i>
	Acc.	Masc.	<i>etšino</i>	<i>avudo</i>	<i>adona(n)</i>
		Fem.	<i>etšinina</i>	<i>avudin, auden</i>	<i>adena(n)</i>
		Neut.	<i>etšino</i>	<i>avudo</i>	<i>adonan, adono</i>
Pl.	Nom.		<i>etšin(i, e, a)</i>	<i>avuda.N</i> <i>avudin.F</i> <sup>80</sup>	<i>ata</i>
	Acc.		<i>etšin-us/-os,</i> <i>etšin(in),</i> <i>etšinena</i>		<i>adonusine</i>

Firstly, the genuine demonstrative pronoun is *etšino*; see ex. (249) for neuter and (250) for feminine. It also exists with the demonstrative prefix as (*h-*) + *atšino*.<sup>81</sup> It is used as modifier and anaphoric pronoun (for the latter use, see Section 3.2.2.1). It agrees in number, gender, and case, although reduced endings and spread of neuter often not allow to identify case inflection.

(249) *met etšino to yanaxtija kal ipejname*  
‘With this fatigue we went again.’ (04\_01072019F\_1; 153)

(250) *atšin i batsi utš efitše=me*  
‘This girl bothered me.’, lit. ‘did not leave me’ (02\_2906019F\_1; 04)

Secondly, the 3<sup>rd</sup> person personal pronoun is used as a demonstrative pronoun, especially together with the prefix *h(a)-*: (*h-*)*ado.N* (as modifier in 251), *h-ado(s).M* (252), *h-ade.F* (anaphoric function in 253); and a plural in *h-ada*. If used as modifier, the neuter form (*h-*)*ado*

<sup>79</sup> But cf. the Cappadocian demonstratives: *atos* (proximity), *etšinos* (distal), *etuto* (far distance, out of sight) (M. Janse, p.c.).

<sup>80</sup> Masculine plural form not attested.

<sup>81</sup> Drettas (1997: 179, Fn.2) notes the occurrence of both forms of the demonstrative pronoun *etšinos* and *atšinos*.

seems to be used insensitive to gender, as Neocleous (2020: 43) suggests. As oblique, see *h-ade(na(n)).F.ACC* (see ex. 220 above) and a rare form *h-adono.N/M.ACC*<sup>2</sup> (in anaphoric function in 254).

- (251) *hado do musluxi aniðum=a*  
 ‘We opened this faucet.’ (08\_04072019M\_1; 079–080)
- (252) *anne as eberum hado ton arkon as eberum*  
 ‘Mother, let us take it, let us take this bear!’ (02\_2906019F\_1; 05)
- (253) *eyo jelo hade duru dže esta tše deri=me*  
 ‘I laugh, she stays and keeps looking at me.’ (01\_28062019F\_2; 27–28)
- (254) *hadono.ACC<sup>2</sup> nad leyume*  
 ‘Will we say this?’ (04\_01072019F\_13; 01)

A third type of demonstrative pronoun is *avudos*; often with the prefix *(h)a-*. This pronominal form goes probably back to AG *οὗτος* [hûtos] ‘this’ < MedGr *οὗτος* [ûtos] and can be etymologically analysed as *ha-udo(s)* with a hiatus filler /v/ or the velar fricative /ɣ/, > *ha-[v]udo(s)* (M. Janse, p.c.). Like this, only *-udo(s)* is of Greek origin; in this case, the regional deictic element must be *ha-*. Note that in Cappadocian this pronominal form is also used as manner adverb *ha-vud-dža* (cf. R. *hajes*). The demonstrative *habu* also exists in regional Trabzon Turkish (ex. 255), whereby it is not obvious in which ways the Turkish and the Romeyka form have influenced each other (Brendemoen 2002, see also discussion of invariant *(h)avu* at the end of this section). Özkan (n.d.) assumes that, what he calls the demonstrative pronoun *avudos* or *afto* ‘this’ (in the present corpus also *audu*), can only be used as anaphoric pronoun referring to a previously introduced antecedent, although deictic use is possible as well, (e.g., 256). Irrespective of its use as anaphoric pronoun or modifier, *avudos* inflects for number, gender and case, e.g., *(h-)avudo/audu.N*, *avude.F* (as modifier in 257), *auden.F.ACC* (anaphorically in 258), *avudin.F.PL* (anaphorically in 259) and syncretic *avudin.F.ACC* (as modifier 260), *(h-)(a)vuda/auda.PL* (anaphorically in 261, as modifier in 262). But note, that like with other modifiers, *avudos* may follow neuter declension while the noun is of feminine gender (263, but cf. 257). Like with the other demonstrative pronouns described above, if used as modifier, the definite article of the noun is obligatory, dependent upon the definiteness of the NP (cf. 258).

- (255) *jani habu mahallenun alişveriş işlerini ben görüjordum*  
 ‘Well, this neighbourhood’s shopping business I have overseen.’ (08\_04072019M\_2; 052–054)
- (256) *avudo to sağan epero.*  
 DEM the.ACC plate.ACC take.1SG  
 ‘I take this plate.’ (A1)
- (257) *eba avude i aişe*  
 ‘Listen, Ayşe...!’ (01\_04022016F\_1; 011)
- (258) *auden aişe leyume*  
 ‘We call her Ayşe.’ (01\_04022016F\_1; 068)

- (259) *avudin ul dinlejevune=me*  
 ‘They are all listening to me.’ (01\_15022015F\_1; 16)
- (260) *avudin batsi eši ðiyome so jan=is*  
 ‘This girl I am giving to you.’ (01\_02022015F\_1; 03)
- (261) *auda ndo ine auda ndo ine*  
 ‘What are these, what are these?’ (03\_30062019F\_11; 060–061)
- (262) *b=esane haoda da šimundos*  
 ‘Where have these rains been?’ (03\_30062019F\_6; 44)
- (263) *havudo din göli dibu dž en edži ama*  
 ‘But in this lake there is nothing.’ (04\_01072019F\_12; 26)

Note that in (264–266), an invariant demonstrative (*h*)*avu*, i.e., without pronominal ending, is used as modifier. The definite article of the noun occurs either alone (264, also, e.g., *havu i sevim* ‘Sevim’ (01\_28062019F\_3; 01)), is left out at all (265, as is often the case for the word *batsi* ‘girl’, though), or is preceded by a preposition (266). These examples give rise to the assumption of an invariant demonstrative modifier (*h*)*avu* as opposed to pronominal (*h*)*avudos* described above. Finally, the invariant demonstrative (*h*)*avu* also combines with spatial expressions like *-tšes*, e.g., *havu-tšes/djes* ‘this side’ (see Section 3.1.4.3.).

- (264) *havu o ðormos etši son barxari bai*  
 ‘This road there goes to the pasture.’ (08\_04072019M\_1; 210)
- (265) *havu batsi bal xastas edune*  
 ‘This girl was sick.’ (01\_28062019F\_2; 13)
- (266) *havu si musafiris ba aboliyume tšebedži na bame haðabahen n=efteme duvari*  
 ‘When we split from these guests, later, we will go up and build a wall.’  
 (08\_04072019M\_3; 022–025)

Similar to invariant (*h*)*avu*, in some cases, *ha* occurs directly together with the article (267) or the demonstrative can be reduced to (*h*-)*ao.M*, *oi.F* (268).

- (267) *ha i batsi bal rumdže utš ekse*  
 ‘This girl does not know Rumca.’ (01\_28062019F\_2; 25)
- (268) *oi patsi tšoh.. bola xastas en*  
 ‘This girl is very sick.’ (07\_04072019F\_5; 15)

Finally, an archaic pronoun is *it(en)os.M/itena.F/ito.N* (cf. *itina* in Sitaridou 2014b: 30). The etymology of this pronoun is not clear; M. Janse (p.c.) suggests a relation to the demonstrative pronoun *atos*, that expresses proximity. This pronoun inflects at least for gender (ex. 269 for masculine, 270 for neuter, feminine *idena* in (01\_04022016F\_1; 050)) and number (271). Note, however, that M. Janse (p.c.) suggests for the form *ita* an analysis not as plural but as distal demonstrative pronoun (see Fn. 79; but cf. Tursun 2019: 269 for an analysis as plural). Ex. (272) suggests that *it(o)* can receive a determiner and be thus treated as noun (also 271). Otherwise, it can occur in the NP together with a noun and following modifiers. In this case, it occurs in the position of the noun and is preceded by the definite article of the NP (273). This

also extends to indefinite NPs (274, 275). The pronoun does not always refer anaphorically to a referent but can substitute unnamed referents (276, also 277). It is frequently used as a filler in search for the right word (278, 279; also Tursun 2019: 269), which does not rule out its analysis as demonstrative pronoun. Unfortunately, the corpus does not show many examples for feminine *idena* but likely the form *edeni* (280) is related to it.

- (269) *si yoryora ido zias en*  
‘In Gorgoras, there is a certain Ziya.’ (01\_04022016F\_1; 018)
- (270) *to ido thena son bešgo pan*  
‘I put this thing above the stove.’ (04\_01072019F\_2; 106–107)
- (271) *layo da ida layo dune*  
‘How have these been, how have they been?’ (04\_01072019F\_1; 077)
- (272) *asa id eftayo na fai*  
‘I make a meal from them.’ (03\_30062019F\_6; 62)
- (273) *tadine t=ito serinluxi haðadžega bere*  
‘Their cooling (effect) reaches till here.’ (08\_04072019M\_3; 140)
- (274) *so spidi embro enan ido ormi*  
‘In front of the house [is] a river.’ (04\_01072019F\_13; 04)
- (275) *baxar kiš hada utš exun ida jani ida aledera onemada*  
‘“Autumn”, “winter” - They did not have them [=these terms]. They had different names.’ (04\_01072019F\_5; 01–02)
- (276) *n=ebine xordare dže, n=epine ida dže*  
‘We had to make hay, we had to make, ...thing.’ (04\_01072019F\_2; 135–136)
- (277) *epuliya do tšege ebedžega eberena ida dibu haidika dž=erxumune [...]*  
‘I sold it around there, from down there I bought something like this and returned.’ (04\_01072019F\_2; 111–112)
- (278) *as valume sin kofa ida opsara*  
‘Let us put this thing in the bucket ...fish.’ (04\_01072019F\_12; 34)
- (279) *opis ido ixame stali son baxari*  
‘Behind [it] we had, say..., a barn at the pasture.’ (04\_01072019F\_1; 154)
- (280) *hare ap edeni ejinedžikse*  
‘Now he married this one.’ (translation unclear; 01\_06042017F\_4; 139)

### 3.2.2.3 Possessives

This section discusses free and bound possessive pronouns in Romeyka; for adnominal possession see Section 5.3; for predicative possession see Section 5.2.1.5.5. In the following, first the grammatical properties of possessives are outlined; second, variation in the individual forms is presented.

Possessive pronouns in Romeyka are either anaphoric enclitics, or free possessive pronouns (Table 15), i.e., both pronominal forms are in complementary distribution (Drettas 1997: 150).<sup>82</sup> Free possessive pronouns originate from AG possessives, i.e., the neuter definite article plus the object pronoun (Mackridge 1987; Bortone 2009; Sitaridou 2014b). The 3<sup>rd</sup> person possessives (both strong and weak) derive from the (deictic) personal pronouns and therefore inflect in singular in agreement with the gender of the possessor (cf. Drettas 1997: 148); otherwise, full possessive pronouns have no gender agreement (cf. Mackridge 1987: 128 on *temeteron*) but agree with the possessee in number (see also Drettas 1997: 150; for examples see below). According to their grammatical behaviour in the NP, the possessive pronouns are in the Greek linguistic tradition considered pronominal adjectives (cf. van Emde Boas et al. 2019: 92). However, the term “pronominal adjective” is not adopted here, for it is found misleading that adjectives can occur in multiple number in nominal modification, while only a single possessive determiner can occur. Diachronically, the definite (genitive) article has become part of the full possessive pronoun.<sup>83</sup>

<sup>82</sup> But note that in translation tasks from Turkish, occasionally the free and bound possessive are used simultaneously, probably as a calque of the Turkish prompt with emphatic possessive pronoun (ii, iii).

- (ii) *aso istanbol op erthe ineka temo.. e.. jenge=m e*  
 ‘The woman who came from Istanbul is my aunt.’ (B1)
- (iii) *tema ta dört tane ta kosaras=mu efeyane*  
 ‘My four chickens escaped.’ (C1)

A rare construction is also (iv) with the clitic possessive co-occurring with the personal pronoun. Even if possibly triggered by the self-correction, the use of the personal pronoun instead of the possessive pronoun is rare. However, the confusion of free possessives with (subject) personal pronouns occurs frequently in Turkish copies like in nominalizations (v) and predicative possession modelled on Turkish existentials of the type free possessive + *var* ‘exist’ (see Section 5.2.1.5.5).

- (iv) *eyo to ða.. to şeri=m bola bonise*  
 ‘My hand hurt much.’ (02\_2906019F\_1; 29)
- (v) *esi.. teson t=erθinimon eyabo*  
 ‘I want you to come.’ (H3)

<sup>83</sup> But cf. (vi) below for the 3SG demonstrative pronoun and (vii) for the 3PL demonstrative pronoun used as possessives.

- (vi) *atşino inega le eduna mi bas*  
 ‘His wife said to him, “don’t go!”’ (04\_01072019F\_13; 11)
- (vii) *edžin do yarðeli xasdax edun*  
 ‘Their child was sick.’ (08\_04072019M\_3; 042)



**Table 15:** Full and bound possessive pronouns in Romeyka

	Clitic pronoun	Full possessive pronoun
1SG	= <i>(i)m</i> <sup>84</sup> , = <i>mu</i>	<i>temo(n)</i> , <i>tema.N/FPL</i>
2SG	= <i>(i)s</i> <sup>85</sup>	<i>teso(n)</i> , <i>tesa.N/F.PL</i>
3SG	= <i>((n)a)t.N/M</i> , = <i>((n)a)des.F</i> = <i>((n)e)θ(e)</i>	<i>tatinu.M/N</i> , <i>tatines.F</i> <sup>86</sup>
1PL	= <i>(e)mun(a)</i>	<i>temetero(n)</i> , <i>temetera.N/F.PL</i>
2PL	= <i>(e)sun(a)</i>	<i>tesetero(n)</i> , <i>tesetera.N/F.PL</i>
3PL	= <i>(a)tun(a)</i> <sup>87</sup> = <i>ana</i>	<i>tatinetero</i> , <i>tatinuna</i> <sup>88</sup>

The initial vowel of the enclitic pronouns changes according to the phonological shape of the noun they are attached to. However, their status as clitics is arguable, although there is some morphophonological evidence for their cliticness (see also Section 3.2.2.1): (i) their phonological form is adapted to the word length and sentence stress (cf. ex. 7 vs. 8 below); (ii) they can merge with the noun (see ex. 9); (iii) 1<sup>st</sup> and 2<sup>nd</sup> person singular clitics show some form of vowel harmony (sensitive to backness), albeit only attested for the noun *tšur* ‘father’; (iv) their distribution in post-posed position is clearly different from pre-posed full pronouns that behave like modifiers, i.e., genitive modifiers. On the other hand, the 3<sup>rd</sup> person post-posed possessives (see ex. 23) as well as the post-posed possessive *mu* in 1<sup>st</sup> person singular (ex. 4) can take primary stress like free pronouns. The fact that SMG lacks possessive affixes, which are the only form of possessive marking in Turkish (see Dryer 2013c), suggests some dynamic in the development of possessive marking in Romeyka. The vast majority of possessives in the Romeyka corpus are clitic pronouns (265 tokens of clitic pronouns vs. 53 free pronouns). This suggests that weak possessive pronouns are the default markers of possession in Romeyka; free pronouns seem to carry an emphatic meaning, although further research is required here. However, a contact explanation for the dominance of possessive affixes alone reaches too short: firstly, in the Turkish type, possessive suffixes are obligatory and can co-occur with full pronouns for reasons of comparison, (distinctive) focus, or topic shift (Göksel & Kerslake 2005: 243); secondly, the existence of post-posed possessives in AG and the use of the (SMG-like) full post-posed pronoun *mu* in a conservative variety of Romeyka (i.e., ROF as spoken in Karaçam) suggest an internal dynamic that could have been reinforced by contact with Turkish, though.

As for their syntactical behaviour, free possessive pronouns precede their head (281a; but cf. Drettas 1997: 151 for PG, where they may occur after the head) and occur at the beginning of the NP (Drettas 1997: 166), while bound pronouns are enclitic on the head (281b). For a general overview on possession, incl. postnominal genitives, see Section 4.2.2. When the pronominal possessive adjectives modify a NP, the possessed noun requires a definite article (281a), like with other forms of modification. (If the clitic is used, the definite noun requires its definite article as well, 281b.) The grammatical role of the possessor NP is case-marked on the definite article (and the noun), the genitive modifier is not affected (281d, 281c; see also

<sup>84</sup> The 1<sup>st</sup> person singular clitic occurs also in the variant =*um* in *tšur=um* ‘my father’ (> *tšir(i)* ‘father’), see also Section 2.3.2.

<sup>85</sup> See Fn. 84; *tšur=us* ‘your father’. Neocleous (2020: 45) also notes the archaic form =*su* (cf. =*mu* for 1SG).

<sup>86</sup> Also, *tetšines/tatšine(s).F* (and *tetšinu.M/N* for PG, Drettas 1997: 148).

<sup>87</sup> Also the forms =*(a)dona*, =*dena* attested.

<sup>88</sup> Also, *tetšinuna/tatšinune* (cf. *tetšino* for distant possessors, Drettas 1997: 148). Neocleous (2020: 45) lists the form of the 3PL possessive pronoun as *tatinús*. The form *tatinetero* seems an archaic form which shows plural agreement with the possessee in -*a* (cf. the comparative suffix *tero(n)/tera.PL*), also based on the demonstrative: *tatšinetera.PL*.

Drettas 1997: 151). Possessives do not only modify NPs but also occur as anaphoric pronouns (282). As modifiers, their scope seems to be limited to the NP, so emphatic personal pronouns can co-occur with possessive clitics (283).

- (281) a. *temetero*            *t=ospī*  
           POSS.1PL        the=house  
           ‘our house’ (A1)  
       b. *t=ospit=emona*  
           the=house=POSS.1PL  
           ‘our house’ (A1)  
       c. *omon temetero i batsi*  
           like POSS.1PL the girl  
           ‘like our girl’ (01\_04022016F\_1; 071)  
       d. *tsi patsis =im to ketši*  
           the.GEN daughter =POSS.1SG the goat  
           ‘my daughter’s goat’ (A1)

- (282) *tadinedero aspro en*  
       ‘Theirs [=the cow] is white.’ (B1)

- (283) *ate ti ðulia=tis epitše tši hare tšimate tše kahete*  
       ‘She has done her work and is sleeping now.’ (A1)

Apart from nouns, possessive clitics can attach to adverbs, apparently not only if nominalized (ex. 284). According to Drettas (1997: 166), in PG, possessive clitics can be attached simultaneously to different elements of a modified NP, i.e., to both the adjective and the head noun. Furthermore, the clitic can just appear at the adjective, e.g., *to kalon=emun to peðin* ‘our good child’ (Drettas 1997: 166), which is not attested for Romeyka.

- (284) *ejnedžikse epan=am*  
       ‘He married above my head.’, lit. ‘above me’ (01\_06042017F\_4; 133)

Prepositions indicating direction/location, such as *s* ‘to, at/in’, fuse with the full possessive pronoun, like in *s+temetero* > *s=emetero* (for all forms see Section 3.2.5).<sup>89</sup>

In the following, examples are provided for the free and bound possessive pronouns, including variation.

Examples of the 1<sup>st</sup> person singular free and bound pronouns are presented in (285–287); an example for the 2SG clitic in (288). A ROF speaker from Karaçam uses the archaic form of the postposed 1SG possessive pronoun *mu* (cf. SMG), whose clitic status is dubious (289).

- (285) ***temo*** *i aðelfi as leyom*  
       ‘Let’s say my sister.’ (04\_01072019F\_5; 53)

- (286) ***dema*** *ti emsale ulinus endyona*  
       ‘I beat all my peers.’ (02\_21042018M\_2; 16)

<sup>89</sup> But cf. the rare construction in (vi).

(vi) *temetero tso xorio bola mejveðes ine*  
       ‘At our village there were many fruit trees.’ (02\_02022015F\_1; 180–181)

- (287) *to peði=m bal eš inedžis*  
 ‘My son is also going to marry.’ (01\_06042017F\_4; 056)

- (288) *ti mana=s bola eyapesa*  
 ‘I loved your mother much.’ (03\_07072019F\_1; 10)

- (289) *na dero da baxdžeðes mu*  
 ‘I will look after my gardens.’ (04\_01072019F\_17; 56)

For 3<sup>rd</sup> person singular, an invariant clitic  $=((n)e)\theta(e)$  (290) exists alongside the gender sensitive bound pronouns  $=((n)a)t$  (291, 292) for masculine/neuter and  $=(n)ades$  (293) for feminine. Like with enclitic object pronouns (Section 3.2.2.1.2), it is not clear what determines the form of the syllable onset (cf. ex. 291 vs. 292,  $=nades$  vs.  $=des$  in 293), although prosody plays likely a role. The status of  $=(n)ades$  (and to a lesser extent  $=((n)a)t$ ) as clitics is not fully clear, considering their etymology:  $=((n)a)t$  goes back to the AG pronominal form *atu* > *ato(s)* (M. Janse, p.c.) and for  $=(n)ades$ , probably a similar etymology can be assumed.

- (290) *exorepsane ama i mana=θe mono merakliena dune*  
 ‘They danced but her mother was a bit thoughtful.’ (01\_06042017F\_4; 061–063)

- (291) *n=eyval do bezi=nat*  
 ‘It is about to take off its nappy.’ (02\_29062019F\_2; 25)

- (292) *anda erde do tuvaletu=t na bai etši*  
 ‘When he needs to go to the toilet, he will go there.’ (04\_01072019F\_13; 07)

- (293) *andra=des bal exaθe hare*  
 ‘Her husband has died (by) now.’ (01\_04022016F\_1; 058)

According to Özkan (2013: 146),  $=(n(e))\theta e$  is in RSür only used for inanimate objects, animals, and babies, although ex. (290) (in ROF) seems to counter this (but see 296 with an animate referent).<sup>90</sup> Overall, the distribution of  $=(e)\theta e$  is not fully clear since it is apparently not only used for nouns, but in nominalized infinitives together with *emun* (294),<sup>91</sup> with other nominalizations (295, with a “non-referential” reading), and with spatial adverbs (296, although an object pronoun would be more likely here). Further research is required on the use, function and etymology of  $=(e)\theta e$ .

- (294) *afti i ðulian<sup>92</sup> to maθin=emun-eθe yola en*  
 this.NOM the.NOM job.NOM the learning=our-its easy be.3SG  
 ‘It is easy for us to learn how to do this job.’ (ROF; Sitaridou 2014b: 52; glosses modified)

- (295) *t=ejlukin=eθe ekserume ap esena*  
 ‘We know your goodness.’ lit. ‘its goodness of you?’ (01\_02022015F\_1; 11)

<sup>90</sup> Özkan (2013: 147) suggests furthermore that  $=(n(e))\theta e$  occurs predominantly in clause-final position but this cannot be confirmed with the present data.

<sup>91</sup> It is not clear whether  $=emun-e\theta e$  in (294) is a combination of two possessive clitics ‘our-its’ or rather a single complex form (as assumed by Sitaridou 2014b: 52, 2014a: 130; cf. Papadopoulos 1955: 59). Sitaridou (2014a: 130) mentions a potential contact explanation of this form as calque of the Turkish (short) infinitive, e.g., *oku-ma-sin-t* ‘read-INF-3SG.POSS-ACC’. Tursun (2019: 242), however, also mentions that  $-e\theta e$  (in the form  $-e\theta i$ ) attaches to personal pronouns, e.g., *tatin-eθi* ‘his/her’ and demonstratives, e.g., *avutin-eθi* ‘this’.

<sup>92</sup> Janse (2002: 221) notes that in some cases the possessor of a genitive phrase occurs without overt genitive marking, presumably modelled on the Turkish compound-type of possessive construction.

- (296) *vejaxute nejenene to zo epuka=neθe n=epiname fila*  
 ‘Or when the cow gives birth, we will put leaves under it.’ (04\_01072019F\_1; 172–173)

There is also some further variation with 3SG possessives whose analysis is not clear:

- (297) *ula=ði eš do tubi=ðe to tšai=ðe da demlika=θ ula tšiega ine*  
 ‘It has got everything: its tube, its tea, its teapot, all things are there.’  
 (08\_04072019M\_3; 167–168)

- (298) *ama elebo ta ðulia=ede*  
 ‘But I see her works.’ (01\_28062019F\_3; 07)

Examples of two 3<sup>rd</sup> person feminine full possessive pronouns, *tatine* and *tetšines*, are shown in (299, 300).

- (299) *tadine t=ito serinluxi haðadžega bere*  
 ‘Its cooling effect (=of the fog) reaches till here.’ (08\_04072019M\_3; 140)

- (300) *tetšineš d=anepse n=epije*  
 ‘Her niece went.’ (03\_30062019F\_6; 05)

Examples for 1<sup>st</sup> person plural are (301)–(303), and for 2<sup>nd</sup> person plural (304)–(306).

- (301) *temeteron to mahallen onda ertes me ti mana=s taništim esi etrižes ežis*  
 ‘When you came to our neighbourhood and I met with your mother, you were running there.’ (01\_07072019F\_1; 02–03)

- (302) *temetera i palei i drani ason baxari iki sefer xordar ebinane*  
 ‘Our forefathers, the elders, used to make hay from the pastures two times (a year).’  
 (08\_04072019M\_1; 067–071)

- (303) *ti mana=muna eleyame ema epinasame*  
 ‘We said to our mother: “Mom, we are hungry!”’ (02\_02022015F\_1; 114)

- (304) *emis tesetero do xizmedi ba eftam*  
 ‘We will serve you.’ (08\_04072019M\_4; 32)

- (305) *tesetera ta tahadžes mi inane haje bola šimu*  
 ‘Did you also have such a heavy rain on your side?’ (03\_30062019F\_6; 49)

- (306) *to nero=suna axev*  
 ‘Does your water flow?’ (03\_07072019F\_1; 48)

3<sup>rd</sup> person plural bound possessives are =(a)dona (307), =(a)dun(a) (308) or =(a)dena (309). /a/ in the onset occurs [C\_] and is dropped [V\_]. Neocleous (2020: 45) also notes /n/ as filler.<sup>93</sup> It is not clear what determines the occurrence of the three variants; in the present corpus, there is no evidence that the vowels may indicate gender-agreement. A second bound 3PL possessive is invariant =ana, which occurs only twice in the corpus referencing animals (310) or

<sup>93</sup> A hapax legomenon in the Romeyka corpus is /h/ (vii), but it is not clear whether it has a deictic function here.

(vii) *ta kalaθo ba sa rašas=hátuna*  
 ‘The basket was in their pannier.’ (07\_04072019F\_6; 14)

children (311)<sup>94</sup> in a construction resembling ‘possessive datives’ with the object pronoun plus possessive. It is not clear whether the double set of 3PL bound pronouns corresponds to the two 3SG possessive clitics. Further research is in order here.

The free 3PL pronouns also show a variety of forms going on one hand back to invariant personal and demonstrative *tatinuna* (312) and *tetšinuna/tatšinune* (313) but have on the other hand a different formation with number-sensitive *-tero/-tera* for both personal- and demonstrative-based *tatinetero* (see 11 above) and *tatšinetero* (314).

- (307) *med i raša=duna*  
‘with their pannier’ (08\_04072019M\_1; 07)
- (308) *epejnam ekovalinam ta ksila=dona*  
‘We went and carried their wood.’ (04\_01072019F\_2; 188)
- (309) *epiname fila dže epuka=dena eθename*  
‘We made leaves and we spread them under them (=the cows).’ (09\_04072019\_7; 09)
- (310) *efteme=ta ta xismetin=ana*  
‘We do them their maintenance.’ (04\_01072019F\_1; 112)
- (311) *ixa dörd dane xarðelae etšino ta xismetin=ana epina*  
‘I had four children, I cared for them.’ lit. ‘I did their care.’ (04\_01072019F\_1; 146–147)
- (312) *tatinuna tospi mikro en*  
‘Their house is small.’ (A1)
- (313) *tatšinune to zo ba aspron edune*  
‘Their cow was white.’ (H1)
- (314) *tatšinetera ta za aspra*  
‘Their cow [is] white.’ (C1)

#### 3.2.2.4 Reflexives

While an inherited means of expressing reflexives is not attested in the present data (but see ex. 319 below), the Turkish bare reflexive *kendi* ‘self’ is borrowed.<sup>95</sup> While in Turkish the word class of *kendi* is not clear, namely whether it is pronominal or nominal (Kornfilt 1997: 302), in Romeyka, *kendi* occurs as noun, receives the neuter definite article *to*, and carries the Romeyka bound possessive (315; see also Neocleous 2020: 44, ex. 36b). Note, however, that in questionnaire data, the Turkish use of *kendi* may be completely copied including the Turkish possessive suffix, e.g., *kendi-si* ‘himself’ (ex. 316a vs 316b). In (317), *kendi* is used as nominal modifier. Finally, the Turkish function of doubling the reflexive pronoun for emphasis, e.g., *kendi kendi-me* ‘myself’ (Kornfilt 1997: 139) is copied (in the Trabzon dialectal form) one-to-one into Romeyka (318).

<sup>94</sup> Note the apparent lack of number agreement between the object pronoun and the possessive clitic.

<sup>95</sup> In a translation task of Turkish reflexive pronouns, the reflexive is ignored in the first examples of the Romeyka equivalent but for later examples, Turkish inflected *kendi* + possessive suffix is copied.

- (315) *saluxe ba an ešis ađaga tibu tšališevīs do kendi-s teries*  
 ‘If you also have health, you work something down here and you maintain yourself.’  
 (08\_04072019M\_1; 306)
- (316) a. *hatšine pa kendisi epitšen=a*  
 b. *onu kendisi yaptı*  
 ‘She did it herself.’ (C1)
- (317) *to kendi ti đulias*  
 the.N.ACC own the.F.ACC work.F.ACC  
 ‘your own works’ (04\_01072019F\_2; 211)
- (318) *kilavo ipe kendi kendu-m-e*  
 ‘She said, “I pray for myself”.’ (01\_04022016F\_1; 127)

Ex. (319) features a reflexive reading expressed by post-posed ton. It is not clear whether this is the postposed full object pronoun or whether this is a relic of post-posed AG reflexives, e.g., *εαυτου* ‘himself’ (van Emde Boas et al. 2019: 345–349). Furthermore, it cannot be assumed that post-posed free object pronouns are generally provoking a reflexive reading, since they are not necessarily co-referential with the subject.

- (319) *ksilo (o)niden ne krifko=ton epis*  
 ‘He is about to hide (himself) behind the wood.’ (02\_2906019F\_1; 27)

### 3.2.2.5 Reciprocals

Reciprocity is expressed by nominalized pronouns in a fixed expression: *d=enan d=alo* ‘the one the other’ (320, 321). As Drettas (1997: 178) states, *ena* and *alo* form an opposition and belong as such together.

- (320) *alo mia as elebame denandalo*  
 ‘If we only could see each other once more.’ (01\_15022015F\_1; 21–22)
- (321) *bola ayapenide dinandalo*  
 ‘You loved each other much.’ (03\_07072019F\_1; 17)

Mutuality is expressed in ex. (322) by the Turkish borrowing *karşı-mız* ‘between us’, lit. ‘opposite us’, which is nominalized in Romeyka by a definite article and appears as prepositional phrase, *son karši-miz* ‘in between us’. This construction calques Turkish spatial relators which can take possessive suffixes and case suffixes, e.g., *ara-mız-da* ‘between us’ (Göksel & Kerslake 2005: 222).

- (322) *išallah ormoŋa ine asinda son karšimiz n=adon emorŋa yonuševum da teke u=boro na yonuševum emorŋa*  
 ‘Hopefully, it is nice. Actually, if it was in between us, I speak nicely but alone I cannot speak nicely.’ (03\_07072019F\_1; 34–35)

Finally, the example in (323) omits the reciprocal pronoun.

- (323) *rahat utš edune ama eyapesam tš=epiyame*  
 ‘It was not easy, but we loved [each other] and went (on).’ (04\_01072019F\_2; 318–319)

## 3.2.2.6 Relativizers

Romeyka has different items serving as relativizers. Relativizers in Romeyka are in general preverbal and different strategies of relativization apply according to the role of the head noun of the relative clause. Their status as pronouns is dubious as pronouns usually reflect some properties of the relativized NP within the restricting clause, for example, animacy (Payne 1997: 326). This does not apply to Romeyka relativizers which are all invariant, although some go back to indefinite and interrogative pronouns, which is from a typological perspective common (Payne 1997: 333). However, for the sake of practicality, all relativizers are presented here together according to their function. Invariable relativizers in Romeyka are the following: (i) *(n)d(o)*, (ii) *ots*, (iii) *op*, (iv) *pe/pu/pi*, although (v) a zero-relativizer strategy is possible, too (see Gandon 2016). The relativization strategies by means of these relativizers are presented below; but, crucially, much further in-depth research is required on relativization in Romeyka to determine the functions of each relative marker. For the syntax of relative clauses see Section 5.3.4.

(i) The relativizer *(n)d(o)/(n)t(o)* (ex. 324) is derived from the neuter article *to*, which is homophonous with the interrogative *to* ‘what’ (but *nto* ‘what’ in RSür (Özkan 2013), and some varieties of ROF; see also Drettas 1997: 351). It also exists in the form *nde* (325, 326), which seems to be a free variant, i.e., it is not grammatically determined, e.g., by gender.<sup>96</sup> It is not clear what determines the form of the relativizer with initial /n/; /n/ occurs predominantly [V\_] but not in all cases; the occurrence of the initial consonant cluster seems to be diatopically defined (but see a very detailed discussion in Drettas (1997: 351–354), suggesting partly free variation, phonological conditioning, and an influence of valency). The vowel /o/ can merge with a consequent vowel, but it may also remain intact; further (phonological) research is required here. *(N)do* relativizes human (i.e., probably also animate) (324, 327, 328) and inanimate subjects (329–331), also in predicatives, but according to Özkan (n.d.), only animate non-human nouns are modified. According to Neocleous (2020: 46), [±animate] but [-human] referents are relativized by *(n)do* (also Drettas 1997: 354). However, this is countered by the present examples. Apparently, *ndo* is productively spreading to all kinds of RCs. It furthermore relativizes inanimate object NPs both full (325) and pronominal (332). Note that the majority of RCs relativized by *(n)do* in the examples below are spatial adjuncts. Finally, *(n)do* is used as a free relativizer in headless relative clauses (HRCs; 333, 334). In (335), the relativizer occurs in a HRC together with an indefinite pronoun. In (336a), the relativizer is accompanied by a spatial adverb, while the analysis of (336b) is less obvious: (336b) could be interpreted rather than as RC as complement clause (CC), the relativizer may be here interpreted as definite article of a nominalized spatial adverb.<sup>97</sup> (For the use of *(n)do* with embedded questions cf. Section 5.2.3.3.5.) Relativization with *(n)do* is in all examples pre-nominal (except for 332). Like with other modifiers, the noun requires its own definite article (but cf. 328).

(324) [*s=osp* ***ndo*** *en*] *da insana edroyan eb edži*  
 ‘The people who are at the house eat from them.’ (02\_02022015F\_1; 185)

(325) *axmedis [son barxa ***nde*** *ba*] o đormo ediksen=a*  
 ‘Ahmet showed her the road that goes to the pasture.’ (B1)

(326) *as leyome [***de*** *n=eftem*] da đulias*  
 ‘Let us say the works we will do.’ (08\_04072019M\_3; 005)

<sup>96</sup> If ex. (327) would suggest *(n)de* as showing number-agreement, ex. (336) counters this.

<sup>97</sup> Also note further variation on the same prompt, whose analysis is unclear (viii).

(viii) *eyo [đe=b=ođe kahume] to xorion eyabo*  
 ‘I like the village where I live.’ (H3)

- (327) *[aso istanbolin d=örthen] i jinega temon i jenge en*  
 ‘The woman who comes from Istanbul is my aunt.’ (H1)
- (328) *[sin avlea to sanda] patsiðae ekusane araba lali*  
 ‘The girls playing at the street heard a car.’ (C1)
- (329) *mesela [ason baxtše d=efteme] to tsubaðin haidika*  
 ‘For example, the corn that we make from our garden are like this [=organic].’  
 (04\_01072019F\_2; 264)
- (330) *[sin džami eberan do en] do spidi temo e*  
 ‘The house opposite the mosque is mine.’ (C1, also H3)
- (331) *[havuda do lebis] do spidae ula šenlig esane*  
 ‘The houses you see there were all inhabited.’ (08\_04072019M\_2; 031)
- (332) *eyrapses etšinon [do ipamen]*  
 record.AOR.2SG that.ACC REL say.AOR.1PL  
 ‘Did you record what we have said?’ (Gandon 2016: 221; presentation/glosses modified)
- (333) *[har to leyo] inšallah ormofa ine*  
 ‘What I say now is hopefully nice.’ (03\_07072019F\_1; 33)
- (334) *[d=eleyan] edžin epiname*  
 ‘We used to do what they said.’ (02\_02022015F\_1; 098)
- (335) *dibon d=aftes ti lali o šamada n=efte*  
 ‘Anything that you do, makes a sound, a noise.’ (03\_30062019F\_6; 32)
- (336) a. *eyo [to abohondžeka kahum] to xorio eyabo* (C1)  
 b. *eyo [abohe kahume]? d=edžeha gzen* (B1)  
 ‘I like the village where we live.’

(ii) *Ots* is a pronominal relativizer derived from the indefinite pronoun *otinos* ‘who/what(ever)’. The form *ots* is invariant. Özkan (n.d.) also notes the pronominal relativizer *oti* (< *otinan* ‘whom(ever)’), which, however, occurs in the present data only as indefinite pronoun (see Section 3.2.2.7). According to Özkan (n.d.), both are only used for animate/human referents. This can be confirmed for *ots* in the present corpus. Ex. (337) below shows a subject-RC; object-RCs with *ots* are not attested in the present corpus. *Ots* can also relativize HRCs (338). Comparison of exs. (338) and (327) above shows that *ots* is apparently in free variation with (*n*)*do*. The ancient form of the relativizer *itina* is reported in Sitaridou (2014b: 30).

- (337) *[aso istanboli ots erte] o mualimis sin džami tšeka katse*  
 ‘The teacher who came from Istanbul lives opposite the mosque.’ (C1)
- (338) *[i aso istanbuli ots erte] temo i jenge je*  
 ‘The woman who comes from Istanbul is my aunt.’ (C1)



(iii) The (now invariant) indefinite pronoun *op* (< *opios* ‘who(ever)’)<sup>98</sup> is used apart from its use as indefinite pronoun (see Section 3.2.2.7) to relativize animate referents in both subject (339, 340) and object RCs.<sup>99</sup> It occurs in HRCs (341). Comparison of exs. (339) and (337) above shows that *op* is in free variation with *ots*.

(339) [*so istanbul ob erðe*] *o öretmenis sin džami edžega ekatse*  
 ‘the teacher who came from Istanbul lives opposite the mosque’ (B1)

(340) [*so istanbul op ine*] *ul o rumdža u=kserone*  
 in.the Istanbul REL be.3PL all the Rumca NEG=know.3PL  
 ‘Those (being) in Istanbul all don’t know Rumca.’ (A1)

(341) [*aso katoxor op erðe*] *iða=na*  
 from.the Katochor REL came.3SG saw.1SG=3SG  
 ‘I saw the one that came from Katochor.’ (A1)

(iv) Grammaticalized relative markers derived from the AG interrogative pronoun *ὅπου* ‘where’ are *p(u)/p(i)/p(e)* (Drettas 1997: 347–351).<sup>100</sup> It is not fully clear whether *p(i)/p(u)* (Drettas 1997: 347–351, Özkan, n.d.) and *pe* (Sitaridou 2014b: 30) derive from the same root as *op*.<sup>101</sup> A reduced form *p=* exists as well (342). All relativizers are immediately pre-verbal. According to Drettas (1997: 347), *p(i)/p(u)* relativize human subjects (343). These relativizers do not occur in the present corpus; more research is required on their function and etymology, also on what determines the different vowels /i/, /u/, /e/.<sup>102</sup>

(342) *opse iða [alis p=epiren] ineka*  
 yesterday saw.1SG Alis.NOM REL=got.married.3SG woman.ACC  
 ‘Yesterday I saw the woman who Ali married.’<sup>103</sup> (Gandon 2016: 222;  
 presentation/glosses modified)

(343) *olon o istarnos [pu sintišen Romeika] onta xate etotes ta Romeika na ...*  
 ‘If the last Romeyka speaking person dies, then Romeyka will ...’ (Tursun 2019: 515)

(v) Finally, Turkish-dominant Romeyka speaker may apply a gap/null relativizer strategy (344, 345) – which is otherwise not common –, if the head noun is expressed (and with likely co-occurrence of the bound (object) resumptive pronoun, see also Section 5.2.1.4).

(344) [*aso katoxor erðe*] *o mexmetis iða=na*  
 from.the Katochor came.3SG the Mehmet saw.1SG=3SG  
 ‘I saw Mehmet who came from Katochor.’ (A1)

<sup>98</sup> It is not fully clear whether there might be an etymological relation between *op* and AG *ὅπου* ‘where’, like in the case of the relativizers *p(i)/p(u)/p(e)* discussed in (iv) below (and as is suggested by Gandon 2016: 221).

<sup>99</sup> But note ex. (ix) from a heritage speaker in Germany where an inanimate referent is relativized with *op*.

(ix) [*sin džamian beran ob en*] *t=ospitin temon t=ospitin en*  
 ‘The house opposite the mosque is my house.’ (H1)

<sup>100</sup> For an in-depth discussion of the evolution of *pu* see Nicholas (1998a: 11–21).

<sup>101</sup> Sitaridou (2014b: 30) suggests an etymology for *pe* < AG *ὅπερ*.

<sup>102</sup> But consider (x) where *pe* is used as indefinite pronoun in a complement clause.

(x) [*t ado pe at jeft*] *i nife=m eftje*  
 ‘What she’ does, is done by my bride.’ (07\_04072019F\_8; 14)

<sup>103</sup> I interpret ex. (342) as shown, but Gandon’s translation is slightly different.

- (345) *[aso istanbulin erθen] o mualimis sin džamian džega gatsen ga*  
 ‘The teacher who came from Istanbul lives opposite the mosque.’ (H1)

### 3.2.2.7 Indefinite pronouns

Indefinite pronouns mark specific and non-specific indefinite referents. Typologically, indefinite pronouns are often either interrogative-based (such as in SMG) or generic noun-based (like in Turkish) (Haspelmath 2013a). Romeyka has mostly interrogative-based indefinite items going back to pronominal forms, although their present pronominal status is unclear since they only show limited nominal agreement.

Romeyka also has a somewhat dubious indefinite pronoun going back to the ancient numeral *is* (< AG εἷς.M.NOM.SG ‘one’ (cf. the forms *énas/ínas*.M.NOM.SG), P. Mackridge, p.c.; Özkan, n.d.). P. Mackridge (p.c.) suggests it to be an emphasised pronominal of ‘one’, e.g., *an en is* ‘if there is [only] one [of them]’ (ROf as spoken in Saráchos (Uzungöl), example P.M.), see also (346). In Neocleous (2020: 143, ex. 56), *is* is analysed as indefinite pronoun ‘someone’. In the Romeyka corpus, it occurs predominantly in nominal form with definite (347) but also indefinite article (348). While Tursun (2019: 243) translates *is* with the Turkish indefinite pronoun *biri* ‘one’, in present word list data, *is* has been offered as equivalent to Tr. *kışı* ‘person’ (cf. 349).

- (346) *poli θelun=a tše is epere*  
 ‘Many [men] want her, but [only] one takes her.’ (Tursun 2019: 243)
- (347) *o o kadara adamlari var idi o is don ađelfo=m ebare o jalo din ekraba=m ebare*  
 ‘She had so many (candidate) men: one [said], “Take my brother!”, the other: “Take my relative!”.’ (01\_04022016F\_1; 109–111)
- (348) *so xorio=muna ena is edone [...]*  
 ‘At our village there was a person [...].’ (C1)
- (349) *as enan pendžere is teri*  
 ‘From a window, a person watches out.’ (05\_03072019M\_3; 07–08)

To express the meaning of ‘some’, Romeyka uses pronouns based on the interrogatives + pre-fixed *ka-*. Some of the interrogatives show number agreement (see section 3.2.2.8). The corpus shows two examples of these indefinite pronouns: *ka-* + the interrogative *posa* ‘how many’ > *kabosa* ‘some’, what the speaker corrects in a metalinguistic comment (350), and *ka-poθen* ‘somewhere else’ (03\_30062019F\_11; 083). Tursun (2019) additionally lists *ka-pote* ‘sometimes’, *ka-pia* ‘some’, *ka-pion* ‘someone’ (for the verbal indefinite ‘something’ see below). For inanimate entities *tipola* is used in the meaning of ‘some’, e.g., *tipola ðondae* ‘some teeth’ (04\_01072019F\_6; 18). Otherwise, the speakers draw on Turkish constructions, e.g., *birkac tane* ‘some pieces’ (351, 352) and the Turkish interrogative-based *kimi* ‘some’ (353).

- (350) *kabosa patiðaes xali plinun dže kahundane (ekabosa mono janliša intane)*  
 ‘Some girls are weaving carpets. (“ekabosa” – it was a bit wrong.)’ (B1)
- (351) *birkadž dane patsiðae ta xalia plinune*  
 ‘Some girls are weaving carpets.’ (C1)
- (352) *ebiya bir katš dane ebira dže ertha*  
 ‘I went, took some and returned.’ (03\_30062019F\_11; 037)

- (353) *kimi ixame sevdaðes kimi eberene so pudžaris sindižene o jalo saldu budža*  
 ‘Some had lovers, some hid in the corner to talk’, the other hid in the corner.’ (translation unclear; 04\_01072019F\_2; 165–167)

Note that the indefinite adverb of place *buθen* ‘somewhere’ (354, 355) is used in both affirmative and negative contexts (although this is likely not the case for *kapoθen* ‘somewhere else’).

- (354) *buθenjega u=boris ne vriškis=a*  
 ‘You can find this nowhere else.’ (08\_04072019M\_1; 254)
- (355) *utš eθelene buθen ebejene*  
 ‘She did not want to go somewhere.’ (01\_04022016F\_1; 105)

Indefinite persons are referred to by *ka(i)nis* ‘somebody’, also ‘nobody’. *Kanis* is a pronoun which inflects for case (ex. 356 for nominative, 357 for accusative, but cf. 358). If used in the meaning of ‘nobody’, the pronoun does not negate the clause, rather a separate sentence negator is used (359, 360). According to Janse (p.c.), *kanis* goes etymologically back to MedGr *καὶν εἷς* ‘even one’, thus including the numeral *is* ‘one’ (see above).

- (356) *kainis emas jardim utš ebine*  
 ‘Nobody helped us.’ (08\_04072019M\_1; 018)
- (357) *kanina utš eredžes sa ðorme*  
 ‘You did not see anybody on the way.’ (04\_01072019F\_17; 39)
- (358) *s kanis udž eθeles*  
 ‘You did not want to anybody.’ (01\_04022016F\_1; 112)
- (359) *aða opis en har ekabaniefte har kainis tš=en*  
 ‘[It] is behind here, now it is closed, there is nobody now.’ (08\_04072019M\_2; 020)
- (360) *etsakonam tš=etroyame ama laja foyumunesine kanis mi lepe=masine*  
 ‘We cracked them and ate them, but how we were afraid, that nobody should see us.’ (04\_01072019F\_2; 288)

The indefinite pronoun referring to indefinite things is *kati/kate* ‘something’ (361–363), but also ‘somehow’ (364, 365). It does not figure in negative contexts in the corpus.

- (361) *ti mana=m kat na ðiyo*  
 ‘I will give something to my mother.’ (05\_03072019M\_1; 495)
- (362) *so xor ha pies kadi pao eftayo kadi fidevo kadi tsakono*  
 ‘When you go to the field, I go [and] do something, I plant something, I break something.’ (03\_30062019F\_7; 07)
- (363) *o rebiz emena ba kade dūšūnevde*  
 ‘God has thought of something for me.’ (03\_30062019F\_8; 06)
- (364) *kati ta malia=s elaiskusanie etrešes*  
 ‘Somehow your hair waved [while] you were running.’ (03\_07072019F\_1; 15)

- (365) *kate isis d=edžadže dže vreš*  
 ‘It is somehow dreary there and it rains.’ (01\_04022016F\_1; 128)

To express the verbal indefinite ‘something’, *tipo(n)/tipu* is used (366, 367). Although two variants exist (368 vs. 369), the difference seems not to be grammatical (anymore), i.e., there is no inflection, and *tipo(n)/tipu* can strictly speaking not count as a pronoun. It can be modified (370). Most often, *tipu* is used in negative constructions, thus meaning ‘nothing’ (371). When used in negative clauses, *tipu* does not function as clause negator, and therefore needs to co-occur with the clause negator *utš* (372).

- (366) *tipu an en eđone=mase*  
 ‘If there was something, she gave it to us.’ (02\_02022015F\_1; 118)
- (367) *tip=utš eperenam tš=edroyame [...]*  
 something=NEG took.1PL CONJ=ate.1PL  
 ‘We could not buy something and eat it [...].’ (04\_01072019F\_2; 250)
- (368) *na baz eberis tipo dros*  
 ‘You will go, buy and eat something.’ (01\_28062019F\_3; 60)
- (369) *dibon d=aftes ti lali o šamada n=efte*  
 ‘The sound of something that you do will make noise.’ (03\_30062019F\_6; 32)
- (370) *bal aletera dibo xorevame*  
 ‘we also danced something else’ (04\_01072019F\_2; 196)
- (371) *utš edune te.. kaini.. tibu tš=edune*  
 ‘There was nobody.. there was nothing.’ (02\_02022015F\_1; 120)
- (372) *oston evle tipu tše na troyo*  
 ‘I will not eat anything till noon.’ (C1)

*Ob(i)* (373, 374) and *ots* (375) are indefinite pronouns which also function as relativizers (see Section 3.2.2.6); they refer to animate entities and their meaning equals more or less ‘who(ever)’. Furthermore, when referring to a single indefinite person, the speaker in (376) uses a Turkish interrogative-based pronoun *kimi* ‘who.acc’ first, before correcting herself with an inherited strategy of genitive + *ena* (here *ino*) ‘one’.

- (373) *op efoyudune udž enivene ob udž efoyudune enivene*  
 ‘Who(ever) was afraid did not go up, who(ever) was not afraid went up.’  
 (02\_02022015F\_1; 168)
- (374) *to tšofalin=at op etšalištupse [...]*  
 ‘Whoever puts his head to work [...].’ (P. Mackridge, personal materials from Saráchos (Uzungöl))
- (375) *ots edone mo don džir ebiname havešluk*  
 ‘Who(ever) was with father, we aspired.’ (02\_02022015F\_1; 009–010)

- (376) *endoyname me da yarðele kimi... du ino do tšefal iški odun du ino do šere tsakodun do ena.. du ino do poðari*  
 ‘We beat with the children: someone... of one the head split like wood, of one the hand broke, of one the leg.’ (04\_01072019F\_2; 057–060)

When referring to indefinite inanimates, the indefinite pronoun *oti* (*oti* ‘that what’; cf. *otinos/ots* ‘s/he who’, cf. Tursun 2019: 393) is used (377, 378). In (379), the pronoun occurs in the genitive-inflected form *otu* and seems to refer to a human referent. (380) features the pronoun *otesan* which is probably related to *oti* but appears in the definite meaning of ‘all this’. Another pronoun which is likely related but not clear in its analysis is *odene* in (381). (382) features the rare use of *oti* as preposition of purpose; it is not clear whether Turkish might have had an impact here.

- (377) *i mana=muna od eleje etšino*  
 ‘What our mother said, that [was done].’ (02\_02022015F\_1; 091)
- (378) *ode od efdas šaka itšun jela arkadaši=s*  
 ‘Whatever you do for fun, your friend laughs.’ (02\_02022015F\_1; 147–148)
- (379) *o kadar sikinti exa ðio otu udž ine*  
 ‘She had so many problems because it were not two (=children).’ (01\_04022016F\_1; 086–087)
- (380) *odesan epina erxumnje so spidi [...]*  
 ‘All these I did, (then) I came home [...].’ (04\_01072019F\_1; 150)
- (381) *odene n=efdjeme*  
 ‘We make everything.’ (04\_01072019F\_1; 064)
- (382) *ode ederesame tsi mana du džuru do lakirdi*  
 ‘For that we regarded mother’s and father’s word.’ (02\_02022015F\_1; 050–051)

As opposed to indefinite pronouns in *ka(p)*- ‘some-’, pronouns in *o-* + interrogative have an indefinite interrogative meaning, e.g., *o-pote* ‘whenever’ (383). In order to refer to indefinite places, plain interrogatives (384) and spatial adverbials (385) can be used (see also Section 5.2.3.3.5 on embedded interrogatives).

- (383) *obode thelis ela*  
 ‘Whenever you want, come!’ (08\_04072019M\_4; 28)
- (384) *erpu<sup>?</sup> en epejename*  
 ‘Wherever it was, we went [there].’ (09\_04072019\_7; 14)
- (385) *ayome katuhaž apohondžeka thelis ahom gadur*  
 ‘Go down there, go wherever you want and pee!’ (07\_04072019F\_6; 40)

### 3.2.2.8 Interrogatives

Interrogatives in Romeyka are derived from different word classes. Some are pronominal and inflect for number, gender or case, while others (for example those being derived from prepositions/adverbs) are invariant. Like with other pronouns outlined above, the interrogative

pronouns showing nominal agreement could be called “pronominal adjectives” as they can be used as modifiers (see 390 below).<sup>104</sup> In polar questions, Romeyka has borrowed the Turkish interrogative particle *mI* (see Sections 3.2.4.2, 5.2.3.3.1). From a typological point of view, interrogatives are often similar to relative or indefinite pronouns (Payne 1997: 300; also Dryer 2013d); this is also the case for some of the pronominal interrogatives in Romeyka which are related to indefinite pronouns and relative pronouns (e.g., *tinan* ‘whom’, *tilayo* ‘how’, *toson* ‘how much’). Romeyka interrogatives are presented in Table 16 and will be briefly characterized with examples below.

**Table 16:** Interrogatives in Romeyka (presentation adapted from Michelioudakis & Sitaridou (M&S) 2013: 356–358)

	+HUM	–HUM
NOM	<i>píos/pía</i> .PL <sup>105</sup> ‘who’	<i>pío</i> ‘which’
	<i>kaškiši</i> (M&S 2013) ‘how many’	<i>píon</i> (M&S 2013) ‘what’
ACC	<i>tín(a(n))</i> <sup>106</sup> ‘whom’	<i>píon</i> (M&S 2013) ‘what’
	<i>kaškišus</i> (M&S 2013) ‘how many’	<i>(n)d(o), (n)de</i> ‘what’
		<i>(n)doxna(n)/do(h)na</i> ‘what’
Other	<i>p(u)</i>	‘where’
	<i>póte</i>	‘when’
	<i>óðe(n), oðande</i>	‘why’
	<i>ja to pío</i> <sup>107</sup>	‘why’
	<i>pos</i>	‘how’
	<i>(ti)layo.N/layo(s).M/layosa.F/laya.PL</i>	‘how’
	<i>póso(n)/pósa.PL</i>	‘how much/how many’
	<i>(a)tóson</i> (TD)	‘how much’
	<i>kadž dane</i>	‘how many’
With spatial specification	<i>pu-merea</i> <i>pu-tšeka</i>	‘where.side’
With preposition	<i>abohe(n)</i> <sup>108</sup> / <i>-tšeka</i>	‘from.where’
	<i>poθen</i> (TD)	‘from where’
	<i>os-pu-džega</i>	‘until where.side’
	<i>mo-tinan, me dína</i>	‘with whom’
	<i>s-posa</i>	‘?how many.PL’

Pronominal interrogatives – i.e., those that inflect for some properties of the NP – are *píos* ‘who’, *tinan* ‘whom’, *poson* ‘how much/many’, *layo* ‘how’, and probably also *ndo* ‘what’, although they (probably) stem from different word classes. These “pronominal interrogatives” are outlined below, followed by the invariant interrogatives.

The pronoun *píos* ‘who’ inflects for number (386, 387) and case (Table 16) but likely not for gender. According to Michelioudakis & Sitaridou (2013: 357), interrogatives in Romeyka

<sup>104</sup> While the interrogative *pote* used as modifier in (390) carries the definite article of the NP, no article occurs with pronominal *tinós* in (xi).

(xi) *hato tinós peði en*  
‘Whose son is this?’ (Tursun 2019: 179)

<sup>105</sup> Neocleous (2020: 46) also notes *ts(i)ʔ* for human referents, a form which is probably closer to AG *ti(s)* than the forms of *píos*.

<sup>106</sup> Tursun (2019: 478) also notes the existence of the nominative form *tinós* ‘whose’ (see also Fn. 104 above).

<sup>107</sup> Tursun (2019: 218) lists for ‘why’ also the forms *ja pío* and *ja nto*.

<sup>108</sup> Elsewhere: *apoxen*; *apóçen* (with voiceless palatal fricative) in Neocleous (2019: 46).

have grammaticalised [ $\pm$ HUM] restrictions but lack number/gender distinctions altogether; see (388) where the Turkish loan *kaš(i)kiši* ‘how many’ is optionally used together with *pios* ‘who’ to add an indication of plurality. However, exs. (389, 390) show that *pio* is apparently also used for non-human entities, thus meaning ‘which’. Note that in (390), *pio* is treated like a modifier and occurs together with the noun after the definite article (cf. also ex. 411 below).

(386) *havudo pios en tina na pulis=a*  
 ‘Who is this, to whom will you sell this?’ (03\_07072019F\_1; 07)

(387) *akome kseris pia ine*  
 ‘Do you still? know who they are?’ (02\_21042018M\_2; 04)

(388) *Pios kaš(i)kisi erθen?*  
 Who NOM.PL came.3SG  
 ‘Who (=many) came?’ (Michelioudakis & Sitaridou 2013: 358)

(389) *to xorio=s pio en*  
 ‘Which is your village?’ (01\_04022016F\_1; 013)

(390) *drayozo trayoðia me do bio karðia*  
 ‘I sing folksongs, with which heart.’ [folksong] (04\_01072019F\_1; 068–069)

*Tina(n)* ‘whom’ is an object pronoun only used for human entities (391).

(391) *alo din=avres*  
 ‘Whom else did you find?’ (04\_01072019F\_17; 28)

*Poso(n)* ‘how much/how many’ inflects for number, with *posa* as plural form (392, 393).

Furthermore, there is a single occurrence of a pronoun *dosdon* (> *toson* ‘how much’<sup>109</sup>, cf. Tursun 2019: 481) which asks for an inanimate object (394).

(392) *bosa bikane*  
 ‘How many did they make?’ (01\_28062019F\_4; 60)

(393) *s posa i da yarðela*  
 ‘How many children do you have?’ lit. ‘at? how many are the children’  
 (01\_04022016F\_1; 080)

(394) *dosdon exume hajes*  
 ‘What (=how much) do we have like this?’ (03\_30062019F\_6; 68)

*Layos* ‘how’ inflects for gender (395, 396) and number (397, 398; see also Tursun 2019: 333–334). It also occurs as *tilayo* ‘how’ (01\_28062019F\_2; 42; Tursun 2019: 477).<sup>109</sup> Possibly, *layo* is on its way to become invariable as the neuter singular form can apparently be used for both masculine singular referents (399) and plural (400, but cf. 398). The semantic difference between *layo* ‘how’ and *pos* ‘how’ (see below) is difficult to discern, although *layo* seems to be preferably used with copula clauses and thus maybe has a more abstract meaning; further research is required.

<sup>109</sup> The etymology of *layos* is unclear. M. Janse (p.c.) notes that it seems not to be of Greek origin and proposes an etymology of *layo* < Tr. *hal-* ‘state’ + adjective endings.

- (395) *o vahidis layos e*  
the.M Vahit.M how.M COP.3SG  
'How is Vahit?' (04\_01072019F\_5; 25–26)
- (396) *layosa ise*  
'How are you.F?' (01\_04022016F\_1; 001)
- (397) *to tsubađi ta fasüljas laya sane*  
'How were the corn, the beans?' (04\_01072019F\_17; 49)
- (398) *laya yomare fordumune tše erxumunesten*  
'How did we carry the loadings?' (04\_01072019F\_1; 195)
- (399) *adžaba andras asi layo en*  
'How is actually her husband?' (01\_28062019F\_2; 35)
- (400) *ta laxana layo koševume dž=eftame laxana ne*  
'How we cast and make cabbage dish, right?' (03\_30062019F\_11; 081)

Finally, the interrogative *ndo* 'what' is used for non-human objects (401). Its etymology that leads to the complex syllable onset is unclear. The fact that it occurs in two forms (*n*)*do* (402, 405) and (*n*)*de* (402–404) lead to the question whether there is a grammatical difference between the two forms, for example regarding number agreement. This can, however, not be fully verified with the existing data (see, e.g., 402). As becomes evident in ex. (405), *ndo* can also ask for a cause.

- (401) *i mana=s d=eftjej*  
'What does your mother do?' (03\_07072019F\_1; 02)
- (402) *audan do ine iba na laxana nde na ine*  
'What are these? I told him, cabbage, what should it be?' (03\_30062019F\_11; 061)
- (403) *nde na leyo=se*  
'What shall I tell you?' (02\_05072019F\_1; 09)
- (404) *nde na troyume u=ksero*  
'I don't know what we will eat.' (02\_05072019F\_1; 25)
- (405) *mana daha ndo kahese*  
'Mother, why are you sitting here?' (03\_30062019F\_6; 30)

Invariant interrogatives are the "relational adverb" *pu* 'where' (Nicholas 1998a: 1; 406, 407), *pote* 'when' (408), *ođen* 'why' (AG *hoθen* > *ōθev* 'wherefore', Sitaridou 2014b: 30; ex. 409, 410), *pos* 'how' (411), *doxnan* 'what' (412–414), and the Turkish loans *kaškiši/kadž dane* 'how many' (415).

For the use of interrogatives like *pu* and *pote* as adverbial subordinators see Section 5.2.3.3.5 on embedded interrogatives and Section 5.3.3.1 on adverbial clauses.

- (406) *pu je i mana=s feren=ana so janis*  
'Where is your mother? Bring her to your side!' (01\_04022016F\_1; 038)



- (407) *b=esane haoda da šimundos*  
‘Where have these rains been?’ (03\_30062019F\_6; 44)
- (408) *i nife nades pote n=arde*  
‘When will her daughter-in-law come?’ (04\_01072019F\_5; 20)
- (409) *oðande de udž ebidžes sade ena dek de n=eftas=a*  
‘Why didn’t you do [more], what will you do with a single one (=child)?’  
(01\_04022016F\_1; 083)
- (410) *orðo tš=epijes etši*  
‘Why did you not go there?’ (04\_01072019F\_2; 215)
- (411) *to kendi ti ðulias to pos efded edži*  
‘How did you do your own work there?’ (04\_01072019F\_2; 211)
- (412) *dohna na leyo alom*  
‘What else will I say?’ (03\_07072019F\_1; 21)
- (413) *donas en so dolabi*  
‘What is in the wardrobe?’ (04\_01072019F\_17; 10)
- (414) *har esena tš=eyo na sindišenume esi ndona bidžes opse tš=osimero tše*  
‘Now you and me will talk: what did you do, yesterday, today.’ (04\_01072019F\_5; 06)
- (415) *xarðelae kadž dane i mana=m efine=me dž=ebeine*  
‘Children - how many - the mother brought [them] to me and left.’ (04\_01072019F\_1; 145)

The interrogative *doxna* ‘what’ exists according to Michelioudakis & Sitaridou (2013: 358) only in ROF and is only used for non-human entities (but cf. 04\_01072019F\_5; 14 in the corpus). Furthermore, it is assumed only to exist for objects (associated with accusative case). However, ex. (413), features the form *donas* as head of a copular expression. Finally, ex. (414) suggests that *doxna* can also occur with the word-initial consonant cluster /nd/ which is otherwise only known for the interrogative *ndo* ‘what’. From the point of semantic coverage, it is not clear what the difference is between *doxna* ‘what’ and *ndo* ‘what’ (cf. exs. 412 and 403); further research is required here.

### 3.2.3 Quantifiers

The category of non-numeral quantifiers in Romeyka is not large, many of the meanings usually expressed by quantifiers like ‘much’, ‘many’, ‘few’, ‘some’, ‘a lot of’ are covered by few inherited quantifiers (and adverbs of degree), e.g., *bola* ‘very/much’, *mono(s)* ‘a bit’, *eliyo* ‘a little’, and the universal quantifiers *ulon* ‘all’, *olon* ‘all/more’, *alon* ‘more/other’;<sup>110</sup> the category is generally open to Turkish loans.<sup>111</sup> While the universal quantifiers show remnants

<sup>110</sup> Unlike the usual procedure for nouns, pronouns and modifier, the citation form of the lemma is here not the masculine nominative singular, because the forms in *-os* are practically not attested; rather, the neuter (and accusative) ending in *-on* is the most frequent variant and is therefore adopted as citation from here.

<sup>111</sup> It is not straightforward to distinguish quantifiers from adverbs of degree; referring to AG grammars, *monos*, *olos*, *alos* and *eteros* belong to the class of quantifiers, while *bola* and *eliyo* are adverbs (van Emde Boas et al. 2019: 358–359).

of nominal agreement, those quantifiers that are discussed in Section 3.1.4.1 as adverbs of degree do not normally inflect for nominal categories. Universal quantifiers can be used pronominally and appear in attributive or predicative position. If they modify a noun, they usually precede the noun with its article but do not trigger determiner spreading themselves; they can also appear post-nominally, though. From a typological point of view, universal quantifiers are argued to have some affinity with conjunctives, including focus particles (Gil 2013b). The latter could be relevant for the case of Romeyka, if one compares the use of the discourse particle *ba(l)* (Section 3.2.4.5.1) together with the non-numeral quantifier *bola* ‘much’, and the lack of it in the examples of the universal quantifiers below. According to Gil (2013b), universal quantifier and focus particles with a meaning of ‘also’, similar to the Romeyka topicalizer *ba(l)*, may share not only grammatical similarities but also similar meanings or even etymological roots.

In the following, the most frequent elements that function as quantifier are outlined. Quantifier with an indefinite meaning like ‘some’, i.e., *kati* ‘something’, *kanis* ‘somebody’, *tipu* ‘something’, are treated in Section 3.2.2.7 on indefinite pronouns. Note that Sitaridou (2014a: 142, ex. 97) lists an indefinite pronoun *kátha=is* ‘everyone’,<sup>112</sup> which is not attested in the present corpus, though (but see Section 3.1.3. for *ís* ‘(some)one’).

The universal quantifier *ulon* covers the meanings ‘all/everybody/everything’ (416–418), ‘whole’ (419), also ‘always’ (420), and ‘constantly’ (421). It is not fully clear whether *ulon* differs originally from *olon* (see below), which has, however, synchronically a different functional distribution – at least in ROF as spoken in Çaykara (incl. Saráchos (Uzungöl), P. Mackridge, p.c.). According to M. Janse (p.c.), both forms go back to AG *olos* ‘whole, entire’ (> SMG *olos* ‘all’; Capp. *ulo*, also other modern Greek dialects) and there might be still some semantic overlap. Tursun (2019: 387) does also not differentiate between *olon/ulon* ‘whole, all’ (also cf. Papadopoulos 1958).

- (416) *avudin ul dinlejevune=me*  
‘They are all listening to me.’ (01\_15022015F\_1; 16)
- (417) *ulin arajevune=me*  
‘Everybody calls me.’ (08\_04072019M\_3; 065)
- (418) *har ula me din araba*  
‘Now everything [happens] by car.’ (08\_04072019M\_1; 092)
- (419) *din türkian ulon eporpatesame*  
‘We visited whole (=every place in) Turkey.’ (02\_21042018M\_2; 36)
- (420) *ula me ta rašes ta yomare ekovalename*  
‘We always carried the loads with the pannier.’ (08\_04072019M\_1; 063)
- (421) *i mana ulo tin patsi elepe*  
‘The mother constantly watches her daughter.’ (01\_06042017F\_4; 020–021)

*Ulon* is a pronoun and inflects at least for number (422) and case (423–426), see Table 17. While Tursun (2019: 387) lists an inflectional paradigm including gender, gender-sensitive forms could not be clearly attested in the present corpus; however, there is a tentative distinction

<sup>112</sup> For an example of *kátha=is* ‘everybody’ see Neocleous (2020: 153, ex. 73c).

according to [humanness], which has also exceptions, though; therefore, the paradigm in Table 17 should be considered tentative rather than conclusive.

**Table 17:** Inflectional paradigm of quantifier *ulon*

	Sg.	Pl.
Nom.	<i>ulo</i>	<i>ul(a), ulin</i> (+HUM, -HUM?)
Gen.	<i>ulu</i> (TD)	<i>ulinun(a)</i> (cf. TD)
Acc.	<i>ulon</i>	<i>ulisuna</i> .2PL
		<i>ulinus</i> .3PL (+HUM), <i>ulun(u)s</i> (-HUM) (cf. TD)
		<i>ula(n), ulona</i> (-HUM)

- (422) *ta xorafe=muna t=ospidae=muna ula ađatses ine*  
 ‘Our fields, our houses are all here.’ (08\_04072019M\_1; 010)
- (423) *emena an erotai ulu t.. ulan anlateva=dona da hikajeres*  
 ‘If you ask me, I will tell you the whole story.’ (C1)
- (424) *ula tsi mamikandus uluns ayo kovalina*  
 ‘I carried all (things) of the elderly people.’ (04\_01072019F\_2; 146)
- (425) *dema ti emsale ulinus endyona*  
 POSS.1SG the.PL.ACC peer.PL all.ACC/GEN? hit.1SG  
 ‘I beat all my peers.’ (02\_21042018M\_2; 16)
- (426) *eyo ulisuna eyabo*  
 1SG all love.1SG  
 ‘I love you all.’ (A1)

Apart from its pronominal use (427–429), *ulon* functions as nominal modifier (430–432), adverbially (420, 421), it can co-occur with an adverb (433) and it can itself be head of a genitive modifier (419, 424).

- (427) *ulin me ta satšia eberenane t=alevre*  
 ‘They all bought flour with [=in amounts of] a sack.’ (08\_04072019M\_2; 152–153)
- (428) *ulin efteme piknigi*  
 ‘All together we make a picnic.’ (04\_01072019F\_1; 059–060)
- (429) *eyo pina tšen ulona*  
 ‘I made them all.’ (04\_01072019F\_2; 138)
- (430) *t=ulinuna t=ospidae mikra ine.*  
 the.GEN=all the.NOM=houses small be.3PL  
 ‘Everybody’s houses are small.’ (A1)
- (431) *erθane ula t=opsarae efaane*  
 ‘They came and ate all the fish.’ (05\_03072019M\_4; 39)
- (432) *ulin ta yarðele don džiri=nadin eyabo*  
 ‘All children love their father.’ (02\_21042018M\_2; 32)

- (433) *emist ul tomara so mektep pame*  
 ‘We went to school all together.’ (A1)

The quantifier can be combined with a preposition (434, 435; note that in ex. 434, the preposition occurs twice) or a pronominal clitic (436).

- (434) *s=ula sa mejveðes enivename*  
 ‘We climbed all the fruit trees.’ (02\_02022015F\_1; 183)
- (435) *as ula eftam*  
 ‘We use them all.’ lit. ‘We make from them all.’ (01\_28062019F\_3; 49)
- (436) *ula=ði eš*  
 ‘Everything (of it) is there.’ (08\_04072019M\_3; 167)

The invariant universal quantifier *olon* ‘all’ (*ol-on* ‘all-GEN.PL’ < AG *ολος* ‘whole, entire’) has probably diachronically the same root as *ulon* but synchronically, it has a different functional distribution: While *ulon* is mostly used pronominal (but cf. ex. 421 above), *olon* occurs predominantly (i) in temporal adverbials where its meaning is ‘whole’ (437, 438);<sup>113</sup> (ii) in the formation of superlatives in periphrastic comparison with an approximate meaning of ‘most’ (439; see also Section 5.2.1.6). In the latter meaning it also occurs in temporal adverbials (440). But note also (441) where *olon* occurs with the approximative suffix *-ahe(n)* (Section 3.1.4.4.4) for a rough meaning of ‘nearly everything’. The distribution of *ulon* and *olon* needs certainly further scholarly attention.

- (437) *eskiden olon din imera jaralmasija troyame*  
 ‘In earlier times, we used to eat potatoes every day.’, here lit. ‘the whole day’ (C1)<sup>114</sup>
- (438) *eskiden olon to xronon son baxari emunesdin*  
 ‘In earlier times, we used to be at the pasture the whole year.’ (C1)
- (439) *olon tranisa eyo mune*  
 ‘I was the oldest.’ (C1)
- (440) *olon embro(n)*  
 ‘at the very beginning, first of all’ (02\_21042018M\_2; 34)
- (441) *olonahe eferename me ti raša=muna*  
 ‘(Almost) everything we carried with our pannier.’ (09\_04072019\_7; 45)

When modifying adjectives, *olon* appears in pre-posed position. As nominal modifier, *olon* precedes the head noun and its article; it does not trigger determiner spreading, though. Note that *olon* as nominal modifier seems to appear also between definite article and head, e.g., *t=olon embrotikón* ‘the very first’ (P. Mackridge, Saráchos (Uzungöl) materials, p.c.); cf. AG *το ολον πραγμα* ‘the entire matter’ (van Emde Boas et al. 2019: 358).

<sup>113</sup> Note that Drettas (1997: 398) also attests *olon* in the general meaning of ‘whole’, e.g., *olon to psomin* ‘the whole bread’ (cf. ex. 419 with *ulon* above). Tursun (2019: 387) lists in this meaning a quantifier *olú(y)os*, e.g., *oluo faji efaya* ‘I ate a whole bread.’, which is not attested in the present corpus.

<sup>114</sup> In order to express temporal ‘every’ in Romeyka, often a Turkish adverbial is borrowed (xii).

(xii) *her akşam ta za so mandri valo*  
 ‘Every evening, I lead the cows to the stable.’ (C1)

*Alon* ‘other/more’ (< AG *αλλος* ‘other’) is like *ulon* and *olon* what Drettas (1997: 177) calls a ‘nom adjectif’. Its meaning entails the ‘other’ out of similar items, as opposed to the quantifier *edero* ‘other/further’ (see below), which stresses difference (van Emde Boas et al. 2019: 359). In the Romeyka corpus, it occurs predominantly adverbially (442-445), but it also modifies nouns and other pronominal items (446, 447). Furthermore, it modifies adjectives as periphrastic comparative marker (448; see Section 5.2.1.6). Finally, it occurs as nominalization (449, 450; see also as part of reciprocal *d=enan d=alo* ‘each other’, Section 3.2.2.5). While Drettas (1997: 177), also Tursun (2019: 100), list a full paradigm for *alon* sensitive to gender, number and case, in the present corpus only remnants of this occur (e.g., ex. 446 vs. 447). As pre-nominal modifier, *alon* seems to show case inflection (446, but cf. 447). In VPs, it can occur in all positions pre- and post-verbally; see also its intermediate position between both finite verbs in the potential construction in (443); in questions, it can be fronted (445).

- (442) *esena alo utš eftame.. utš anespalume*  
 ‘We will henceforth not make you.. not forget you.’ (01\_02022015F\_1; 09–10)
- (443) *u=borume alon eskumes*  
 ‘We could not stand up anymore.’ (02\_02022015F\_1; 162)
- (444) *dohna na leyo alom*  
 ‘What else will I say?’ (03\_07072019F\_1; 21)
- (445) *alo din avres*  
 ‘Whoelse did you find?’ (04\_01072019F\_17; 28)
- (446) *alon d=ajarevume hatši tše hađa*  
 ‘What else do we look for, here and there?’ (08\_04072019M\_1; 245)
- (447) *alos i dibon erodes=me*  
 ‘Will you ask me something else?’ (08\_04072019M\_1; 321)
- (448) *to šehiri alo masraflı e*  
 ‘The city is more expensive.’ (08\_04072019M\_1; 312)
- (449) *ondan sora t=alo sin dželia=m edune*  
 ‘Later, there was the other [=child] in my belly.’ (01\_06042017F\_4; 125)
- (450) *o is don ađelfo=m ebare o jalo din ekraba=m ebare*  
 ‘That person: “Take my brother!”, the other: “Take my relative!”.’ (01\_04022016F\_1; 110–111)

In the meaning of ‘other’, *alos* competes with the pronoun *edero* ‘other’, although the latter highlights difference from a certain entity while *alos* marks similarity (see van Emde Boas et al. 2019: 359). *Edero* inflects for number (451 vs. 452) and can be apparently combined with *alos* to *al+edero* ‘other’ (453-455). In (452) it is used pronominally, while as modifier, it occurs frequently together with Tr. *çeşit* ‘type’ (451, 456). The Turkish equivalent of *alo*, *başka* ‘other’, may occasionally interfere with *alo*, e.g., [...] *başka mamika erxudune* ‘another grandmother came’ (04\_01072019F\_2; 197); see also *felandžiðes* ‘others’ < Tr. *filan* ‘so and so’ (04\_01072019F\_2; 191).

- (451) *ta batsiðaes edero džesit xizmet exune*  
 ‘The support of daughters is different.’ (03\_30062019F\_7; 32)
- (452) *beker eðelesan=ana edera ðelesan=a*  
 ‘Unmarried wanted her, others wanted her.’ (01\_04022016F\_1; 106)
- (453) *ida aledera onemada*  
 ‘They had other names.’ (04\_01072019F\_5; 02)
- (454) *ena aletero ylossa maðama.. maðenimo bola zori e*  
 ‘Learning a further language is difficult.’ (C1)
- (455) *bal aletera dibo xorevame*  
 ‘Or we danced something else.’ (04\_01072019F\_2; 196)
- (456) *edero džešidin ylossa bola tšedin en*  
 ‘A different language is very difficult.’ (B1)

Finally, the adverb of degree *bola* (Section 3.1.4.1) appears frequently and covers manifold meanings, both qualitative and quantitative, including ‘very’ (457, also 458), ‘many’ (459–462), ‘much/a lot’ (463–466), ‘long’ (temporal (467, 468) and spatial (12)). *Bola* can modify verbs (e.g., 463), nouns (e.g., 459) and adjectives (457). If used as a modifier, it precedes the head it modifies. *Bola* can modify singular (465, 469) and plural (459, 460) nominal heads; it does not trigger determiner spreading, the head has no article. Its form *bola* is usually invariant but (462) features the quantifier as complement of a copular expression likely showing plural (human) inflection.

- (457) *bola makra ise*  
 ‘You are very far away.’ (01\_15022015F\_1; 23)
- (458) *resmije tš=esi bola arkadaši estini*  
 ‘Resmiye and you have been close friends.’ (03\_07072019F\_1; 16)
- (459) *pola ospidæ ine tši*  
 ‘There are many houses there.’ (A1)
- (460) *trana b=obsaræ eplezame bola ba*  
 big TOP=fish.PL caught.1PL many FOC  
 ‘We caught big as well as many fish’ (08\_04072019M\_1; 159)
- (461) *evra bola insani*  
 ‘I met many people.’ lit. ‘I found’ (04\_01072019F\_5; 13)
- (462) *o zaman poli emunestine*  
 ‘At that time, we were many.’ (04\_01072019F\_2; 305)
- (463) *šin bula tšališevune*  
 ‘They work a lot.’ (04\_01072019F\_2; 325)
- (464) *o jaš=im bola en*  
 ‘My age is advanced.’ lit. ‘a lot’ (01\_04022016F\_1; 007)

- (465) *ai bola ilos ebidžen*  
 ‘There was so much sun.’ (03\_30062019F\_6; 47)
- (466) *ton džiri=m bola utš eynorizo*  
 ‘I don’t know my father well.’ (02\_21042018M\_2; 10)
- (467) *bola estaθan edži*  
 ‘They stayed there for a long time.’ (03\_30062019F\_11; 051)
- (468) *u=boro na steko bola*  
 ‘I cannot stay long.’ (03\_30062019F\_6; 02)
- (469) *bola đromos*  
 ‘(It is) a long way.’ (04072019F\_10; 12)

The universal quantifier used to express the meaning ‘every’ is borrowed Turkish *her* ‘every’ which is always borrowed as a whole NP together with the head noun (470–472), sometimes even including Turkish case suffixes (471). As (472) shows, the Turkish construction for ‘everything’ is eventually substituted by the inherited Romeyka quantifier *ula* ‘all/everything’. Note that the manner adverb *tomara* ‘together’ occurs often together with the quantifier *ulon*, but seems to replace it in (473).

- (470) *i šeher kalo fazla masraflı efes tšingi her gün sin džebi so džüzđani=s uraševı [...]*  
 ‘In the city, you spend more money, because every day you deal with your wallet [...].’  
 (08\_04072019M\_1; 314–315)
- (471) *her jere ferume=sas*  
 ‘I will show you around.’ lit. ‘to every place’ (08\_04072019M\_4; 34)
- (472) *her šej.. ula mona makra*  
 ‘Everything is a bit far.’ (07\_04072019F\_6; 31)
- (473) *erθen tomara ksendžen=ata eban dže bič*  
 ‘He came, pulled up everything and left.’ (07\_04072019F\_8; 17)

### 3.2.4 Particles

Particles are phonologically free forms which show no agreement and convey a certain meaning to a related phrase. Romeyka uses particles in (i) negation, (ii) questions/answers, (iii) complementation and subordination, (iv) modality, and (v) discourse. The majority of particles in Romeyka is inherited, but a question and an answer particle as well as discourse markers are borrowed from Turkish. Frequent particles in Romeyka will be briefly characterized according to their function below. Note that this Section also contains information on (mainly borrowed) elements that are not strictly speaking particles; for example, Section 3.2.4.3. focuses on borrowed clause linkage markers, which are not necessarily particles in Turkish (subordinating conjunctions are otherwise discussed in Section 3.2.6.2). For the pan-Anatolian deictic element *ha* whose grammatical status is not finally agreed upon see Section 3.2.2.2.

## 3.2.4.1 Negation particles

Romeyka uses four negation particles: (i) *utš* for sentential negation in indicative sentences, (ii) *mi(n)* in imperatives and wishes,<sup>115</sup> (iii) *xe* in subjunctives,<sup>116</sup> (iv) *mutš* in counterfactuals (Sitaridou 2014a: 121). Chatzopoulou & Sitaridou (2014) also mention *miðen* and *tšas* in non-veridical clauses. In existentials, the Turkish negative existential particle *yok* is borrowed, phonologically frequently realized as /jox/ (474). For the other borrowed functions of Tr. *yok* in Romeyka see the following Section. For clausal negation and a detailed description of the individual particles see Section 4.3.6.1.

- (474) *kainis jox*  
 ‘There is nobody.’ (01\_28062019F\_2; 21)

## 3.2.4.2 Question/answer particles

The Turkish question particle *mI* is borrowed in an invariant form *mi* among others in polar questions and alternative questions (Section 5.2.3.3). According to Michelioudakis & Sitaridou (2016: 26), it occurs optionally in indirect questions and obligatorily in (polar) direct questions “of total ignorance”. They analyse invariable *mi* as phonological clitic since it obligatorily requires a head to its left.

The Romeyka positive responsive particle is (like in Greek) *ne* ‘yes’, or in discourse more emphatically *he* ‘yes’ (475a). Question particles can be doubled. As a negative responsive particle, the Turkish existential *yok*, which is also used as an answer particle in informal Turkish, is borrowed (476; also Mackridge 1987: 134). *Jok/jox* also functions as question tag (475a), albeit without the question particle which would be obligatory in Turkish question tags (475b); note, however, that the Standard Turkish question tag is formed by the negative copula *değil* + interrogative *mI*, so *yok mu* seems to be probably an informal variant.

- (475) a. *he he heralda n=arde jok*  
 ‘Yes, yes, we will probably come, won’t he?’ (04\_01072019F\_17; 64)  
 b. *Herhalde gelecek, yok mu?*  
 ‘He will probably come, won’t he?’

- (476) *jok jok ade si yoryora edune*  
 ‘No, no, she was in Gorgoras.’ (01\_04022016F\_1; 056–057)

## 3.2.4.3 Borrowed clause linkage markers

The particle *ki*, borrowed from the Persian complementizer *ke*, is an areal phenomenon reaching far beyond Anatolia. In Romeyka, *ki* is used as complementizer of verbs of saying in indirect speech (477; also 01\_04022016F\_1; 096), a strategy which also exists in Turkish (Kornfilt 1997: 3; Göksel & Kerslake 2005: 356; cf. code-switching in 478), but see complementation of verbs of saying without complementizer in (479, 480). Possibly, the use of the complementizer *ki* in indirect speech is more prominent in Turkish-dominant speakers. For a detailed description of complementation strategies in verbs of saying see Section 5.3.2.1.2.; the multi-purpose modal particle *na* is a productive complementizer; see Section 3.2.4.4.

<sup>115</sup> Neocleous (2020: 63) also lists *me(n)* as a variant of *mi(n)* for imperatives and wishes.

<sup>116</sup> Sitaridou (2014a: 121) states *xe* for subjunctives; this analysis is followed by Neocleous (2020: 63–67) although he lists /tš/ as variant of *xe*. In the light of the present examples – and the examples in Neocleous (2020), it seems, however, that *xe* is a palatalized variant of the clausal negator *utš* which is used in subjunctives like in most other negative contexts.



- (477) *ipe ki eleya ki ej bejuk allah-um emena sin almanja nesib bisun*  
 ‘She said, I used to say, oh good God, predestine me to Germany.’ (01\_04022016F\_1; 118–119)
- (478) *sen de ki dib udž erlaepsa aso genšluk=im*  
 ‘You can say that I have not understood anything from my youth.’ (02\_02022015F\_1; 012)
- (479) *eb-edode iba as prao so xoraf*  
 ‘Later I said, I should go to the field.’ (03\_30062019F\_11; 030)
- (480) *lege as dero ine mi tibo na troyume*  
 ‘He said, let us see whether there is something to eat.’ (04\_01072019F\_12; 10–11)

Note that apart from its use as complementation particle, *ki* is – like in Turkish (Göksel & Kerslake 2005: 102) – used as a modality particle (“discourse connective” in terms of Göksel & Kerslake) indicating that a content is obvious to both interlocutors (481). Even if no further examples are available in the Romeyka present corpus, it can be assumed that other discourse functions of *ki* are also borrowed from Turkish (see also 3.2.4.5.2 below).

- (481) *etšine har u=fanerui ki*  
 ‘It can (obviously) not be seen now.’ (07\_04072019F\_6; 02)

The Turkish subordinating particle *diye* ‘in order to’ (actually gerund of the verb ‘say’, i.e., ‘saying’) is borrowed into Romeyka together with some of its syntactic features (Göksel & Kerslake 2005: 400–401). *Diye* appears in Romeyka (like in Turkish) in phrase-final position of finite adverbial subordinate clauses of reason (482, 483). The particle also appears in the form *dejine* to express purpose (484); cf. also the borrowed past participle TTr. *deduum* (< Tr. *de-diğ-im* ‘having said’) in (485), probably also expressing purpose. With the use of *diye* for adverbials of reason and purpose, only part of the many functions of Turkish *diye* are copied to Romeyka (see Göksel & Kerslake 2005: 400–401). Furthermore, in Turkish *diye* always occurs at the end of the subordinate clause followed by the main clause (Göksel & Kerslake 2005: 400), while in Romeyka, the subordinate clause is post-posed (see also Sections 5.3.3.1.4 and 5.3.3.1.5).

- (482) *o kadar eyabo=se [i batsi=muna esena eyaba dije]*  
 ‘I love you so much, because our daughter loves you.’ (01\_14012015F\_1; 5)
- (483) *ište bazilarina jardimdže epi inumunestine [ixane sevdaris dije]*  
 ‘We assisted some for they had a lover.’ (04\_01072019F\_2; 170–171)
- (484) *panda si ðulia muna epejname [ti ðulian n=eftjame dejine]*  
 ‘We always went to work for the work to be done.’ (09\_04072019\_7; 25)
- (485) *jane me so borbade ma deduum [...] si komsiðaes=muna so zijareti na bame*  
 ‘Having said going for a walk, we will go to visit our neighbours.’ (08\_04072019M\_3; 107–109)

Another borrowed particle in adpositional phrases is the Turkish postposition *göre* ‘according to’ (Göksel & Kerslake 2005: 217–218), which is in (486) used in its function as marking

comparison. Although no other occurrences of *göre* are attested in the present corpus, it is likely that it also occurs in its other functions in Turkish-dominant speakers.

- (486) *har esi ba ona göre emena ifeti jardimi bisun*  
 ‘now according to this, do help me this year’ (08\_04072019M\_3; 99–100)

#### 3.2.4.4 Modal particles

Romeyka has the following modal particles: *na*, *as*, *an*. The multi-purpose modal particle *na* is used for a wide range of functions. It also occurs in the free variant *ne*. It is used (i) for a variety of non-indicative clause types, including “subjunctives”<sup>117</sup> (487, 488; Section 4.3.6.2), conditionals<sup>118</sup> (489, 490) and counterfactuals (491; Section 5.3.3.2); (ii) as complementizer in volitionals (492, 493; Section 4.3.6.4), potentials (494, 495; Section 4.3.6.5), emotive, mental perception and some other verbs like ‘wait’ (496, 497), and in temporal adverbial clauses with *os* ‘until’ (Section 5.3.3.1.1).<sup>119</sup> For the syntax of *na*-complements see Section 5.3.2.1.3. Apart from its modal functions, *na* is used in Romeyka (and Cappadocian) as a future marker (498, 499; Section 4.3.2.4). The particle is always followed by a finite verb (as complementizer it occurs between two finite verbs). Unlike in AG, *na* is not followed by the subjunctive but rather by indicative mood.

- (487) *ta patsiðes utš ejollaevane n=eðuilevame*  
 ‘They did not send the girls, we should work.’ (02\_02022015F\_1; 044)
- (488) *ti mana=muna jardim n=epiname*  
 ‘We should help our mother.’ (02\_02022015F\_1; 045)
- (489) *sena dibu ne baðane [...]*  
 ‘If something happens to you [...].’ (01\_04022016F\_1; 098)
- (490) *emeklis ba na inese na stetšis aðatšeka*  
 ‘When you retire, you can stay here.’ (08\_04072019M\_1; 308–309)
- (491) *asinda du mintanis n=jedune da šašdim ona*  
 ‘Actually it should have been “mintanis”, I was mistaken.’ (03\_07072019F\_1; 30)
- (492) *ade beði eš batsi eyapume n=efda*  
 ‘She has a daughter, we wait for her to make a son.’ (01\_04022016F\_1; 091)

<sup>117</sup> Note that in line with the general typological terminology, the term “subjunctive” should be actually reserved for specific morphological forms of the verb themselves and should not be used to denote a clause type used for wishes, let alone to be used synonymous with non-indicative clause types complemented by the modal particle *na* (as done for example in Sitaridou 2014b). Still, the term “subjunctive” is as a preliminary decision used here to refer to verbal constructions with the particle *na* used for wishes to set them apart from other forms of modality, although the subjunctive is not a morphological form (see also 4.3.6.2 on subjunctive mood).

<sup>118</sup> Conditionals are usually formed by the particle *an* (see below). However, (non-/)veridical conditionals seem also to occur with *na*-clauses (Sitaridou 2014a: 122, Table 3). Both may be even used interchangeably (see x and Sitaridou 2014a: 136, ex. 73a/b).

(xiii) *para na ešis saluxe ba an ešis aðaga tibu tšalishevis do kendis terijes*  
 ‘If you have money, if you also have a good health, you work something here and you sustain yourself.’  
 (08\_04072019M\_1; 305–307)

<sup>119</sup> For PG, only part of these functions are described by Drettas (1997: 324–325).

- (493) *hidž udž eyapena n=esteka*  
‘I did not want to stay at all.’ (01\_06042017F\_4; 138)
- (494) *tevekela tib=utš eborenam n=epiname*  
‘Without asking, we could not do anything.’ (02\_02022015F\_1; 089)
- (495) *i paranta utš iš itšen ba u=boris n=epejnis*  
‘If you do not have money, you cannot buy it.’ (02\_05072019F\_1; 34)
- (496) *na bermenune na erde argos*  
‘They (will) wait whether the bear comes.’ (05\_03072019M\_4; 15)
- (497) *erazune na skodonone t=argures*  
‘They keep watch to kill the bears.’ (05\_03072019M\_3; 23–24)
- (498) *na bao so xorio*  
‘I will go to the village.’ (02\_05072019F\_1; 12)
- (499) *de n=eftas=a*  
‘What will you do with it?’ (01\_04022016F\_1; 083)

The optative particle *as* (also called exhortative particle,<sup>120</sup> Chatzopoulou & Sitaridou 2014) is used for assertions which are close to irrealis mood, i.e., no claim is made as to whether an action took place (Payne 1997: 245; cf. Drettas 1997: 308 for PG). This crucially includes wishes (Dobrushina, van der Auwera & Goussev 2013) both of the suggestive ‘let us’-type (500–502) and the elliptical conditional ‘if only’-type (503–505), but also deontic assertions like ‘should’ (506, 507),<sup>121</sup> including questions that are usually irrealis (508). However, *as* is also used in counterfactuals (509). *As* is followed by the finite verb usually in present or aorist tense (the latter fosters a more counterfactual reading; see also Sitaridou 2014a)<sup>122</sup>, but it also occurs with *ixa* ‘had’ complemented by the infinitive (Sitaridou 2014a: 136).

- (500) *elate as troyume*  
come.IMP.2PL OPT eat.1PL  
‘Come, let us eat.’ (A1)
- (501) *as derume ndo inumist*  
‘Let us see what will be.’ (03\_30062019F\_8; 05)
- (502) *an thelide as bam etši janina doboxu me terem=a tš=erxumis*  
‘When you like we could go over there, we look through the binoculars and return.’  
(08\_04072019M\_3; 144)
- (503) *alo mia as elebame denandalo*  
‘If only we could see each other once more.’ (01\_15022015F\_1; 21–22)

<sup>120</sup> The particle *as* is referred to as optative particle here as this term roughly covers the most important of the particle’s functions. However, since the optative has a specific meaning in the Greek grammatical tradition, exhortative particle might be the more precise term.

<sup>121</sup> Drettas (1997: 308) notes for PG that the difference between the imperative for the second persons and the optative is that the imperative expresses obligation while the optative is close to potentials. The continuum from desire to exhortative is gradual, though.

<sup>122</sup> Drettas (1997: 309) notes in PG a second optative (OPT<sub>2</sub>) formed by *as* + imperfective, which did not figure in the Romeyka data, though.

- (504) *as ixa nan bats-obon omo hatena*  
 ‘If only I had a little daughter like her.’ (03\_07072019F\_1; 14)
- (505) *as eleba ti sevda=m soyun do xrono mia*  
 ‘If only I had seen my beloved one at least once a year.’ [folksong] (04\_01072019F\_1; 070–071)
- (506) *eb-edode iba as prao so xoraf*  
 ‘Later I said, I should go to the field.’ (03\_30062019F\_11; 030)
- (507) *ap esena as mathan ade ba usta-ena as inede*  
 ‘She shall learn from you (so that) she shall become a master, too.’ (01\_02022015F\_1; 04–05)
- (508) *as arajev adina*  
 ‘Shall I call her?’ (04\_01072019F\_17; 68)
- (509) *as eftayo=se faji, ama utše na troyo ipes me.*  
 OPT make.1SG=2SG food but NEG PRT eat.1SG say.2SG 1SG  
 ‘I would have made you food, but you said to me you wouldn’t eat it.’ (A1)

The particle *an* is used for conditionals, both nonveridical and antiveridical (counterfactual). *An* is always followed by a finite verb, either in present tense (510–512) or in the imperfective (513). In both tenses, assertions are usually irrealis. Counterfactual conditionals are formed by *an* + auxiliary *ixa* + non-finite verb/inflected infinitive. If negated, the negation particle is placed immediately before the verb, preceded by the conditional particle (514). In all modal contexts, *an* seems to be used interchangeably with *na* (also cf. Drettas 1997: 326 on PG); it is not clear what determines their distribution. For a detailed discussion and the syntax of conditional clauses see Section 5.3.3.2.

- (510) *an roizune dži na futšiyundane*  
 ‘When they fall into it, they will drown.’ (04\_01072019F\_13; 05)
- (511) *eyo an bayo eksu roizo so nero*  
 ‘When I go outside, I will fall into the water.’ (04\_01072019F\_13; 22–23)
- (512) *an θelide as bam etši*  
 ‘If you like, we can go there.’ (08\_04072019M\_3; 189–190)
- (513) *müslüman an esane tibu tš=aleyane=mase*  
 ‘If they were muslims, they cannot say anything (against) us.’ (04\_01072019F\_2; 292–293)
- (514) *an dži θelide, zamani andž üdž ešide..*  
 ‘When you don’t like, when you don’t have time..’ (08\_04072019M\_4; 05–06)

Note that there is a certain variability in the choice of modal particles in Romeyka: not only appears *na* in the same conditional contexts as *an* (Section 5.3.3.2; see also Fn. 118 above), but there seems to be also variation between *na* and *as* in optatives (515; also cf. Neocleous 2020: 52). As an explanation, Dobrushina, van der Auwera & Goussev (2013) state that the optative might be related to the subjunctive, if the subjunctive is associated with non-actuality which

brings it closer to notions like “irrealis” and “potentialis”. The fact, that so far in Romeyka no variability is detected between *an* and *as* suggests that indeed *na* is the particle that is productive in taking over multiple modal functions; however, there cannot be made any assumption here as to the diachronic developments; further research is required, also comparing developments in other modern Greek varieties.

- (515) *auros eba gözledev d=opsaræ as deri.. na deri argos erd edži mi*  
 ‘The man watches the fish. Let him see, he watches whether the bear came.’  
 (04\_01072019F\_12; 40)

### 3.2.4.5 Discourse particles

#### 3.2.4.5.1 Topic and focus particles

The topicalizer<sup>123</sup> *ba(l)* (< AG adverb *pal(in)* ‘back, again’, Papadopoulos 1961: 130) occurs in the Romeyka corpus with a certain frequency (170 tokens). Its phonological form varies between *ba* and *bal* according to whether the particle is followed by a consonant or vowel in the onset of the following word, although this is not consistent (see, e.g., 521, 536). Following NPs, it needs to be ascertained whether *ba(l)* has potential cliticness; at least, it seems not to impact upon word stress of the head; further research is in order. *Ba(l)* occurs with (i) NPs (516) or individual nouns (517), pronouns (518), quantifiers (519, 520), and adjectives (521, 522);<sup>124</sup> (ii) adverbs, especially adverbs of time (523–525); (iii) VPs (526–528) and even subordinate clauses (529, 530). When topicalizing a NP, *ba(l)* is usually post-posed referring back to the nominal head; when topicalizing a VP, the particle occurs in phrase-initial position and has scope over the following VP (526, 537), probably paying tribute to its adverbial origin. When topicalizing an adverb, *ba(l)* follows the adverb and precedes the VP.

*Ba(l)* can be considered a topic particle at discourse level. It has the following functions: (a) it indicates a change of subject in a sequence of clauses (518); (b) it puts emphasis on a word, e.g., an adverb (523), or emphasises a fronted object (531); (c) it is used to express contrastivity (532); and (d) it functions as enumerating connective (533–535); its meanings including ‘also, too’ (536), ‘again’ (525, 528), ‘nevertheless’ (527) or ‘even’ (530).<sup>125</sup> Note that according to Neocleous (2020: 120), *pal* can only assign contrastive topichood but not aboutness topichood.

Interestingly, both meaning and functions of *ba(l)* overlap with that of the Turkish clitic *dA*. Göksel & Kerslake (2005: 101) describe the functions of *dA* as follows: “The clitic *dA* is a conjunction and discourse connective with additive, adversative, continuative/topic-shifting and enumerating functions. It occurs after stressed constituents, except when it functions as a continuative/topic-shifting connective.” Although this potential interesting candidate for “pattern replication” (Matras & Sakel 2007) cannot be explored here in more depth, the correlation between Romeyka *ba(l)* and Turkish *dA* should be notable (note also the occasional “matter borrowing” of *dA* in Romeyka of bilinguals, see Section 3.2.4.5.2). Haig (2017: 407) suggests that the use of post-posed additive features as candidate feature for a Pan-Anatolian linguistic area; the crucial question for assessing a contact explanation of *ba(l)* in Romeyka

<sup>123</sup> The exact functions of *ba(l)* in Romeyka are still to be investigated. Neocleous (2020: 120) considers *pa(l)* a “topic particle”, and Drettas (1997) defines *pa* in PG as a “contrastive topic marker”. Sitaridou & Kaltsa (2014: 19) argue that *pa* has no topic status but is rather a contrastive marker. So, the present terminology is tentative and awaiting closer scrutiny.

<sup>124</sup> Considering ex. (xiv), which is subject to intense code-switching, *ba(l)* also topicalizes interrogatives.

(xiv) *do ba d=efte detšinuz-un ešeklux-i*  
 ‘Why did you do their stupidity?’ (04\_01072019F\_2; 211)

<sup>125</sup> For the functions of *ba(l)* in PG, see Drettas (1997: 434–448).

being thus whether it occurs in a similar range of functions in other varieties of Greek outside of Anatolia.

If mentioning the functional similarities of *ba(l)* and Tr. *dA*, one cannot go without noting that the coordinating particle *tše* in Romeyka has also similar functions; although it is not clear whether a discourse marker *tše* needs to be analysed as distinct from the coordinating conjunction *tše* (see Section 3.2.6.1.1).

- (516) *ade si yoryora edune ha andra ades **bal** exaθe hare*  
 ‘She was in Gorgoras, her husband has died (by) now.’ (01\_04022016F\_1; 057–058)
- (517) *ta yarðela **bal** sahep tšikmadi*  
 ‘He didn’t adopt the children.’ (01\_06042017F\_4; 140)
- (518) *o argo t=opsarae ebire efaen adin **ba** nunizune leyone to.. harin nde n=eftame*  
 ‘The bear took the fish and ate (them). They wait, they say, “What will we do now?”.’  
 (05\_03072019M\_4; 04–05)
- (519) *ul **pal** eyapesame aet asen*  
 ‘We all wanted it like this.’ (01\_06042017F\_4; 067–068)
- (520) *pola **bal** tšešida exorepsane*  
 ‘They danced many kinds [of dances].’ (01\_06042017F\_4; 008–009)
- (521) *heralda kalabaluk **ba** ejendune*  
 ‘Apparently, it was crowded.’ (03\_30062019F\_6; 15)
- (522) *xastas **bal** emune*  
 ‘I was sick.’ (01\_28062019F\_2; 12)
- (523) *elioya **bal** utš eθelena n=andriza*  
 ‘I did not want to marry early.’ (02\_02022015F\_1; 013–014)
- (524) *exorepsame epištera **bal** eh.. kalabalik edune*  
 ‘We danced, later it became crowded.’ (01\_06042017F\_4; 007)
- (525) *erθame t=akšam **bal** aini hesab epigam=ata*  
 ‘We came in the evening and we made again the same.’ lit. ‘We did them in the same account.’ (03\_30062019F\_2; 37–38)
- (526) ***ba** na bam arajevom do ðromo*  
 ‘We went to seek the way.’ (08\_04072019M\_1; 179)
- (527) *ama jela erode=me **bal** kad exo apes=am*  
 ‘But come, ask me, I have nevertheless something within me.’ (03\_30062019F\_7; 35)
- (528) *erθame sabaxdan **bal** elmeksam dže esirame dže **bal** ebigame=da **bal** diri ebigame džureg*  
 ‘We came, in the morning we milked again, we run [lit. ‘pulled’] (the machine) and made cheese again, we made çörek cheese.’ (03\_30062019F\_2; 13)

- (529) *ha e paranta utš iš itšen ba u=boris n=epejnis*  
 ‘If you don't have money, you can't buy it.’ (02\_05072019F\_1; 34)
- (530) *hane ađio para pa..*  
 ‘even without money..’ (08\_04072019M\_1; 310)
- (531) *esena bal ade bola eyaba*  
 ‘She loves you much.’ (01\_14012015F\_1; 2)
- (532) *esi ba jolaepso=me eyo ba na jolaevo=se*  
 ‘Do send to me, I will (also) send to you.’ (01\_15022015F\_1; 26)
- (533) *trana ba obsarae eplezame bola ba*  
 ‘We caught as well big as many fish.’ (08\_04072019M\_1; 159)
- (534) *pola arapaðes aspra pal ine kotšina pal ine*  
 many.NOM cars.NOM white.NOM TOP be.3PL red.NOM TOP be.3PL  
 ‘Many cars are both white and red.’ (Neocleous 2020: 54, ex. 46e; presentation/glossing adapted)
- (535) *emorfon bal guvetin yarðelin emune*  
 ‘I was a handsome and strong child.’ (02\_21042018M\_2; 15)
- (536) *demo i batsi ba ena dek eše*  
 ‘My daughter also has one [=child].’ (01\_04022016F\_1; 085)
- (537) *bal kadi leyon ta peðia baška ta patsiðaes baška*  
 ‘They say something; the sons are different, the daughters are different.’  
 (03\_30092019F\_7; 45)

Finally, Drettas (1997: 465–466) describes for PG the use of a focalization particle =*ki* (i.e., =*tši* in Romeyka) which functions as an enclitic anaphora to a second or third agent (see also Sitaridou & Kaltsa 2014: 22). According to Papadopoulos (1958), the form might go back to the spatial relator *eki* ‘there’, which can from a diachronic point of view not be ruled out, although Drettas (1997: 466) also mentions that =*ki* could go back to contact with a postposed Georgian semantic intensifier *k'i*. Indeed, in the Romeyka corpus certain examples with a sort of “resumptive” *tši* occur that could be related to a spatial meaning (538–540). Ultimately, it is not clear whether the functions of the coordinating conjunction *tše* mentioned above are linked to the phenomenon described by Drettas (1997) for *tši* (541); further research is required here.

- (538) *to zo so katroi na ðena tši*  
 ‘I will tie the animal at the stable.’ (A1)
- (539) *habaðadžega as=emetero din džami buga gaheves ine tši*  
 ‘Here, below our mosque, there is a café.’ (08\_04072019M\_3; 157–159)
- (540) *ama aslinda sa komata n=ebejn edži*  
 ‘But actually, it should come till the open spaces.’ (08\_04072019M\_3; 142)
- (541) *ate ti ðulia=tis epitše tši hare tšimate tše kahete*  
 ‘She has done her work there?/and? is sleeping now.’ (A1)

## 3.2.4.5.2 Other discourse markers

Many prominent discourse markers in Romeyka are borrowed in form and function from Turkish. They will be briefly characterised below. Before presenting the borrowed discourse markers and connectives, though, it has to be mentioned that Romeyka seems to have preserved the AG postpositive connective discourse marker *γάρ*, /gar/, which is used to indicate how one part of a text correlates to the larger discourse of a story (van Emde Boas et al. 2019: 668–670), in the form of the adverbial *har* ‘now’. *Har* occurs in contexts beyond its general temporal adverbial function (542, 543), although further research is needed here to confirm the initial parallels. Furthermore, it should be noted that the Turkish adverbial *şimdi* ‘now’ seems to fulfil similar functions in discourse (see ex. 552).

(542) *har sabale ibe n=arxume*  
 ‘She said, “Tomorrow I will come.”’ (07\_04072019F\_5; 30)

(543) *ama leyo=se ama har bal ejendu s ena šimu ta xorafae ula ha eθedž ega*  
 ‘But I tell you, (now) it rained a lot, the fields all collapsed.’ (03\_30062019F\_6; 54)

*Ja* used in Romeyka goes back to the Turkish clitic *ya* (see Göksel & Kerslake 2005: 105) with a rough meaning of ‘surely, indeed, you see’. Like in Turkish, clause-final *ja* is used for facts which are common knowledge or obvious to both interlocutors (544, 545; “reminding discourse connective” in Göksel & Kerslake 2005: 105). In clause-initial position, *ja* poses an alternative (546).

(544) *lastiga=m endane ena tšamur ja*  
 ‘My shoes became full of mud (, you know).’ (03\_30062019F\_11; 039)

(545) *so dolabin do ine mono džidin ebsindane džüngü marendane ja so dolab*  
 ‘Those who are at the repository cooked a bit difficult, because they got (obviously) pale at the repository.’ (03\_30062019F\_11; 090–091)

(546) *ja arajepson adina*  
 ‘So, call her!’ (04\_01072019F\_17; 67)

*Jane*, or less frequently *hane*, (< Tr. *yanı* ‘i.e., that is to say, namely’ (Nemeth 1962/2020: 110; Göksel & Kerslake 2005: 451)) is used for explanatory purposes, e.g., to further elaborate on a just stated point (547, 548) or to conclude (549, 550). It occurs usually in phrase-initial position, although it may also occur clause-finally.

(547) *jane habaladžega obi bai kalo utš eftej jane aboro efte*  
 ‘Well, the one who departs from here doesn’t make it good, well, he does it bad.’ (08\_04072019M\_1; 301–303)

(548) *to mektebi ađatšeka xujepsame jane hađa so mehele jok tše so merkes ebejnam tš=exumestine*  
 ‘We went to school here, well, in this neighbourhood there was no school, we went back and forth to the center.’ (08\_04072019M\_1; 013–014)

(549) *jane emistine haets eđevasam da imeras*  
 ‘Well, we spent the days like this.’ (08\_04072019M\_1; 045–046)



- (550) *ne **jane** ena imera andža kahuvane son baxari*  
 ‘Yes, well, on a single day, however, they left to the pasture.’ (08\_04072019M\_1; 107–108)

*Išd(e)* (< Tr. *işte* ‘see’, an “organizational discourse connective” linking previously mentioned items to present a statement, Göksel & Kerslake 2005: 455) is used to specify new information that has been signalized before (551, 552) or to indicate that a certain topic has come to an end (553). Due to its co-referential/linking function, it seems to be able to appear in all clause positions, usually preceding the phrase or clause that is elaborated on (cf. 552).

- (551) ***ište** eb ordan ebedžeha etšeli edžes*  
 ‘See, from there, from there he turned over there.’ (03\_30062019F\_8; 22–23)
- (552) *ne šimdi emis ba ađatšes **ište** s=opsarema bejname bola*  
 ‘Yes, we went fishing a lot here.’ (08\_04072019M\_1; 138–139)
- (553) *efiga dže be ađaha katsa ga **išd***  
 ‘I left it and sat down here, see.’ (03\_30062019F\_6; 37)

*Nejse* (< Tr. *neyse* ‘anyway, never mind’ (see Nemeth 1962/2020: 106, Göksel & Kerslake 2005: 455, 457)) is used to express letting go of an (often negative) fact or to end a conversation topic (554).

- (554) *bori=nata jani **nejse***  
 ‘She cannot do these, never mind.’ (07\_04072019F\_5; 28–29)

*Zate* (< Tr. *zaten* ‘in any case, well’, Göksel & Kerslake 2005: 456) is used to make a previous statement more predictable by corroborating or overriding it (also cf. Drettas 1997: 422 for PG). It occurs in initial (555) or final position (556).

- (555) ***sate** ula da đomea jaja bename todjes*  
 ‘In any case, we went all our ways by foot back then.’ (08\_04072019M\_2; 130)
- (556) *evradžes bola **sade***  
 ‘Did you got wet, in any case?’ (04\_01072019F\_5; 14)

Finally, the Turkish clitic *dA* has a variety of discourse connective functions (see Göksel & Kerslake 2005: 101) and can mean ‘also’, ‘and’, ‘moreover’, ‘again’, ‘even’, ‘although’ (Nemeth 1962/2020: 94). It is furthermore used as a conjunction, in which function it also occurs in Romeyka (see Section 3.2.6.1) In Romeyka, discourse connective *dA* occurs in adversative (557) or continuative function (558), although often only in conjunction with other borrowed words or in code-switching. Interestingly, although the functions of Turkish *dA* and the Romeyka topicalizer *ba(l)* often overlap (see Section 3.2.4.5.1), the contexts in which *da* is borrowed cannot be occupied by *ba(l)*.<sup>126</sup> Note, however, also the use of the conjunction *tše* ‘and’ as additive particle which resembles the use of Turkish *dA* (Section 3.2.6.1.1).

<sup>126</sup> There is another particle that occurs as a hapax legomenon in the position of *ba(l)*, namely, the form of address *eba* ‘hey’ (xv = ex. 515, this chapter). It is not clear whether this is erroneous or conditioned by the phonological similarity.

(xv) *auros eba gözledev d=opsaræ*  
 ‘The man is watching the fish.’ (04\_01072019F\_12; 40)

- (557) *enan inega exo – jok da tesera*  
 ‘I have a wife.’ – ‘No, four!’ (06\_03072019M\_2; 16–17)
- (558) *ai bola ilos ebidžen belki da hadžan udž efd=a*  
 ‘There was [lit. ‘it made’] so much sun, maybe that is why it? didn’t do it.’  
 (03\_30062019F\_6; 47–48)

### 3.2.5 Prepositions

Romeyka is a prepositional language (like SMG, and unlike Turkish, Dryer 2013e);<sup>127</sup> the syntax of a prepositional phrase is P–NP. For the most frequent prepositions and their function see Table 18, examples for the prepositions listed in Table 18 are presented in exs. (559)–(567) below. Some prepositions like *ek-* ‘out’, *an* ‘up’, *kata* ‘down’ merely exist in compound forms, for example, in spatial adverbials or compound verbs (Section 4.1.1.4). There is uncertainty about whether some spatial adverbials may be considered postpositions, namely, *(e)ban* ‘above’, *(e)buga* ‘under’, *embro* ‘before/in front of’, *opis* ‘behind’, *eperan* ‘opposite’; consider (for AMG) the distinction in Karatsareas (2016: 62, Table 5 and 6) between (primary) prepositions and (secondary) postpositions, the latter correspond to the multisyllabic forms that are treated here as spatial adverbials in Section 3.1.4.4.1. Some prepositions like *os* ‘until’ and *ðae* ‘for (the sake of)’ also function as subordinating conjunctions; they are discussed in Section 3.2.6.2. alongside other subordinating conjunctions like *prin* ‘before’.<sup>128</sup> While *omon* goes back to an adverb, it is used both as preposition and subordinating conjunction.

**Table 18:** Prepositions in Romeyka

Preposition	Function	Translation equivalent
<i>s</i>	locative	in, at, to
<i>as</i>	ablative	from, of, by, from ... on, since
<i>ap</i>	source	from
<i>os(t(on))</i> <sup>129</sup>	temporal, locative	until
<i>ja(t)</i> <sup>130</sup>	purpose	for (RSür, ROf)
<i>ðae</i> <sup>131</sup>	purpose	for (the sake of)
<i>me(t)</i>	comitative	with
<i>me(t)</i>	instrumental	with, through, by
<i>aðio</i>	privative	without
<i>omo(n)</i>	comparison	like

<sup>127</sup> The Romeyka corpus features a single occurrence of a post-posed preposition (xvi).

(xvi) *an ðelide as bam etši janina doboxu me terem=a tš=erxumis*  
 ‘If you like, let us go there, we watch through the binocular and return.’ (08\_04072019M\_3; 144)

<sup>128</sup> *prin* ‘before’ can be also used as a preposition, e.g., *prin kantria xrone* ‘a few years ago’ (Tursun 2019: 254).

<sup>129</sup> It is not fully clear whether the long form *os-ton* is actually part of the preposition (< AG *ós* ‘to’ + DET.ACC, van Emde Boas et al. 2019: 397) or rather a (reanalysis of the) definite article *ton*.

<sup>130</sup> Both forms *ja(t)* and *ðae* go back to the AG preposition *ðia* ‘through, by means of, because of’ (M. Janse, p.c.); while *για* is widespread in other (synchronic) varieties of Greek (including PG), the fact that the /ð/ of *ðae/ðe* has survived in ROF as spoken in Caykara is remarkable (P. Mackridge, p.c.).

<sup>131</sup> Sitaridou (2014b: 30) lists *ðae* as retention of the AG preposition *ðia* for ROF as spoken in Çaykara (see also previous Fn. 130). The preposition is not attested in the present corpus (neither in Tursun 2019 with speakers from Saráchos (Uzungöl) and Ödžena (Karaçam)); in the examples from Sitaridou (2014b), however, *ðae* seems to function more like an adverbial subordinator rather than a preposition; therefore, see also Section 3.2.6.2.

- (559) *asa id eftayo na fai*  
 from.the 3PL make.1SG INDET bread  
 ‘I make a bread from them.’ (03\_30062019F\_6; 62)
- (560) *eθaresa abemetera so xorio isen*  
 ‘I thought you are from our village.’ (01\_15022015F\_1; 10)
- (561) *edodes džumartesi oston oile ba xujevame*  
 ‘In these times, we studied on Saturday till noon.’ (08\_04072019M\_1; 030–031)
- (562) *os pudžega pijede*  
 ‘Until where did you go?’ (04\_01072019F\_17; 33)
- (563) *oston evle tipu tše na troyo*  
 ‘I will not eat anything till noon.’ (C1)
- (564) *ja t=apið epejnam ekrivame s=oros*  
 ‘For the pears, we (went and) hid at the forest.’ (04\_01072019F\_2; 289)
- (565) *sade jad emena faji utš eθelisa ps.. pseθinimo*  
 ‘I do not want to cook just for myself.’ (C1)
- (566) *hane ađio para pa*  
 ‘even without money’ (08\_04072019M\_1; 310)
- (567) *ta muskara omo yarðelja tero*  
 ‘I care for the calves like for children.’ (08\_04072019M\_1; 234)

The local prepositions *s* ‘to, in’, *as* ‘from’, and *ap* ‘from’ merge with the definite articles and possessive pronouns (see Section 2.3.5.3. on crasis);<sup>132</sup> *s* + *to* > *s=o* ‘in the’,<sup>133</sup> *as* + *tin* > *sin* ‘from the’, *s* + *temetera* > *s=emetera* ‘to us’, *as* + *ta* > *as=a* ‘from the’ (see ex. 559), *ap* + *temetero* > *ab=emetero* ‘from us’ (cf. 560). *Me* and *ja* become *met* (574) and *jat* (565) [ \_V ].<sup>134</sup> The preposition *ap* also appears as *ep* as a product of assimilation to a following front close-mid unrounded vowel (e.g., 03\_30062019F\_11; 067). A sort of distant assimilation in labial environments also occurs with the vowel of the preposition *me(t)* ‘with’ which shows some variability according to the vowel of the following word, i.e., either the definite article or a pronoun (568; see also Section 2.3.2). While the preposition occurs in the majority of cases with the vowel /e/, sometimes the vowel seems to vary according to the feature [±front/back], although with considerable deviation. A tentative pattern is provided in (569). This distant assimilation between the preposition and the following word resembles Turkish vowel

<sup>132</sup> Apparently, also non-contracted locations are possible (xvii).

(xvii) *as to pislukinen eðevam eka*  
 ‘We suffered from the pissing.’ (02\_09062019F\_1; 08)

<sup>133</sup> But cf. the rare forms *sto* (xviii) and *tso* (xix).

(xviii) *temetero tso xorio bola mejveðes ine*  
 ‘At our village, there were many fruit trees.’ (02\_02022015F\_1; 180–181)

(xix) *sto ðormo so spidi bai*  
 ‘(At) the road leads to the house.’ (translation unclear; 04\_01072019F\_12; 05)

<sup>134</sup> Since *me(t)* goes back to AG *meta* ‘together with/ accompanied by’, retention of the /t/ in some contexts is likely (M. Janse, p.c.). However, this explanation does not account for the ‘filler’ /t/ in *ja*, although it might be caused by analogy with *me(t)*. Note that the form *meta* appears in one example of the corpus (08\_04072019M\_1; 233) but its analysis remains unclear.

harmony, although in Turkish, the vowel of the clitic suffix is variable and not that of the stem. Like with the variability in the word *tširi=m* ‘father’ (Section 2.3.2), the “stem” variability in Romeyka reminds one also of an Umlaut phenomenon.

- (568) *bazen **mi** di manam, bazen **mo** do džirim*  
 ‘sometimes with my mother, sometimes with my father’ (04\_01072019F\_2; 075–076)
- (569) o / [ \_to, ta]                      e.g., (570a/b), but cf. *motinan* ‘with whom’  
 a / [ \_ta]                                e.g., (570c)  
 i / [ \_t(s)i]                              e.g., (570d)  
 e / [ \_ti., te., ta]                      e.g., (570e/f), but (570g/h)
- (570) a. *mo don džiri=m*  
 ‘with the father’ (04\_01072019F\_2; 067)  
 b. *ula mo da šeri=una mo da rašes=muna*  
 ‘everything with our hands, with our backs’ (04\_01072019F\_1; 200)  
 c. *ula ma da makinās*  
 ‘everything with the machines’ (04\_01072019F\_2; 243)  
 d. *mi tsi aulus udž epikan=a*  
 ‘We had not done it with the men.’ (01\_06042017F\_4; 068)  
 e. *ekoftame me ti kerenti*  
 ‘We cut with the scythe.’ (09\_04072019\_7; 11)  
 f. *me ta sturatša*  
 ‘with the sticks’ (03\_30062019F\_7; 16)  
 g. *me to sirti=muna*  
 ‘with our pannier’ (08\_04072019M\_2; 009)  
 h. *da kadas me do limo ini*  
 ‘The cats are hungry.’ (03\_30062019F\_11; 068–069)

Most of the prepositions take accusative case, e.g., *s*, *as*, *ap*, *me* (Özkan 2013: 145; but cf. the genitive construction in 571). But due to the syncretism of nominal inflection, the accusative case is merely visible in the feminine (572)/masculine (573) singular of nouns and in the use of object pronouns (574).

- (571) ***su** deirmen to xark erθe*  
 ‘He came to the water canal of the mill.’ (03\_30062019F\_8; 24)
- (572) *emistine ađa **sin** odžena jašaevume*  
 ‘We live here in Ogene.’ (08\_04072019M\_1; 003–004)
- (573) ***ason** barxari asi mazira me t=aleyo ta yomarae=m eferena*  
 ‘From the pasture, from the village square I carried my loads with the horse.’  
 (08\_04072019M\_2; 136–137)
- (574) *ipen eyo na fero=se **met esena***  
 ‘He said, “I will bring you (with you?).”’ (translation unclear; 02\_21042018M\_2; 24)

Note that like it is the case with most prepositions, the preposition *me* is used for several functions including comitative (575, 576) and instrumental (577, 578; additional functions are exemplified in 579–585). The use of a single preposition for both comitative and instrumental applies also to the equivalent Turkish postposition *ile/-(y)la* (Göksel & Kerslake 2005: 67).

- (575) *mijes ebiya s=opsarema habađadžega **me** di mim don beđa*  
 ‘Once we went fishing down here, with the son of my uncle.’ (08\_04072019M\_1; 152–154)
- (576) *embro ebejname **me** ta za*  
 ‘In earlier times, we went with the cows.’ (08\_04072019M\_3; 082)
- (577) *har ula **me** din araba*  
 ‘Now everything [is done] by car.’ (08\_04072019M\_1; 092)
- (578) *ta kabanae **me** to tšagudža siθađo*  
 ‘I straighten<sup>?</sup> the walls with the ?.’ (translation unclear; 08\_04072019M\_3; 047)
- (579) *ulin **me** ta satšia eberenane t=alevre*  
 ‘Everybody bought flour with (i.e., in amounts of) a sack.’ (08\_04072019M\_2; 152–153)
- (580) *jela ađelfi=m ela **med** emena*  
 ‘Come, my sister, come to me.’ (01\_15022015F\_1; 18–19)
- (581) *anda utš en **mo** to limo beftenam eka*  
 ‘When there was nothing, we went to bed hungry.’ (02\_02022015F\_1; 118–119)
- (582) *ejendumune nife **med** etšino orašefta*  
 ‘I became a wife, I dealt with it.’ (02\_02022015F\_1; 016–017)
- (583) *jane **me** so borbade ma deduum [...]*  
 ‘Having said going for a walk, [...].’ (08\_04072019M\_3; 107)
- (584) ***met** o đromo na drežis na gadivenis utš saatluk đromo*  
 ‘If you go [lit. ‘run’] with the road, you will descend a three-hour way.’ (01\_28062019F\_3; 57–59)
- (585) ***me** din kaza biren=ana*  
 ‘He took him with [i.e., by means of] the accident.’ (03\_30062019F\_8; 11–12)

All prepositions in Romeyka are inherited although in code-switching in ex. (586), a Turkish PP is inserted.

- (586) *ode od efdas šaka **itšun** jela arkadaši=s*  
 ‘About what you do in jest, your friend laughs.’ (02\_02022015F\_1; 147–148)

If a preposition governs a NP, the preposition occurs only before the first element of the NP (587), although the preposition can apparently be doubled before a pronoun or quantifier and the head noun (588). Possibly, the latter is a sign of attrition (see also the position of the preposition only before the head noun in (xviii) in Fn. 133, instead of before the pronoun). If a spatial adverbial is involved, apparently no local preposition is required (589).

- (587) *s=emetero to mehele mektebi utš edon*  
 ‘At our neighbourhood there was no school.’ (08\_04072019M\_1; 024–025)

- (588) *s ula sa mejvedes enivename*  
 ‘We climbed on all trees.’ (02\_02022015F\_1; 182)

- (589) *n=epinam=a ta za buka*  
 ‘We put it under the cows.’ (04\_01072019F\_1; 170)

As for the syntax of prepositional phrases, their position in the clause is variable and linked to information structure: Neocleous (2020: 148) states that pre-verbal PPs are in focus position. In the Romeyka corpus, the majority of PPs occur in preverbal position (590, but cf. 575, 576 above), so it is not straightforward to determine the “unmarked” word order of PPs (see Section 5.2.1.1). The WOWA dataset (Schreiber 2021) at least suggests that the speaker is a decisive variable, indicating that word order may depend possibly upon the linguistic profile of the speaker. Furthermore, according to the WOWA dataset, 47% of obliques like beneficiaries, instruments, and comitatives are post-predicate and 42% of locations. However, interestingly, 78% of goals appear to be post-predicate, in line with a general tendency noted for the languages of eastern Anatolia, though Armenian may be an exception. It needs to be noted, though, that these semantic categories are not always flagged with prepositions. The overall rate of prepositional marking in the WOWA dataset is 61%. Several PPs may be included within a single clause (590); their individual position is variable according to what is stated above.

- (590) *ason barxari asi mazira me t=aleyo ta yomarae=m eferena*  
 ‘From the pasture, from the village square, I carried my loads with the horse.’  
 (08\_04072019M\_2; 136–137)

### 3.2.6 Conjunctions

Conjunctions are grammatical particles that function as linking elements at the level of the NP, VP, or clause. Their function can be separated into coordinating (Section 3.2.6.1) and subordinating (Section 3.2.6.2) conjunctions.

#### 3.2.6.1 Coordinating conjunctions

##### 3.2.6.1.1 Coordinating conjunction *tš(e)*

The most prominent coordinating conjunction is *tš(e)* ‘and’. Its functions are: (i) word, phrasal or clause coordination, (ii) coordination of finite verbs in double (or serial) verb constructions, (iii) a periphrastic progressive formed by a (coordinated) double word construction, (iv) as a complementizer in potentials, (v) as a focalizing discourse connective (like *ba(l)*).<sup>135</sup> The conjunction occurs in the form *tše* before a vowel in the onset of the following word; before consonants, the form is *tš*. The functions of *tš(e)* are outlined below.

(i) *Tš(e)* is used for coordination of content words (591, 592), phrases (NP (593, 594), VP (595), PP), and clauses (596). A zero strategy for coordination is allowed at phrase (597) and clause level (Section 5.3.1.1; also cf. Drettas 1997: 430).

- (591) *resmije tš=esi*  
 ‘Resmiye and you’ (03\_07072019F\_1; 16)

- (592) *exbase tše xase*  
 ‘ruptured and lost’ (02\_05072019F\_1; 22)

<sup>135</sup> Note also the use of pre-posed *tš(e)* in PG comparatives as *tš=alo* ‘once again’ (Drettas 1997: 156, §191), which is not (yet) attested for Romeyka.

- (593) *epejna perena da xuljera se do tasi*  
 ‘I went and took the spoons and the cup.’ (07\_04072019F\_6; 42)
- (594) *havudo do ješiluxi tš=avudo kalo do hava itšiega butšega ne vriškis=a*  
 ‘This green and this good air, where do you find them there?’ (08\_04072019M\_3; 123–125)
- (595) *etroede dž=eskustini*  
 ‘You ate and got up.’ (03\_07072019F\_1; 27)
- (596) *fordonum din araba tše ferume son barxari*  
 ‘We load the car and bring [it] to the pasture.’ (04\_01072019F\_1; 021)
- (597) *ebejna tin pođea=m do tasi=m*  
 ‘I went [with] my apron and my cup.’ (07\_04072019F\_6; 50)

(ii) A special case of VP coordination are serial verb constructions coordinated by *tše* (for a detailed discussion see Section 5.3.2). *Tš(e)* links at least the final two finite verb forms (598, but cf. 599). A special type of double verb construction frequently involves a finite form of the motion verbs *pao* ‘go’ (600) or *erxume* ‘come’ (601, also 599) that is attached to another VP by means of *tše*. Both verbs involved show the same inflection with regard to person and tense (see, e.g., 605a). The meaning of these constructions is still more or less linked to a motion but considering the frequency of the construction and the option of semantic bleaching, this construction appears to be partially grammaticalized (cf. ex. 602). It should be noted, however, that Neocleous (2020) considers the focus particle *tše* to be different from the coordinating conjunction *tše*; further research is required here. In (602), the double verb construction appears in initial position although it is otherwise almost always clause final. The double verb construction is also possible with a variety of non-motion verbs (e.g., frequently *fero* ‘take’, 603a, and *eftayo* ‘make’, 604). It is argued that this construction is a calque of the Turkish *-(y)Ip* converb which links two verbs with a meaning of simultaneity, and which reflects also in the Turkish translation of the examples (603b, 605b). Alternatively, the construction can also translate as the Turkish *-ArAk* participle, e.g., *fiso tše troyo*, Tr. *üfleyerek yiyorum* ‘I eat blowingly’ (Tursun 2019: 212). Importantly, some of the examples (e.g., ex. 600) suggest that this serial verb construction is not only mirrored on Turkish *-(y)Ip* converbs but also on their specific use *-(y)Ip + dA* which expresses an adversative relation between both verbs (Göksel & Kerslake 2005: 439); see below for a discussion of the similarity of Tr. *dA* and Romeyka *tše*.

- (598) *ta xordare=mun do thelis bas koftjis tše feris*  
 ‘Our fields: you go, cut and bring what you like.’ (08\_04072019M\_1; 219)
- (599) *da xordarae kofteme tše eberum tš=erxumis*  
 ‘We cut the meadows, we take [it] and return.’ (08\_04072019M\_1; 200–201)
- (600) *hajese efikame tš=epiyame*  
 ‘We left [it] like this and we went.’ (04\_01072019F\_2; 323)
- (601) *manaxe ebenam tš=erxumis*  
 ‘We went and returned alone.’ (08\_04072019M\_1; 018)
- (602) *pao so raši ebero tše bayo ena demligi*  
 ‘I go to the mountain, and take a tea pot (with me).’ (08\_04072019M\_1; 264–265)

- (603) a. *me t=aleyo forduna dže ferina ađadžega*  
 b. *ata yükleyip getirirdim buraya*  
 ‘With the horse packed, I carried them here.’ (08\_04072019M\_2; 160)
- (604) *ta laxana layo košovume dž=eftame laxana ne*  
 ‘How do we cast and make cabbage dish, right?’ (03\_30062019F\_11; 081)
- (605) a. *am pao so raši n=epero liyo diri tše na fero*  
 b. *Dağlara gittiğim zaman, biraz peynir alıp götürüyorum.*  
 ‘When going to the mountains, I take a little cheese with me.’ (C1)

(iii) A double verb construction with *tše* + finite *steko* ‘stay’ or *kahume* ‘sit’ is used to express progressive or continuous aspect (Özkan 2013: 148; see Section 4.3.3.3). This is a calque of the Turkish *-(y)Ip* converb construction with the verb *durmak* ‘stay’ (606). While the construction in (607) is ambivalent as to its state of grammaticalization with the semantics of *steko* still transparent, the example in (608) shows a more grammaticalized progressive function with semantic bleaching.

- (606) *çalışıp duruyorum*  
 ‘I keep working.’ (Özkan 2013: 148)
- (607) *hae ul hae ospid exune tše stekune*  
 ‘Now, they all have houses (and they stay).’ (03\_30062019F\_7; 48–49)
- (608) *ae kolaieve da tše stetš ama eš erde argo*  
 ‘He is shielding them like this but the bear is coming.’ (04\_01072019F\_12; 19–20)

(iv) *Tše* is used as complementizer in (antiveridical) potentials with *boro* ‘can’ (609–611), which usually take the particle *na* (612); see Sections 4.3.6.5 on potentials and 5.3.2.1.3 on *na*-clauses.<sup>136</sup>

- (609) *ne faje borum tše droyum ne rahatje borum tš=eftjeme*  
 ‘We can neither eat bread, nor can we make it easily.’ (07\_04072019F\_6; 58–59)
- (610) *ne boro tš=erxume ne esi*  
 ‘Neither I can come, nor you.’ (01\_15022015F\_1; 23)
- (611) *tše ne borum tš=erxumesine*  
 ‘We cannot come.’ (07\_04072019F\_5; 17)
- (612) *ađahandžeka tip=u=borum n=eberume*  
 ‘We cannot buy anything here.’ (01\_28062019F\_3; 55)

(v) *Tše* appears postposed in NP coordinations (613–615), where it functions as a sort of discourse particle, albeit more detailed research is required here. According to Neocleous (2020: 100–101), *tše* functions as “additive particle” like SMG *ce* ‘also’ that has been classified as “focal associate operator”; it is thereby not clear, whether Neocleous (2020) distinguishes

<sup>136</sup> Note that Drettas (1997: 429–430) states that *tše* is in PG the obligatory complementizer of certain verbs like PG *arxino* ‘start’.



the coordinating conjunction *tše* from the “additive particle”.<sup>137</sup> From a phonological point of view, *tše* seems to cliticize with the preceding noun and does not carry any stress. It resembles the topicalizer *ba(l)* and yields a meaning of ‘also’.<sup>138</sup> However, it is striking that post-posed *tše* occurs frequently with numerals and Turkish numeral classifiers like *dane* ‘piece’ or *kere* ‘instance’ (615) but also in the enumeration in (616). Its use is not restricted to quantities, though (617).

- (613) *on dane đaxtila xo so šerim, on dane tše sa bođarem*  
 ‘I have ten fingers on my hand and also ten on my feet.’ (01\_28062019F\_4; 57–58)
- (614) *dört dane beš dane tše yarđele*  
 ‘four or five children’ (04\_01072019F\_2; 093)
- (615) *beš kere dört kere tše sin tšaikaran ebiyame*  
 ‘Four times, five times, we went to Çaykara.’ (02\_29062019F\_2; 09)
- (616) *mo ta zagoda tše mo do sorani tše*  
 ‘with Zagoda herbs and with onion’ (04\_01072019F\_1; 189)
- (617) *ndu dž=epika*  
 ‘What did I (not?) do?’ (01\_28062019F\_2; 10)

Moreover, postposed *tše* also appears phrase-finally in VPs (618–621) where it possibly puts emphasis or focus on the phrase (also Drettas 1997: 428–429, §700). However, its use with the subordination conjunction *omon* ‘like’ in (622) is intriguing. Neocleous (2020: 101) lists *tše* as additive focus particle in post-verbal position and together with a mass noun (623). In this position, post-verbal and post-nominal *tše* appear to be the same particle.

- (618) *kavurevune mono dže*  
 ‘They stir it a bit.’ (03\_30062019F\_11; 086)
- (619) *evraxa dže*  
 ‘I got wet.’ (04\_01072019F\_5; 15)
- (620) *osa da šindže as es.. asirame tš=etroyame tše o koliva*  
 ‘If there was now any, we could pull and eat boiled corn.’ (04\_01072019F\_2; 273)
- (621) *etšina ekovalesa tš=estivaksa tše espundžesa dže*  
 ‘I collected them and piled them up and swept.’ (01\_28062019F\_2; 08–09)
- (622) *omon dž=ebiyam oneon dodže šimu na drano šimu*  
 ‘When we went, suddenly? it rained, a big rain.’ (03\_30062019F\_6; 41)

<sup>137</sup> Michelioudakis & Sitaridou (2016) term *tše* a “spurious coordinator”. Note also the “adverbial” function of AG *kaí* (> R. *tše* ‘and’) which functions as addition/extension marker to indicate that a certain utterance also applies to the following phrase (van Emde Boas et al. 2019: 693). This includes meanings such as ‘also’ and ‘too’, roughly corresponding to the functions of the Turkish particle *da* (see discussion below in this section); although AG *kaí* seems to appear predominantly before the “added” phrase which forms a crucial test for the contact hypothesis.

<sup>138</sup> Note that Oikonomidis (1958) apparently already mentioned the similar functions of *ba(l)* and *tše* (in Drettas 1997: 435). For complementary occurrences of *ba(l)* and *tše* in PG, see Drettas (1997: 428).

- (623) *θénumen dže álas*  
 ‘We also put salt.’ (Neocleous 2020: 101, ex. 34)

The functions of postposed *tše* can be made evident in comparison with the functions of the equivalent conjunction *ce* ‘and’ in Cappadocian which is used in accordance with the functions of the Turkish discourse particle *dA* (624a/b; E. Daveloose, p.c.; cf. also Section 3.2.4.5.1 on *ba(l)*). Cappadocian *ce* is used, like Romeyka *tše*, (a) as (“proclitic”) connective, and in accordance with Turkish *dA*, (b) as post-posed scope particle (equivalent to additive function in Göksel & Kerslake 2005: 101), and (c) as topic shift marker. In line with this, it can be assumed that post-posed *tše* in Romeyka also functions as a discourse connective (very similar to *ba(l)*!; but cf. Drettas 1997: 429 for differences) in accordance with Turkish *dA*: ex. (616) above would be an example of the enumerating function, whereby like in Turkish (Göksel & Kerslake 2005: 101), *tše* is attached to each enumerated phrase; ex. (617) might be an example of the additive function, whereby like in Turkish the preceding syllable bears stress; exs. (622), (624) might be examples of *tše* as topic shift marker, whereby stress is (like in Turkish) on the following word. While it cannot be shown for all examples how they correspond to the functions of Turkish *dA*, it is likely that the use of the conjunction as discourse connective might be modelled on a Turkish strategy.

- (624) a. *t=ejenton tše exolesten tš=epien*  
 b. *Ne oldu da küsüp gitti?*  
 ‘What happened that he left angrily?’ (Tursun 2019: 216)

Finally, the Turkish particle *dA* is also used as conjunction *da* in Romeyka in different contexts at phrase and clause level which all mirror the Turkish functions: (a) coordinating in the Turkish expression *bir de* ‘(lit.) and one’ (625); (b) adversative (626); (c) as topic-shift marker (627); see Göksel & Kerslake (2005: 101).

- (625) *kurdž en bir da tšureg en*  
 ‘It is kurç and also çörek cheese.’ (03\_30062019F\_2; 14)
- (626) *asinda son karšimiz n=adon emorfa yonuševum da teke u=boro na yonuševum emorfa*  
 ‘Actually, if it was in between us, I speak nicely but alone I cannot speak nicely.’  
 (03\_07072019F\_1; 34–35)
- (627) *bir zamanlar so kulin da bejname [...]*  
 ‘Once when we went to school [...].’ (08\_04072019M\_1; 028–29)

### 3.2.6.1.2 Disjunctive conjunctions

Disjunction is solely expressed by loan words.<sup>139</sup> Most common is Turkish *(ve)jaxod(e)* ‘or’ (628, 629) which coordinates clauses. Other conjunctions also operating at phrase level are *(ve)ja* ‘or’ (630, 631) and *jada* ‘or’ (632). According to Göksel & Kerslake (2005: 445), these three conjunctions are used interchangeably in Turkish; the same seems to apply to Romeyka. However, while *ve yahut da* seems to be used less frequently in spoken Turkish; it occurs frequently in Romeyka, probably as a loan from an earlier stage, i.e., Ottoman Turkish. *Yoksa* ‘if not, or, otherwise’ is used for antiveridical disjunction at clause level (633–635; for the Turkish disjunctions see also Nemeth 1962/2020: 110).

<sup>139</sup> But note that Neocleous (2020: 177) presents in ex. 122 *i* as disjunctive particle ‘or’ but it is not clear whether this is actually attested from a speaker; *i* is not listed in this function in Tursun (2019).

- (628) *xandilaza=na jahoda ekuntena=na*  
‘I tickle her or I hug her.’ (02\_02022015F\_1; 144–145)
- (629) *vejaxute nejenene to zo epukan=eθe n=epiname fila*  
‘Or when a cow gives birth, we put leaves under it.’ (04\_01072019F\_1; 172–173)
- (630) *jane ta za to enan etšilizane veja ta dio tšilizane*  
‘One of the cows tumbled or two tumbled.’ (08\_04072019M\_1; 110–111)
- (631) *havu to peđin habađahandžeka eparen ja do hükümetin na diyum=a jaxota hao to peđin n=eperis*  
‘Take this child from here or we will give him into a children's home or you will take this child from here.’ (02\_21042018M\_2; 19–21)
- (632) *d=akšemi n=arde ište jada ja... belki na bai so xorio*  
‘He will come in the evening, see, or, maybe he will go to the village.’ (04\_01072019F\_5; 60)
- (633) *ina tina joksas mutš eyriga*  
‘[unclear translation] or does she not understand?’ (01\_28062019F\_3; 32–33)
- (634) *adžaba maksuz efte=a mi joksas mi*  
‘Does she actually do this intentionally or not?’ (01\_28062019F\_3; 30–31)
- (635) *os pudžega pijede dodtšes eperan joksas ebedžeberan erde bireza sinora o dursuni*  
‘How far did you go? Did you go over and back or from opposite until Dursun’s garden line?’ (04\_01072019F\_17; 33–34)

### 3.2.6.1.3 Adversative conjunctions

Adversative coordination is at clause level expressed by the Turkish loan *ama* ‘but’ (636–638).<sup>140</sup> Furthermore, Turkish *ancak* ‘yet’ is borrowed (639, 640). Note, however, that *andžak* in exs. (639) and (640) is used as a temporal adverbial meaning ‘only just’. Both functions are copied from Turkish including their syntactic properties: the adverbial occurs before the predicate while the conjunction links two phrases (see Göksel & Kerslake 2005: 205, 446).

- (636) *exorepsane ama i mana=θe mono merakliena dune*  
‘They danced but her mother was a bit anxious.’ (01\_06042017F\_4; 062–063)
- (637) *ama leyo=se ama har bal ejendu s ena šimu ta xoraḡae ula ha eθedž ega*  
‘But I tell you, now it rained a lot, the fields all collapsed.’ (03\_30062019F\_6; 54)
- (638) *ama so šeher erθame bola rahatluk eš*  
‘But when we came to the city, there was a lot of comfort.’ (09\_04072019\_7; 26)
- (639) *fazla magri zamani utš ešide na glošge andžak hazirlanjestestine*  
‘If you don’t have so much time, you will return and just get ready.’ (08\_04072019M\_4; 14–16)

<sup>140</sup> Note that in Cappadocian the same loan word is stressed *amá* (M. Janse, p.c.), which is not the case in Romeyka where the stress is more evenly distributed like in Turkish but tends to be more on the first syllable than on the second (see e.g., 08\_04072019M\_1; 066).

- (640) *ena imera **andža** kahuvane son baxari*  
 ‘Only just on a single day, they went up to the pasture.’ (08\_04072019M\_1; 107–108)

Example (641a/b) suggests that also the coordinating conjunction *tše* seems to be interpretable in adversative function, as the Turkish translation evidences.

- (641) a. *poli θelun=a tše is epere*  
 b. *Onu herkes ister ama birisi alır.*  
 ‘Everybody wants her but only one takes her.’ (Tursun 2019: 243)

#### 3.2.6.1.4 Correlative conjunctions

All correlative conjunctions are borrowed from Turkish: *ja ... ja* is used for ‘either ... or’ at word (642), phrase (643), and clause (644) level, *ne ... ne* for ‘neither ... nor’ at least at clause level (645–647), and *hem ... hem* for ‘as well as’ at least at clause level (648, 649).

- (642) *aso xorio ba fortununesine **ja** jaralmasija **ja** laxana **ja** haidika **ja** tsubađe*  
 ‘And from the village, we carried either potatoes or cabbage or such corn cobs.’  
 (04\_01072019F\_2; 230–233)
- (643) *n esexpanam eka s=orus **ja** sa xortare **ja** sa ksila*  
 ‘We would go down to the forest either for the herbs or for the wood.’  
 (04\_01072019F\_2; 180–181)
- (644) ***ja** dajaxi n=etroyame **ja**..*  
 ‘Either we got the bashing or.. .’ (04\_01072019F\_2; 291)
- (645) ***ne** boro tš=erxume **ne** esi*  
 ‘Neither I can come nor you.’ (01\_15022015F\_1; 23–24)
- (646) *edode evelden **ne** makina edune **ne**..*  
 ‘At that time, there was neither a machine nor.. .’ (04\_01072019F\_2; 218–219)
- (647) ***ne** faje borum tše droyum **ne** rahatje borum tš=eftjeme*  
 ‘Neither can I eat bread, nor can we make it easily.’ (07\_04072019F\_6; 58–59)
- (648) ***hem** trana sani **ham** bola*  
 ‘They were as well big as many.’ (08\_04072019M\_1; 160)
- (649) ***hem** d apiđe ba fa **hem** do dajaye fa*  
 ‘Eat the pears as well as the bashing.’ (04\_01072019F\_2; 296–297)

#### 3.2.6.2 Subordinating conjunctions<sup>141</sup>

While subordinating conjunctions of time are usually inherited, other subordinating conjunctions, like for example that of cause, are borrowed from Turkish. The subordinating

<sup>141</sup> Although subordinating conjunctions can be considered part of the functional category “complementizer”, they are termed here ‘subordinating conjunction’ to express in a narrow sense the part-of-speech they are and the function they take over. One reason for this is that the term “complementation” (and consequently also the term “complementizer”) is used here in the strict sense as grammatical element selected by a predicate as opposed to “subordination” which includes adverbial phrases and other non-obligatory phrases (see also the terminology in

conjunctions available at the corpus will be briefly exemplified below; further research on the different kinds of subordinating conjunctions is required.

The conjunction of cause Tr. *cünkü* ‘because’ is borrowed into Romeyka either as *tšüngü* or in the unround Trabzon regional variant *džingi* (650, 651).

- (650) *do getšinima mono zori edune džingi ađa so mehele emistine đormo utš ixame*  
 ‘The living was a bit difficult, because here in the neighbourhood we did not have a road.’ (08\_04072019M\_1; 059–062)

- (651) *so dolabin do ine mono džidin ebsindane džüngü marendane ja so dolab*  
 ‘Those who are at the repository cooked a bit difficult, because they got pale at the repository.’ (03\_30062019F\_11; 090–091)

The subordinating conjunction of purpose *đae*, which is actually a preposition (see Section 3.2.5) does not figure in the present Romeyka corpus but is according to Sitaridou (2014b) attested for ROF (652).

- (652) *đæ temon t=erθan=im xavitsin eřtes=me*  
 for POSS.1SG the=coming=POSS.1SG pudding.ACC make.2SG=OPN.1SG  
 ‘For the sake of my coming, you will make me some pudding.’ (Sitaridou 2014b: 41, exs. 27, presentation/glosses modified)

For expressing ‘therefore’, inherited strategies (653, 654) exist, although incidentally the Turkish expression is used as well (i.e., TTr. *onun itšun* ‘therefore’, 02\_02022015F\_1; 123). Ex. (653) includes the preposition *ja* plus pronoun. The construction in (654) is not fully transparent but seems to be formed by deictic *ha-* and a spatial adverbial.

- (653) *jad ado iba e.. erθa s ađaha katsa ga*  
 ‘Therefore [lit. ‘for this’] I said I came here and sat down.’ (03\_30062019F\_6; 33–34)

- (654) *belki da ha-džán udž eřd=a*  
 ‘Maybe that is why it<sup>?</sup> did not do it.’ (03\_30062019F\_6; 48)

Subordinating conjunctions of time are usually inherited like *prin*<sup>142</sup> ‘before’ (655), and *os* ‘until’, which occurs with the subordinator + *na* (cf. subordinator *úsna* in Drettas 1997: 381); see also Section 5.3.3.1.1. Ex. (656) features a conjunction *avid* ‘after’ whose analysis is uncertain. Likewise, the analysis of *tša* in (657) is unclear.

- (655) *so bazar prin bao di fadime zijared n=eřtao*  
 ‘Before going to the market, I will visit Fatime.’ (B1)
- (656) *o džiris ades avid eřaen do fajido as=osbe ekserin*  
 ‘Her father left the house without eating the meal.’ (B1)
- (657) *ton ovleas tu troyu tša embefstoka tši tšimume*  
 ‘After having eaten lunch, I lay down for a while.’ (C1)

Section 5.3). Furthermore, for the sake of the basic grammatical description of complex phrases in Romeyka, there is no claim made here about “null-complementizers” or the syntactic slot filled by a complementizer.

<sup>142</sup> Drettas (1997: 380) describes the subordinating particle in PG as *prina*, including the modal particle *na*.

In temporal adverbial clauses, the clausal subordinators *anda/onda*, *omon*, and *(o)sat(i)* are used. For a detailed description of temporal adverbial clauses see Section 5.3.3.1.1.

*Anda/onda* ‘when’ is realised in immediately pre-verbal position (658–660). The etymology of this form is unclear. Drettas (1997: 377–378) describes this subordinator in the form *ontan*; Neocleous (2020) confirms for ROF the form *anda*. Assuming that *anda/onda* is probably not derived from Greek, it could be analysed as a Turkish form consisting of the noun *an* ‘moment’ + locative maker *-DA*. As an alternative scenario, the Greek conditional particle *an* could be combined with the Turkish locative marker (M. Janse, p.c.). Phonologically, the form resembles the MedGr participle *-onda* ‘being’ (Manolessou 2005: 248); a reanalysis of this ancient gerund as (temporal) subordinating particle would be probably not totally unlikely. According to Drettas (1997: 377), *ontan* is used to express simultaneity with a potential generic meaning.

- (658) *adin i dran anda xandane do yarðeli=s düšünef*  
 ‘When the elders die, you think of your child.’ (01\_04022016F\_1; 088)
- (659) *anda θelename bola kala insana anda ine.. erthane panda epejnam edži*  
 ‘When we wanted, when very nice people came, we always went there.’  
 (09\_04072019\_7; 21–23)
- (660) *temeteron to mahallen anda ertes me ti mana=s taništim esi etrižes ežis*  
 ‘When you came to our neighbourhood, when I met your mother, you were running there.’ (01\_07072019F\_1; 02–03)

In similar environments in the subordination of clauses *omon* or *emon* (02\_29062019F\_2; 32–34) ‘like’ + relativizer *to* is used. Drettas (1997: 378–379) discusses this subordinator in the compound form of *amont(o)* (> *omon* ‘like’ + the relativizer *to*). Although its differences from *anda* are subtle, *omon* refers to a punctual moment when two events overlap (Drettas 1997: 378) and possibly includes a semantic aspect of manner, which exclusively temporal *anda* does not exhibit. Tursun (2019: 226) also translates *omon* with Turkish *kadarı-yla* ‘until-like’, which also combines a temporal and a manner aspect, e.g., *omon t=elepo*, Tr. *gördüğüm kadarıyla* ‘as I see’ (Tursun 2019: 226).

While exs. (661–664) have primarily a temporal reading, (665) has more a relativizing function; similarly, the PP in (666).

- (661) *omon d=erakse=me išd ebiyame son doxtorin*  
 ‘When it bit me, we went to the doctor.’ (02\_29062019F\_2; 32–34)
- (662) *omon dž=ebiyam oneon dodže šimu na drano šimu*  
 ‘When we went, suddenly? it rained, a big rain.’ (03\_30062019F\_6; 41)
- (663) *ebedž ümon d=udž evra do džesdani to mural=im bozulefde*  
 ‘But then when not finding my wallet, I became sad.’ (08\_04072019M\_1; 173–175)
- (664) *omon d=eroise s=ormi ejendune na-tsurula*  
 ‘When she fell into the water, she got very wet.’ (04\_01072019F\_13; 45)
- (665) *emon d=erabas o rabis ajeli=se*  
 ‘May God grant you what you like.’ (01\_15022015F\_1; 14)

- (666) *eyo esena hats leyo ama [haets **omon do** leyo] yolajin utš edun*  
 ‘I told it you like this, but it was not so easy like the way I told you.’ (02\_29062019F\_2; 47–49)

Another particle used for temporal subordinate clauses is *(o)sat(a)* (also *osati*, Tursun 2019: 580). Tursun (2019) translates it with the Turkish converb *-(y)ken*, e.g., Tr. *yapar-ken* ‘while doing’. Although there is only a single example available at the corpus which does not allow to finally conclude on the function of this construction (667, cf. also 668), *osata* may have a durative meaning that distinguishes it from the other temporal subordinators. Note that Drettas (1997: 379–380) discusses the temporal subordinator *sít(ae)* expressing temporal simultaneity. The example in (669) suggests that in copula clauses with *osata* no overt copula is required (also possibly 03\_30062019F\_11; 105). This would be in line with the use of the Turkish converb *-(y)ken* which already includes the copula (Göksel & Kerslake 2005: 79). Otherwise, however, the form in (669) could be also possibly due to the phonological similarity of the particle *osata* to the Turkish conditional construction *olsa da* ‘if there was’. Finally, (670) shows another particle of unclear etymology which also seems to function as a temporal subordinator (perhaps, it might be related to a form of the PG future marker *tha* although in form it is homonymous with the deictic particle *ha-*).

- (667) *sad erxumune t=aleyo=m ege estrofi*  
 ‘When coming, my horse turned down.’ (08\_04072019M\_2; 100–103)
- (668) *osati erxome so spiti*  
 ‘when coming home’ (Tursun 2019: 580)
- (669) *osada šindže as es.. asirame tš=etroyame tše o koliva*  
 ‘If there were any, we could eat [lit. ‘pull and eat’] boiled corncobs now.’  
 (04\_01072019F\_2; 272)
- (670) *so xor ha pies kadi pao eftayo kadi fidevo kadi tsakono*  
 ‘When you go to the field, I go [and] make something, I plant something, I break something.’ (03\_30062019F\_7; 07)

### 3.2.7 Interjections

Interjections are a part of speech that can stand alone (i.e., unlike particles), does not modify another part of speech (i.e., unlike adverbs) and does not inflect. Interjections include greetings, response particles, exclamations, curses, and hesitation markers and form as such a word class involving discourse markers as well as fillers. Some frequent interjections in Romeyka are briefly presented below.

Greetings that reflect in the Romeyka corpus are *eba* for an informal greeting between women (also at the telephone, 671) and the form *ebats* (> *batsi* ‘girl’) with which an elder woman can address a girl or young women (672). For opening a telephone call, Turkish *alo* ‘hello’ is used (671). On the level of exclamation, *eh* can be used to catch attention (673; also 04\_01072019F\_1; 074, i.a.; and Tr. *eh giz=im* ‘oh my girl’ in 02\_29062019F\_2; 45, probably equivalent to *ebats*).

- (671) *alo eba bu.. abohendžeka stedžis*  
 ‘Hey there, where.. at which side are you staying?’ (01\_15022015F\_1; 03–04)

(672) *ebats arzu, i mana=s d=eftjej*  
 ‘Hey girl, Arzu, what does your mother do?’ (03\_07072019F\_1; 01–02)

(673) *e babaanne b=isin*  
 ‘Hey, grandmother, where are you?’ (03\_30062019F\_7; 12)

Response particles of agreement and confirmation are *he* (674), *ha* (675), or stronger reduplicated *hehe* (676) or *haha* (01\_04022016F\_1; 104); see also Section 3.2.4.2 above.

(674) A: *etšinena eynorizez me*  
 B: *i šexrije*  
 A: *he, ineka*  
 ‘Do you know her?’ – ‘Şehriye?’ – ‘Yes, a woman.’ (01\_04022016F\_1; 020–022)

(675) *ha adšijan adšijan so boyazi*  
 ‘Yes, up there at the canyon.’ (08\_04072019M\_2; 087)

(676) *hehe he isa isa*  
 ‘Yes yes, yes, correct, correct.’ (01\_04022016F\_1; 075)

*Ha* is otherwise used with a falling tone (i) to confirm (677) or put emphasis on a just made statement (678, 679); (ii) if something new comes to the speaker’s mind (680, 681); or with a rising tone (iii) in the case of non-understanding meaning ‘hugh’ (06\_03072019M\_2; 50). Prolongued *ha:* with a falling tone demonstrates comprehension (682). The ideophone *afkur* in ex. (678) is a regional Trabzon Turkish imperative of the verb *afkur-mak* ‘bark, howl’ (Standard Tr. *havla-mak* ‘bark’), which is used in a pejorative way as to say ‘shut up’; *ilaxsu/ilakson* ‘bark.IMP’ is the Romeyka equivalent.

(677) *araba joli eš ha embro utš išen ðromo*  
 ‘There is a car road; before, there was no road.’ (08\_04072019M\_1; 211–212)

(678) *ha afkur, afkur o škilo ilaxsu*  
 ‘Bark, bark like a dog!’ (01\_28062019F\_3; 25–26)

(679) *efa laxana ha*  
 ‘I eat cabbage!’ (01\_28062019F\_3; 61)

(680) *ha eyo sas šeher selame n=aboliyo=sas*  
 ‘I will send you.’ (unclear translation; 08\_04072019M\_4; 01–02)

(681) *har ero ndje ha eyo epejna so tafi tše*  
 ‘Now I don’t know.. . Ha! I went to the grave.’ (07\_04072019F\_6; 35–36)

(682) *ha: hado bal abe mešhur*  
 ‘Ha, this is also well-known!’ (01\_28062019F\_2; 17)

Other response particles are *ah* for confirmation and comprehension (683; also 04\_01072019F\_13; 49), and *aha* for a confirmation of an assumption (684).



- (683) *ah ebijen jeksu eroise so nero dže kamenos endune nanero*  
 ‘She went outside, she fell into the water and, poor her, became wet.’  
 (04\_01072019F\_13; 24)

- (684) *aha ebije evre d=opsarae*  
 ‘Aha! He went and found the fish.’ (04\_01072019F\_12; 28)

*Ja* is used like the areal interjection *ya* (< Tr./Ar. *yahu*, Göksel & Kerslake 2005: 105) for attracting attention (685) and as a demand (686). *Ja*: can apparently be also used as a strong affirmation (687).

- (685) *aha da jisi=nad ja de ta is ebidže*  
 ‘Aha, his footprints! Look, he made footprints!’ (04\_01072019F\_12; 24)

- (686) *e batsi=mu ja elade*  
 ‘Oh my girl, come!’ (04\_01072019F\_2; 198)

- (687) *ja: olon tranisa eyo mune*  
 ‘Yes, I was the oldest.’ (04\_01072019F\_2; 069–070)

Exclamations are basically formed by all vowels: *oh* is used (i) for emotional attachment both negative (with a falling tone; 688) and positive (starting with a high tone, then falling and rising again; 689); or (ii) a feeling of wellbeing and enjoyment (short *oh*; 690, also 02\_21042018M\_2; 52); *hao*: for admiring confirmation (04072019F\_10; 23). *Uh* is used for astonishment (691). *Eh* can be used for pitying (692) and is otherwise used as hesitation marker.

- (688) *oh do jaš=im eđeve hare*  
 ‘Oh, my age has passed now..’ (03\_30062019F\_8; 03)

- (689) *ksero tabi ah utš eksero ooooo*  
 ‘I know, of course, how could I not know, oh!’ (04\_01072019F\_1; 065)

- (690) *epikame džahedest to keifi=muna oh*  
 ‘We rejoiced in it, there we felt comfortable, oh!’ (03\_30062019F\_2; 36)

- (691) *u bola emorfo batsi en*  
 ‘Uh, she is a very beautiful girl!’ (01\_28062019F\_2; 34)

- (692) *eh demo i batsi ba ena dek eše o kadar sikinti exa*  
 ‘Oh, my daughter has also one child, she had so many problems!’ (01\_04022016F\_1; 085)

The Arabic secondary interjection *vallahi* (< Ar. *wa-llāh* ‘by God’) is used as a strong confirmation of a statement (01\_28062019F\_3; 29). Apart from this, many other Arabic exclamations are used in informal speech, e.g., *gali inallillah* (< Ar. *kālū innā lillāhi* ‘we came from God’), roughly translating as ‘my God’ (02\_2906019F\_1; 07); *ej bejuk allahum* ‘oh good God’ (< Tr. *büyük allah-ım* ‘my great God’; 01\_04022016F\_1; 119).

Curses are likely to be Turkish, e.g., *kimu sikin* (< Tr. *kimin sikine*) roughly meaning ‘who cares’ (03\_30062019F\_11; 041).

Discourse markers are usually borrowed from Turkish, e.g., *olsun* ‘never mind’ (< Tr. *ol-sun* ‘it shall be’; 693); *haidi* ‘come on’ (694); *aini* ‘the same’ expressing consensus is used at least in a Turkish string of code-switching (04\_01072019F\_1; 084). The Turkish secondary interjection *ej giddi* ‘how it went’ is frequently uttered by women of a certain age revelling the past (695).

(693) *mono fazla konuŝevume olsun*

‘Do I speak too much? Never mind!’ (01\_28062019F\_3; 17–18)

(694) *haidi anne son doxtorin na pame*

‘Come on, mother, let’s go to the doctor!’ (02\_2906019F\_1; 30)

(695) *epsenam tŝe etroyame layamis da insane e gidi*

‘We cooked it and ate it, happy people – how the time has passed.’ (04\_01072019F\_1; 188)

## 4 Morphology

### 4.1 Word formation

Before describing word formation by derivation, the “core” of a word needs to be defined. The smallest unit of a word in Romeyka which is used as the basis of word formation is the root. Van Emde Boas et al. (2019: 261) distinguish a root from a nominal or verbal stem by the fact that the latter can contain “elaborations”. An example are verbal stems which are linked to a single root, *fero* ‘bring’, *foro* (TD) ‘dress’, *fortune* ‘load’.

This section describes basic strategies in Romeyka word formation, including derivation and compounding. However, while word formation is common and well-documented in AG, the corpus evidence for word formation in Romeyka is relatively thin and traces of historical developments are difficult to establish (for an untrained eye). Apart from the typical word formation strategies, mechanisms of loan word integration are treated in this section as well, since in a language shift scenario - as it is the case for Romeyka - borrowing is a productive means of increasing the lexical material.

#### 4.1.1 Derivation

##### 4.1.1.1 Nominal derivation

Nouns can be derived from nouns ( $N \rightarrow N$ ), such as, for example, the name of inhabitants from a place name like in *Ótšena* (Ogene/Karaçam köyü) > *otšenote* ‘person from Ogene’ (Tursun 2019: 579). Furthermore, a variety of other nominal derivation suffixes are attested in the literature on PG, e.g., *aryad-ia* ‘hired workers’ > *aryátes* ‘worker’ (Karatsareas 2014: 94).

Furthermore, nouns can be derived from verbs ( $V \rightarrow N$ ). Examples of deverbal nouns are:<sup>143</sup>

*foro* (TD) ‘dress’ (< AG *foreo* ‘carry habitually’) > *i foresia* ‘cloths’  
*yrafo* ‘write’ > *i yrafî* ‘scripture’, *to yrama* ‘writing/letter’ (Tursun 2019: 225)  
*psefkume* (TD) ‘lie’ (< AG *psevðome* ‘lie’) > *to psema* (TD) ‘lie’  
*fovume* ‘fear’ > *o fovos* (TD) ‘fear’  
*fovume* ‘fear’ > *fovorizo* ‘scare’ (Özkan 2013: 147) > *foverixtra* ‘scarecrow’  
*xareno* ‘rejoice’ > *i xara* (TD) ‘happiness’

Verbs are productively nominalized to form action nouns. There are two nominalization strategies: (i) deverbal nouns are formed by means of the suffixes *-ma(n)*, *-mo(n)*, *-io*, or *-ia* that are attached to the aorist stem (without initial augment /e/); (ii) nominalized infinitives where the ending *-mo(n)* is attached to the aorist stem + infinitival ending *-ini/-eni* (see also Section 5.3.2.2.2). It is not clear what determines the selection of the suffix in deverbal nouns. Examples of deverbal nouns with the suffix *-ma* (*-man* in other varieties, cf. Tursun 2019) are *opsarema* ‘fishing’, *gölimbema* ‘swimming’, *porpatema* ‘walking’, *xorema* ‘dancing’, *landžema* ‘jumping’, *fajima* ‘catering’, *spundžema* ‘swiping’. Examples for deverbal nouns with *-mo(n)* are *fanimo* ‘eating’, *kopsimo* ‘tree logging’, *panimo/banemo* ‘going’, *yrapximo* ‘writing’ (TD); nominalized infinitives with *-mo(n)* are *kseraθ-ini-mo* ‘drying’, *pis-eni-mo* ‘making’. Examples of verbs nominalized by the suffix *-io* are, e.g., *tšimeθio* ‘sleeping’,

<sup>143</sup> Note that Tursun (2019) provides plenty of derived forms for the lemmata in his dictionary. However, as most of them do not figure in the present corpus and can thus not be confirmed in the present research, forms only found in Tursun (2019) are referenced as such by (TD).

*xaθio* ‘dying’ (TD), *lastio* ‘straying’ (TD); an example for a deverbal noun in *-ia* is *yanaxtizō* ‘get tired’ > *yanaxtija* ‘fatigue’.

Turkish loanwords are readily nominalized as well drawing on the *-ma* suffix: *gonuřema* ‘speaking’, *egutema* ‘education’, *getřinima* ‘getting on well’, *třalijeřema* ‘working’. The selection of the nominalization stem *ma-* is likely induced by a Turkish nominalization strategy (so called “short infinitives”) which makes use of homonymous *-mA*, e.g., *oku-mak* ‘read’ > *okuma* ‘reading’. All nominalizations, native or borrowed, receive neuter gender.

Finally, nouns can be derived from adjectives (Adj → N), e.g., *kalos* ‘good’ > *o kalos* ‘good man’. In this case, the gender-sensitive adjectival forms simply receive the respective definite article.

#### 4.1.1.2 Adjectival derivation

Adjectives, which predominantly occur in predicative attribution (Section 3.1.2.2), are derived from participles in *-menos*, e.g., *binasmenos* ‘hungry’, *đipsasmenos* ‘thirsty’, *xaremenos* ‘happy’, *xamenos* ‘dead’ (see Section 4.3.7.3). The adjectival form may in turn lead to a null derived noun, e.g., *o kamenos* ‘the poor’.

#### 4.1.1.3 Adverbial derivation

Many frequently used adverbs of place and time are multi-word constructions involving prefixed prepositions which are perceived as a single phonological word due to pausing and stress. Such adverbs of time which involve the preposition *eb-/ap-* ‘from’ are for example *ep-iřtera* ‘later on’ and *(tř-)eb-etře-nahen* ‘thereafter’, lit. (probably) ‘and from this later’, shorter forms being *eb-edř-ae* ‘later’, *edře-naen* ‘later’, *(e)tř-eb-edři(n)* ‘later’. Another form *hado-bahen* ‘then?’ is probably a compound of the third person pronoun + demonstrative *h(a)-*, *h-ado* ‘this’ and a temporal suffix *-?ahen* ‘later’.

Adverbs of place involving a preposition form part of the diverse spatial expressions in Romeyka. The following examples involve the preposition *eb-/ap-*:

*eb-edřega* ‘from there’, *h-ab-adřega* ‘from here’, *h-ab-ađa-han-dřega* ‘from here’, *h-ab-ađa-ha* ‘from around here’,  
*eb-ebuka* ‘from below’,  
*eb-eban* ‘from up’,  
*h-ab-adři-beran* ‘from opposite there’, *ap-atř-eperan-dřes* ‘from around opposite there’,  
*ab-adř-ablan* ‘from over there’,  
*ab-a(du)-dřan* ‘from up here’.<sup>144</sup>

#### 4.1.1.4 Verbal derivation

Verbal derivation is productive in Romeyka. A variety of derivational affixes is used for verbalization, causativization, or integration of Turkish loans. A root stem to which derivational suffixes are attached forms the verbal base on which the inflection suffixes are attached. According to Özkan (2013:147), a single affix can have several functions: For example, the suffix *-(i)zo* is used: (i) for verbalization of nouns, e.g., *fos* ‘light’ > *fosizo* ‘to be lit/to shine’; (ii) to increase transitivity or as a causative marker, e.g., *maθano* ‘to learn’ > *maθizo* ‘to teach’ (see also Janse 1999); (iii) for the formation of onomatopoeic verbs such as *tsatsalizo* ‘to weed’; (iv) to grammatically adapt Turkish verbs like in *dajanizo* ‘to endure’ from Tr. *dayan-*

<sup>144</sup> Probably also *apohon-dřega* ‘where.REL’.

‘to endure’. Another suffix, which is especially in loan word integrations (Section 4.1.4.) very productive is *-evo* (see also Mackridge 1987: 127). Two forms of verbalization are addressed in the following.

Firstly, verbs can be derived from nouns. Examples for denominative verbs are *trayoḏia* ‘folksong’ > *trayoḏo* ‘sing folksongs’, *ineka* ‘woman’ > *ineḏžizo* ‘marry’, *đulia* ‘work’ > *đulevo* ‘to work’ (AG *đoulos* ‘slave’), *to faji* ‘meal’ (in some varieties ‘bread’) > *faizo* ‘feed’, *opsar(i(n))* ‘fish’ > *opsarevo* (TD) ‘to fish’, *pezo* ‘play’ > *peđi* ‘child’ (cf. van Emde Boas et al. 2019: 274). Van Emde Boas et al. (2019: 274) report that the suffix *-azo/-izo* was used in AG for formation of denominative verbs which corresponds to many Romeyka denominative verbs (see also Özkan 2013: 147).

Secondly, adjectives can be verbalized as well, e.g., *ksero* ‘dry’ > *ksereno* ‘to dry’, *fovero* ‘scary’ > *fovorizo* ‘to scare’ (ex. from Özkan 2013: 147).

Finally, compound verbs are formed by prefixation of prepositions. Such prefixes are *kata-* ‘down’, *ek(s)-* ‘out’, *sin-* ‘together’, *em-* ‘on’, *an(a)-* ‘up’ (for the corresponding AG prefixes see van Emde Boas et al. 2019: 275-276; also Drettas 1997: 490). A case in point are motion verbs going back to AG *veno* ‘go, walk’, *ey-veno* ‘leave’, *kate-veno* ‘descend’, *ane-veno* ‘go up’, *se-veno* ‘enter’, *đe-veno* ‘pass’. Some further examples of compound verbs are listed below:

*sin-* ‘with/together’ > *sindišenum̱e* ‘chat’  
*ek-* ‘out’ > *eyvalo* ‘take out’, *eksimnizo* (TD) ‘wake up’  
*em-* (< *en-* / [ \_b ]) ‘in(to)’ > *embeno* ‘enter’  
*apo-* ‘from’ > *aboliyo* ‘send’  
*peri-* ‘around, about’ > *bermeno* ‘wait’

#### 4.1.2 Compounding

In Romeyka nominal compounds, the modified (head) noun occurs in final position (like in Turkish nominal compounds, e.g., *ağaç yaprağ-ı* ‘leaf of a tree’). Gender is assigned in congruence with the gender of the head noun. The corpus contains only few examples of nominal compounds, e.g., *andrađelfo* ‘brother-in-law’ > *andras* ‘husband’, *ađelfo* ‘brother’; *tsubađalevro* ‘corn flour’ > *tsubađi* ‘corn’, *alevre* ‘flour’; *opserangišo* ‘fishhook’ > *opsari* ‘fish’; *liftokari* ‘hazelnut’ > ? ‘hazel?’, *kari* > *kariđ* ‘nut’ (cf. Bortone 2009: 86).<sup>145</sup>

According to the compound forms provided in Tursun (2019), compounding is realized by insertion of a semantically void linking element /o/ between the modifier and the head noun, e.g., *tupađopsomi* ‘corn bread’, *đentrofilo* ‘leaf of a tree’ (see also Ralli 2008, 2020 for compound marking by means of *-o-* in AMG).<sup>146</sup> Interestingly, in some examples from the Romeyka corpus, the linking element /o/ seems to be rather attached to the head noun at the end of the compound, e.g., *tsubađalevro* ‘corn flour’.

Compounds with an adjectival modifier do not figure in this corpus. Compound verbs are formed by prefixation, see Section 4.1.1.4. above.

<sup>145</sup> The corpus features a single nominal compound probably mixed with Turkish (i). While the form of the compositum is clearly Turkish, namely modifying noun + head noun + POSS.3SG suffix, the use of the genitive seems to indicate an inherited genitive possessive, which is still rare, given the fact that the preposition *os* governs accusative case.

(i) *os tu mehelle baši*  
 ‘until the begin of the neighbourhood’ (08\_04072019M\_1; 150)

<sup>146</sup> The present corpus has only a single compositum of this kind, *auropeđi* ‘boy’ (04\_01072019F\_2; 192), whereby it is not clear whether the /o/ is part of the first noun *auro(s)* ‘man’.

### 4.1.3 Other strategies of word formation

Reduplication may occur as a strategy in word formation although it is not frequent and seems to be linked to onomatopoeia, e.g., *xorxorizo* (TD) ‘snore’. Özkan (2013: 147) lists two onomatopoeic words of RSür formed by reduplication, *tsatsalizo* ‘peel’, *dsundsurizo* ‘perforate’. For reduplication in manner adverbs, reminiscent of the corresponding construction in Turkish, see Section 3.1.4.3.

### 4.1.4 Loan words

This section deals with the integration of Turkish loanwords of different word classes, namely verbs, nouns, adjectives and adverbs (especially temporal) but also some function words, in particular particles related to discourse (question particle, answer particle, discourse particles) and coordinating and subordinating conjunctions.

Function words are usually borrowed without phonological integration but possibly in the regional Trabzon Turkish phonological variant, e.g., *džingi* ‘because’ (< Tr. *çünkü* ‘because’). Probably there needs to be distinguished between borrowings and more spontaneous code-switching, though. Furthermore, loan words of all word classes seem to be frequently borrowed from older stages of Turkish, i.e., Ottoman Turkish, and are probably less frequently used in modern Turkish, e.g., *mektep* ‘school’ instead of modern standard Tr. *okul* ‘school’; *mualimina* ‘teacher.F’ vs. *öğretmen*, *vejaxut(e)* ‘or’ instead of *veja* ‘or’. This may provide evidence about the time period in which the loans were adopted.

In sum, predominantly content words are phonologically integrated into the Romeyka system. However, this excludes adverbs of time which are usually not phonologically integrated; they need to be distinguished, though, into those temporal adverbs that seem to occur more spontaneously for pragmatic reasons via code-switching, e.g., *şimdiden sonra* ‘later’, *sonradan* ‘thereafter’, *yeniden* ‘again’, and actual long-time lexical borrowings that often resemble in their use function words, e.g., *hiç* ‘never’, *kadar* ‘until’ (but also *sabaxtan* ‘in the morning’). Having said this, only those content words seem to be phonologically integrated that became part of word classes with inflectional features.

Adjectives (predominantly attested with predicative adjectives but also attributive adjectives) are integrated by a word-final /i/ (see also Section 3.1.2.1). As /i/ is also (regionally) used as inflectional ending for feminine adjectives, it is not quite clear, whether integration with /i/ is phonological or possibly grammatical integration.<sup>147</sup> The majority of Turkish adjectives is integrated by /i/ which is attached to the Turkish full word form, e.g., *temiz-i* ‘clean’, *kalabalux-i* ‘crowded’, *hazir-i* ‘ready’, *zor-i* ‘difficult’, *tik-i* ‘narrow’, *ixtiar-i* ‘elder’, *doyal-i* ‘natural’. However, some Turkish loans are integrated by /e/, e.g., *geniş-e* ‘broad’, *şenlig-e* ‘inhabited’. Given the same word shape of *geniş-e* and *temiz-i*, it is not clear what triggers the final vowel. Turkish loan adjectives that contain the Turkish derivation suffix *-lı* receive an /s/ in Romeyka, e.g., *boşlı-s* ‘indepted’ (Tr. *borçlu*), *emekli-s* ‘retired’, *hevesli-s* ‘keen’; however, also *xasta-s* ‘sick’. It can be therefore concluded that Turkish adjectives ending in a vowel are integrated by word-final /s/, while those ending in a consonant receive /i/ or /e/. However, not all loan adjectives are phonologically integrated, e.g., *džidin* ‘tough’ (Tr. *çetin*). Further research is required to estimate whether for example the kind of bilingualism of the speaker plays a role.

In general, Turkish loan adjectives inflect according to the Romeyka system, e.g., *tehlikeli* ‘dangerous’ > *tehlikeli*.PL.

<sup>147</sup> This differentiation is slightly easier with nouns where inanimates may receive /i/ or /s/, but nouns denoting human masculines receive the nominative singular masculine ending *-is*.

Nominal loan words are assigned a gender in part based on transparent inherited principles, such as feminine gender for loans ending in a vowel, e.g., *i džami* ‘the mosque’ (see Section 4.2.1.1). Occasionally, their gender may be variable, e.g., *i para, to para* ‘the money’. They inflect for case and plural like native nouns, e.g., *ta pendžere-des* ‘the windows’, *i avdži-ðes* ‘the hunters’, *ta doxtor-ja* ‘the doctors’, *ta tšešida* ‘the variants’. Loans denoting humans usually receive the ending of nominative singular masculine nouns, e.g., *o šoför-is* ‘the driver’, *o öretmen-is* ‘the teacher’ (see Section 4.2.1.1).

Nominal inanimate loans ending in a vowel may be integrated by a word-final /s/ (like adjectives ending in a vowel), e.g., *dünja-s* ‘world’ (but cf. also *dünja* ‘world’). Inanimate nominal loans ending in a consonant are integrated by word-final /i/ or /in/, thereby probably assigning them to the class of neuter gender nouns in *-(i)n*, e.g., *to džam-i* (< Tr. *çam* ‘windowpane’). This integration is applied even if it results in homonymy, i.e., *to džam-i* ‘the windowpane’ vs. *i džami* ‘the mosque’. The appearance of /i/ seems not to be systematic even for individual words, though, e.g., *düyün* ‘wedding’ (< Tr. *düğün*), *düyün-i* ‘wedding’; *do dofedž* (< Tr. *tüfek*) ‘the gun’, *do dofedž-i* ‘the gun’. Although the fact that the Romeyka typical palatalization of /k/ takes place [i] in the latter example suggests the underlying existence of the integration marker, even if omitted. The adaptation of Turkish loans to Romeyka phonological processes like palatalization of /k/ is not systematic, though, e.g., *t=ejluk-in* ‘the goodness’, but *kaimax-i* ‘cream’ (< Tr. *kay-mak* ‘cream’). In general, the majority of loan nouns seems not to receive any integration in the nominative singular, e.g., *to iz* ‘the footprint’, *to gülfe* ‘plague’ (< Tr. *külfet*).<sup>148</sup>

Internally complex Turkish nouns are integrated according to the principles outlined above irrespective of their internal structure. Complex Turkish nouns are those which are formed by derivation or composition. Examples of derivation are the Turkish profession suffix *-çl* and the suffix *-llık* which nominalizes nominal or adjectival roots: e.g., *i avdži-ðes* ‘the hunters’ (< Tr. *av-çı* ‘hunter’), *orendžiris* ‘student’ (< Tr. *öğren-ci* ‘student’), *ješilux-i* ‘verdure’ (< Tr. *yeşil-lik* ‘greenery’), *ešeklux-i* ‘stupidity’ (< Tr. *eşek-lik* ‘stupidity’, lit. from *eşek* ‘donkey’), *to genšluk* ‘the youth’ (< Tr. *genç-lik*). An example for an integrated Turkish compound is *jermalmasia* ‘potatoes’ < Tr. *yer-elma-sı* ‘potato’. These complex Turkish nouns are phonologically adapted to the Romeyka system and predominantly integrated by word-final /i/ for inanimates and the ending *-is* for nouns denoting masculine persons. A plural is assigned according to the inherited rules (see Section 4.2.1.2). Loan words can also take case marking, e.g., *as to pishuxinen.ACC* ‘from the dirt’ (02\_09062019F\_1; 08).

Finally, Turkish nominal loans can appear as part of light verb constructions. The formation of light verb constructions in Romeyka largely follows the Turkish pattern, i.e., noun + verb ‘make’/‘do’, the noun may be integrated, e.g., *jardim-i bisun* ‘do help me’ (< Tr. *yardım etmek* ‘help’, lit. ‘do help’).

Simple loan verbs are integrated either by the suffixes *-evo* or *-izo* (for *-izo* see Özkan 2013: 147). While in other AMG varieties, for example, Cappadocian, loan verb integration by *-izo* is frequent, in Romeyka, *-evo* appears to be far more productive. Tursun (2019: 42) notes that the use of the verb marker *-evo* to integrate loan verbs is only productive for twenty years. This is an interesting fact which may give a hint on the process of language shift, maybe as a consequence to the migration waves in the 1980, which gave a visible rise to shift after twenty years. For integration with *-evo*, an active and a passive verb stem exist: *-evo* or *-efkume* whereby *-evo* is far more productive. It is not clear what determines the choice of a passive or active stem for integration although the semantics of the verb (also probably reflecting in the complexity of the Turkish verb stem?) might play a role. Loan verbs in the passive stem follow

<sup>148</sup> There might be also a form of integration of Turkish loan words by means of a hypothetical suffix *-dži*, e.g., *eyilepsa do nemazdži=m* ‘I rendered my prayer’ (01\_28062019F\_2; 03). Due to lack of sufficient examples, it is not clear, though, whether this is a phonological effect or indeed a morphological form of integration.

the inflection paradigm for verbs in *-me*, while verbs integrated by *-evo* follow the inflection paradigms for verbs in *-o* (see Section 4.3.1). Examples for verbs integrated by *-evo* are, e.g., *kavur-evo* ‘mix’ (< Tr. *kavur-mak*), *jasa-evo* ‘live’ (< Tr. *yaşa-mak*), *uraş-evo* ‘deal with’, *ara-jevo* ‘seek’, *tsaliş-evo* ‘work’, *kabat-evo* ‘close’, *göster-evo* ‘show’. Verbs in *-efkume* are, e.g., *düşün-efkume* ‘think’ (< Tr. *düşün-mek*), *konus-efkume* ‘chat’ (< Tr. *konus-mak*), but also *konus-evo* ‘speak’. Often a glide is inserted before *-evo* to prevent a hiatus, e.g., *demle-j-evo* ‘brew’, *izle-j-evo* ‘watch’, *dinle-j-evo* ‘listen’, *hazirla-i-evo* ‘prepare’. This seems to be, however, variable: *jasa-evo* ‘live’, *ara-evo* ‘seek’, *jola-evo* ‘send’, *başla-evo* ‘start’. Note that Tursun (2019: 42) lists loanverbs with stem-final /e/ always without glide, e.g., *rentele-evo* ‘grate’ (also Mackridge 1987: 127). The phonological process of epenthetic /j/ in onsets, especially after /n/ (see Section 2.3.6), extends also to loan verbs, e.g., *kullan-j-evo* ‘use’.

Turkish loan verbs integrated by the mechanisms described above adhere to the inflectional paradigms like native verbs, i.e., they inflect for tense, e.g., *e-giz-ebs-e* ‘he became angry’ (< Tr. *kızmak*), or form an (inherited) infinitive, e.g., *arajepson* ‘call.IMP’. Even Turkish adjectives can become loan verbs, e.g., *e-pisman-epsane* ‘they regretted’ (< Tr. *pişman* ‘regretful’). Complex Turkish loan verbs are integrated by the same principles as simple loan verbs irrespective of their internal structure. Examples of Turkish derivational affixes that are borrowed as part of the verb – often in lexicalized verb stems whose internal composition is not even transparent in Turkish anymore – are listed below (see also Section 3.1.5). Note that often these borrowed complex verbs are the only means to express morphological functions like causatives or passives in Romeyka (see Section 5.2.2).

-DIr-, causative of often monosyllabic stems (Nemeth 1962/2020: 87):

e.g., *gez-dur-evo* ‘walk sb. around’

-t-, causative of often polysyllabic stems (Nemeth 1962/2020: 87):

e.g., *anla-t-evo* ‘tell’

-lA- with a meaning ‘provide/deal with’:

e.g., *jol-la-evo* ‘send’, *sabax-la-jevo* ‘sleep through the night’, *baş-la-epso* ‘started’, *demle-jevo* ‘brew’, *izle-jepso* ‘watched’, *dinle-jevo* ‘listen’, *destele-fsame* ‘bundled.1PL’, *hazir-la-ievo* ‘prepare’, *pesle-vo* (< Tr. *besle-mek*) ‘nurture’

-lAn-, expanded lA with passive and reflexive suffix:

e.g., *jeş-lan-evo* ‘grow old’, *kullanjevo* ‘use’, *dik-len-evo* ‘become steep’

-Il- passive: e.g., *boz-ul-evde* ‘spoiled’

## 4.2 Nominal morphology

This section discusses nominal inflection (Section 4.2.1) as well as possession (Section 4.2.2), definiteness (Section 4.2.3), and diminution (Section 4.2.4). Noun classes, that exist in AG, remain in Romeyka as underlying relics although more recent developments in the gender system lead to syncretism. Therefore, it will be rather dealt with the grammatical and semantic gender system in Romeyka (Section 4.2.1.1). Animacy and humanness play a certain role (especially in gender and definiteness) but are not discussed in a separate chapter. Case marking will be focused upon in detail (Section 4.2.1.3) since grammatical roles such as the subject of an intransitive clause, transitive clause, and object of a transitive clause are expressed in Romeyka by means of case marking (and word order) in an accusative system, i.e., the subjects of both intransitive and transitive clauses are treated alike via case marking (cf. Meakins, Green & Turpin 2018: 131). Determiner and adjective agreement are discussed in the Sections (3.2.1) and (3.1.2), respectively.



### 4.2.1 Nominal inflection

Number, gender and case are the main categories of nominal inflection. Nouns were historically assigned to declensional classes (related to phonological shape); the analysis of this system (or its remnants) in Romeyka remains incomplete, though. A tentative classification based on AG declension classes is provided below. Within the noun phrase, articles agree with the nominal head in all three categories, while adjectives show mainly number agreement with remnants of gender agreement, case agreement seems to have faded in the present data (Section 3.1.2), and at the class of numerals (Section 3.1.3), only the numeral ‘one’ *ena*, which functions as indefinite article, agrees with the nominal head in gender and case (see Section 3.2.1.2).

In general, the state of the nominal declension system in Romeyka can be compared with that of PG: Drettas (1997: 129) states that the declension system of PG has irregularities and some “deficiencies” like for example in plural or case assignment where the genitive is “relatively rare” in Drettas’ (1997) corpus. In Romeyka, the changes may be considered to go even further in some terms like increasing syncretism, although inherited properties are preserved elsewhere.

Romeyka nouns can be classified according to their original AG declension classes. A tentative overview based on van Emde Boas et al. (2019: 40–62) is provided in the following:

#### 1<sup>st</sup> declension/ a-declension: stems ending in *-a*:

(i) Fem. nouns in *-a* (also in *-i* with underlying /a/), e.g., *i xora* ‘land’; according to Drettas (1997: 125) also the following examples belong to this group, e.g., *i jineka* ‘woman’, *i karðia* ‘heart’, *i xara* ‘joy’, *i patsi* ‘girl’, *i mana* ‘mother’, *i ftira* ‘louse’, *i ðulia* ‘work’, *i kosara* ‘chicken’)

(ii) Masc. nouns in *-is/-as*, e.g., *o andras* ‘husband’ and some Turkish loans, e.g., *o öretmenis* ‘the teacher’

#### 2<sup>nd</sup> declension / o-declension: stems ending in *-o*:

(i) Masc. (and some fem.) nouns in *-os*, e.g., *o potamos* ‘river’, *o anthropos* ‘man’

(ii) Neut. nouns in *-on* (and contracted neut. nouns), e.g., *to ostuð* ‘bone’ > *to ostoun* > *osté-on*, *to astron* ‘star’; according to Drettas (1997: 126) also the following, e.g., *to xoraf* ‘field’, *to ospit* ‘house’, *to neron* ‘water’, *to xorion* ‘village’, *to peðin* ‘child’, *to milon* ‘apple’, *to ksilon* ‘wood’, *to xtinon* ‘cow’, *to provaton* ‘sheep’, *to vuteron* ‘butter’.

#### 3<sup>rd</sup> declension: Masc., neut., fem., stems ending in [C] or *-i*:

e.g., *to peði* ‘child’, *to yala* ‘milk’, *i šeri* ‘hand’, *o alas* ‘salt’, *o vuð(ias)* ‘ox’, *to yonaton* ‘knee’ > *to yoni* ‘knee’, *to ðenðron* ‘tree’ (van Emde Boas et al. 2019: 59); according to Drettas also, e.g., *to kreas* ‘meat’, *to xoma(n)* ‘soil’, *to stoman* ‘mouth’

Drettas (1997: 119) notes the PG specialty that the nominative of masculine nouns in *-os* (also some in *-as/-is*) occurs in the accusative ending *-on* (instead of *-os*), i.e., *o skilon* ‘dog.NOM’, *o aðelfon* ‘brother.NOM’, *o ðromon* ‘road.NOM’, *o arθopon* ‘man.NOM’. Holton et al. (2019: 247) mention the same PG idiosyncrasy and add that this phenomenon mostly takes place when the noun is preceded by the definite article.<sup>149</sup> The same phenomenon is referred to by Janse (2004) and Revithiadou & Spyropoulos (2012) as Differential Subject Marking (DSM). In line with Mackridge (1987: 124), in Romeyka the masculine nouns in question have the endings *-o(n)* and *-os*, if definite; whether the accusative ending is *-o* or *-on* depends on the variety and whether the variety keeps full endings (for a deeper discussion see Section 4.2.3).

<sup>149</sup> This phenomenon may be linked to the masculine(/neuter) genitive ending in *-u* (Holton et al. 2019: 248).

## 4.2.1.1 Gender

For agreement purposes, Romeyka distinguishes three genders: masculine, feminine, and neuter (see 1). Gender assignment goes back to AG declension classes (see just above), whereby feminine and neuter nouns can have different endings, but masculine nouns end frequently in *-os/-as/-is* (cf. noun classes in SMG, Karatsareas 2011: 142). This three-way distinction pertains to the singular; the plural has a two-way distinction neuter and masculine/feminine (see Section 4.2.1.2).

- (1) *i ineka*.F ‘the woman’  
*i ylossa*.F ‘the language’  
*o anθrobos*.M ‘the man’  
*o δromos*.M ‘the road’  
*to stroma*.N ‘the bed’  
*to xtino*.N ‘the cow’

Originally, gender in AG was lexically fixed (i.e., primarily morphological rather than phonological, following Corbett 1991: 33) and hence in principle not fully predictable from semantics. Underlying semantic rules assigned masculine gender to male humans and many higher animals, feminine gender to females, and neuter gender to inanimates. In PG (and also in Romeyka), gender shifted toward a more semantically-oriented assignment based on animacy, so that targets low at the animacy scale became associated with neuter gender (see Karatsareas 2011: 154, 2014). This development, starting in AG and going further in Romeyka than, for example, in SMG, includes the spread of neuter forms to masculine and feminine paradigms in both singular and plural declension of nouns, determiners, and adjectives (see 2) (Mackridge 1987; Özkan 2013).<sup>150</sup> Note in this regard that the frequency of the neuter article in the present corpus is much higher than that of the other genders (Table C.1 in Appendix C). At present, however, the spread of neuter gender seems to extend beyond [+ANIM, -HUM] since even human targets are affected.

- (2) a. *o vuðias*.M.SG – *ta vuðia*.N.PL ‘the ox(en)’ (RSür; Özkan 2013: 144)  
 b. *to vuð/vuið*.N.SG ‘the ox’ (ROf; Tursun 2019: 206)

Although there seems to be a general shift in plural declension towards a unified system based on the neuter paradigm, the shift does not affect all nouns simultaneously; rather, animacy seems to be a decisive variable in the change. Karatsareas (2011: 155) describes for AMG that elements low on the animacy scale trigger semantic agreement. Furthermore, the spread of the neuter paradigm in nominal plural declension may be supported by the syncretism of the plural masc./fem. definite article with the fem.sg. definite article. The spread of the neuter paradigm applies especially to the plural of [–HUM] feminine nouns, e.g., *da*.N.PL *kadas*.F.PL ‘the cats’ (03\_30062019F\_11; 065); *da*.N.PL *ðulias*.F.PL (08\_04072019M\_2; 144); *da*.N.PL *imera*.F=des.POSS.3SG.F ‘her days’ (01\_28062019F\_2; 32), and inanimate masculines. For an example, consider also the reanalysis of the masculine determiner *o* of the masculine noun *fengo(s)* ‘moon’ as part of the word and subsequent assignment of neuter gender in *to*.N.SG *ofengos*.M.SG ‘the moon’ (04\_01072019F\_12; 39). The spread of neuter declension extends to [+ANIM/(+HUM?)] masculine/feminine nouns, e.g., *do*.N.SG *yaiðuran*.M.SG ‘the donkey’

<sup>150</sup> The spread of the AG diminutive suffix *-ðion* to many masc./fem. nouns has caused the shift to neuter gender of many nouns, e.g., *o vuðias* ‘the ox’ > *to vu+ði(on)* ‘little ox’ > *to vuið* (incl. metathesis). The spread of neuter has, however, gone further in PG than in SMG (Mackridge 1987: 128).

(02\_29062019F\_2; 04) and may in Turkish-dominant speakers even affect masculine proper nouns, e.g., *t=ali dio yarðelæ iše* ‘Ali had two children’, T1).<sup>151</sup>

Regarding agreement between head and modifiers of the NP, in [–HUM] feminine nouns, the nominal modifier appears in neuter gender; see, for example, in (3) the use of a neuter determiner and interrogative for a feminine noun. According to Neocleous (2019: 35), modifier only agree with their head when the head is [masc.; +HUM]; however, it needs to be made clear that this only applies to the singular. For an attempt at an explanation of the change in gender agreement see Section 3.1.2.1.

- (3) *me to pio karðia*  
 with the.N which.N heart.F  
 ‘with which heart’ [folksong] (04\_01072019F\_1; 204)

Nominalized adjectives/participles seem to remain in the masculine lexical form when used for feminine referents (4).

- (4) *kamenos endune nanero*  
 ‘The poor (girl) got soakingly wet.’ (04\_01072019F\_13; 24)

Turkish loan nouns are generally assigned to a gender class and are inflected accordingly: [+HUM] masculine nouns add *-is* (*-s* following a vowel), e.g., *o öretmen-is* ‘the teacher’ (A1), *o musafir-is* ‘the guest’ (08\_04072019M\_3; 155; also see Mackridge (1987: 127) for the treatment of Turkish loans). Some inanimate loans are also assigned masculine gender, e.g., *o dünja(s)* ‘the world’ (04\_01072019F\_2; 243), *o baxtje(s)* ‘the garden’ (04\_01072019F\_17; 45), albeit [–HUM] loan nouns are often assigned neuter gender, especially if they include Turkish derivative affixes such as *-llk*, which evokes a more abstract meaning, e.g., *do jesilux-i* ‘the verdure’ (08\_04072019M\_3; 123); see also Section 4.1.4. Neuter nouns add the neuter suffix *-i* (or *-in* following a consonant), e.g., *do dofedži* ‘the gun’ (05\_03072019M\_4; 28), *to goemleg-in* ‘the blouse’ (03\_07072019F\_1; 28), *do muslux-i* ‘the faucet’ (08\_04072019M\_1; 079). According to Mackridge (1987: 127), inanimate nouns ending in a vowel are treated as feminines (unlike SMG which assigns to these nouns masculine gender), e.g., *i para* ‘money’, *i džami* ‘the mosque’. Like in native inanimate feminine nouns, in inanimate loan nouns mixed declensions may occur, e.g., *i makina*, *da makinās* ‘the machine(s)’ (04\_01072019F\_2; 243). Some loan nouns also seem to have variable gender either neuter or masculine, e.g., *o havas* (08\_04072019M\_1; 239) vs. *to hava* ‘the weather’ (08\_04072019M\_1; 244); *o mahales* (08\_04072019M\_3; 177) vs. *to mahalen* ‘the neighbourhood/municipality’ (01\_07072019F\_1; 02), whereby – at least in the latter example – there may be a semantic difference as to whether *mahale* refers to an area, i.e., meaning ‘the neighbourhood’ (neuter), or to an agent, i.e., ‘the municipality’ (masculine).

#### 4.2.1.2 Number

The feature of number in Romeyka has the values singular and plural. The plural is formed by adding a plural suffix, which is selected according to the declension class (see Table 19;<sup>152</sup> cf. also Drettas 1997: 118–128 for PG): in a very rough summary, neuter nouns inflect with *-æ* (> *ia*) or *-a*, e.g., *ta raši-a* ‘the mountains’; and feminines and masculines by *-ðæs* or *-ðes*, e.g., *i patsi-ðæs* ‘the girls’. Note that most loan nouns (except for units and numeral classifiers

<sup>151</sup> Özkan (2013: 145) states that (occasionally) feminine proper nouns may be assigned the singular neuter article *to*.

<sup>152</sup> The overview in Table 19 cannot cover the classificatory specialties of all nouns but rather presents an overview of the most common nominal declension classes.

such as *tane* ‘piece’, *kiši* ‘person’) are integrated into the Romeyka system by being assigned gender and a declension class (probably, based on their ending and semantic features) and inflecting for plural according to the system (but cf. *ta iz* ‘the footprints’, 04\_01072019F\_12; 24).

**Table 19:** Plural declension in Romeyka of Of (ROf)

Declension class	Gender	Examples singular	Plural ending	Examples plural	Turkish loanwords
1 <sup>st</sup> declension	M in <i>-as</i> (exs. from Özkan 2013: 143–144)	<i>andra(s)</i> ‘husband’ <i>babugas</i> ‘grandfather’	<i>-ðæs/-ðes</i> <sup>153</sup>	<i>i andriuðes</i> ‘the husbands’ <sup>155</sup> <i>i babugaðes</i> ‘the grandfathers’	
		<i>babugas</i> ‘grandfather’	<i>(-and)</i> <sup>154</sup>	<i>i babugand</i> ‘the grandfathers’	
	M in <i>-is</i>	<i>komšis</i> ‘neighbour’	<i>-ðæs/-ðes</i>	→	<i>i komši-ðæs</i> ‘the neighbours’
	F in <i>-a</i> (also in <i>-i</i> with underlying /a/)	<i>nife</i> ‘bride’ <i>patsi</i> ‘girl’ <i>mana</i> ‘mother’	<i>-ðæs/-ðe(s)</i>	<i>ta nif-aðe</i> ‘the brides’ <i>ta patsi-ðes</i> ‘the girls’ <i>i mana-ðes</i> ‘the mothers’	<i>i para-ðes</i> ‘the money.PL’ <i>ta pendžere-ðes</i> ‘the windows’
		<i>mana</i> ‘mother’ <i>ineka</i> ‘woman’ <i>kosara</i> ‘chicken’ <i>kata</i> ‘cat’ <i>ðulia</i> ‘work’	<i>-(e)s</i>	<i>i mana-s</i> ‘the mothers’ <i>i inetš(i)/-(es)</i> ‘the women’ <i>ta kosara-s</i> ‘the chicken.PL’ <i>ta kada-s</i> ‘the cats’ <i>ta ðulia-s</i> ‘the works’	<i>ta makina-s</i> ‘the machines’ <i>ta fasulija-s</i> ‘the beans’
2 <sup>nd</sup> declension	M in <i>-os</i> <sup>156</sup>	<i>anθropos</i> ‘human’ <i>likos</i> ‘wolf’ <i>argos</i> ‘bear’	<i>-i</i>	<i>i anθrop-i/arθob</i> ‘the humans’ <i>i litš(-i)</i> ‘the wolves’ <i>i artš-i</i> ‘the bears’	<i>i muyalim-i</i> ‘the teachers’
		<i>argos</i> ‘bear’ <i>auros</i> ‘man’	<i>-ðes</i>	<i>i argo-ðes</i> ‘the bears’ <i>i aura-ðes</i> ‘the men’	<i>i av(u)dži-ðes</i> ‘the hunters’ <i>i deliganli-ðes</i> ‘the young people’ <i>ta baxtje-ðes</i> ‘the gardens’ <sup>157</sup> <i>ta kaja-ðes</i> ‘the rocks’

<sup>153</sup> The occurrence of the vowel /æ/ in some masculine and feminine plural endings (vs. /e/) is dependent upon the variety (see Section 2.1.1).

<sup>154</sup> *-And* is a plural suffix used for masculines, which is not productive anymore; many forms have a second plural form with the productive ending *-ðæs/ðes* (Özkan (2013: 144)).

<sup>155</sup> Cf. *t=andruðe* ‘the husbands’ in Neocleous (2020: 34).

<sup>156</sup> This applies also to PG masculine nouns in *-on* that inflect in the plural also in *-i*, however, some of the masculine nouns in *-on* that Drettas (1997: 119) lists for PG receive a neuter plural in Romeyka, i.e., *o škilo* > *i škilija* ‘the dogs’, *o ðromo(s)* > *ta rðome* ‘the roads’; the noun *o aðelfos* ‘the brother’ has no distinct plural anymore and has collapsed together with the feminine form *i aðelfi* ‘the sister’ in the neuter form *aðelfe* ‘siblings’ (see also Drettas 1997: 130).

<sup>157</sup> According to Drettas (1997: 124) the gender of *baxtše* ‘garden’ is feminine, although it is treated as masculine in the Romeyka corpus.

		<i>auros</i> ‘man’	<i>-ini</i>	<i>i a.úrini</i> ‘the men’	
		<i>topos</i> ‘ground’	<i>-e</i>	<i>ta dob-e</i> ‘the places’	<i>i arkadaš-e</i> ‘the friends’
		<i>đromos</i> ‘road’		<i>ta rđom-e</i> ‘the roads’	
	N in <i>-on</i> (& contracted N nouns, e.g., N in <i>-in</i> )	<i>astron</i> ‘star’	<i>-a</i>	<i>t=astr-a</i> ‘the stars’	
		<i>neron</i> ‘water’		<i>ta ner-a</i> ‘the waters’	
		<i>ksilon</i> ‘wood’		<i>ta ksil-a</i> ‘the woods’	
		<i>milon</i> ‘apple’		<i>ta mil-a</i> ‘the apples’	
		<i>provato(n)</i> ‘sheep’		<i>ta provat-a</i> ‘the sheep’	
		<i>to peđín</i> ‘child’		<i>ta peđi-a</i> ‘the children’	
		<i>aloyo</i> ‘horse’	<i>-ta</i>	<i>t=aloyo-da</i>	<i>ta šea(-da)</i> ‘the things’
		<i>stroma</i> ‘bed’		<i>ta stroma-ta</i>	
	N in (underlying) <i>-ion</i>	<i>xoraf</i> ‘field’	<i>-æ/e</i>	<i>ta xorafæ</i> ‘the fields’	<i>ta ders-e</i> ‘the lessons’ <i>ta tšitšeg-e</i> ‘the flowers’ <i>ta fotoraf-e</i> ‘the photographs’ <i>ta pontol-æ</i> ‘the trousers’
		<i>ospit(in)</i> ‘house’		<i>t=ospidæ</i> ‘the houses’	
3 <sup>rd</sup> declension	M, N, F in <i>-i</i>	<i>kalaθi</i> ‘basket’		<i>ta kalaθe</i> ‘the baskets’	
		<i>opsarin</i> ‘fish’		<i>t=opsaræ</i> ‘the fish’	
		<i>omátin</i> ‘eye’		<i>ta máta/æ</i> ‘the eyes’	
		<i>to đendron?</i> ‘tree’		<i>t=rđende</i> ‘the trees’	
		<i>kariđ</i> ‘walnut’		<i>ta kariđe</i> ‘the walnuts’	
		<i>yarđel</i> ‘child’		<i>ta yarđelæ</i> ‘the children’	
		<i>kariđ</i> ‘walnut’	<i>(-ia)</i>	<i>ta kariđia</i> ‘the walnuts’	<i>ta doxtor-ja</i> ‘the doctors’
		<i>opsarin</i> ‘fish’		<i>t=opsárja</i> ‘the fish’	
		<i>xinari?</i> ‘trace’		<i>ta xinárja</i> ‘the traces’	
	or stem in [ _C ] = in <i>-ma</i>	<i>raši</i> ‘mountain’	<i>-a</i>	<i>ta raši-a</i> ‘the mountains’	<i>ta lastig-a</i> ‘the galoshes’
		<i>mandri</i> ‘stable’		<i>ta mandri-a</i> ‘the stables’	<i>ta demlig-a</i> ‘the tea pots’
		<i>faji</i> ‘bread’		<i>ta fai-a</i> ‘the breads’	
		<i>šeri</i> ‘hand’	<i>-e</i>	<i>ta šer-e</i> ‘the hands’	<i>ta insan-e</i> ‘the people’
		<i>onoma</i> ‘name’	<i>-ta</i>	<i>t=onomat-a</i> ‘the names’	

According to Table 19, many masculine plural nouns – as well as many feminine nouns – inflect in *-ðes*, which is a very productive plural ending. *-ðes* is also assigned to masculine and feminine loan words. It is not fully clear, though, according to which principle loan nouns are assigned masculine gender; certainly, nouns denoting male human referents are treated as masculine. While human masculine nouns – both loan and inherited – consistently receive the [+hum] definite plural article *i*, inanimate nouns (also feminines, see below) receive the neuter plural article *ta*. Some masculine nouns in *-os* inflect in the plural in *-i* and *-e*. These endings also occur in Turkish loans; it seems that the *-ðes* ending is used for loans ending in a vowel and the *e* and *i* endings are assigned to loans ending in a consonant, albeit the trigger of their respective distribution is unclear. The *-i* ending also applies to nominalized adjectives, e.g., *i palei* ‘the elders’, *i drani* ‘the elders’. Another rare plural suffix is *-ini*, e.g., *aúriini* ‘men’.

An older masculine ending *-and* (Mackridge 1987; Drettas 1997) seems to have been superseded by the *-ðæs* ending, as evident in a double set of plural forms, e.g., *babugand* / *babugaðæs* ‘grandfathers’ (Özkan 2013: 144). The plural in *-and* is only used for some nouns often denoting peoples, families, relatives, professions, and some animals (Özkan 2013: 144), often in the meaning of ‘people of’ (Drettas 1997: 121), e.g., *i turkand* ‘the Turks’ but cf. *i turtš* ‘the Turks’ in Deffner (1878: 212). Interestingly, the Romeyka corpus features the feminine plural *i mamikandi* (> *tsi mamikandus.gen*) ‘the grandmothers’ (e.g., 04\_01072019F\_2; 184).

Feminine plural nouns receive the ending *-ðe(s)* or *-s*. The *-s* ending is used for nouns ending in the singular in *-a*. This extends to Turkish loans ending in *-a* that are treated as feminine (although they receive the neuter plural article), e.g., *i makina - ta makinas* ‘the machine(s)’.<sup>158</sup> It is not clear what predicts the occurrence of the plural form *-ðe* or *-ðes*; for example, for the word *patsi* ‘girl’, both forms occur, *patsi-ðae/-ðe* (4 tokens) vs. *patsi-ðæs* (7 tokens) ‘daughters’. Furthermore, the historical *-i* ending of some masculine nouns extends to some feminine nouns where it can cause palatalization: a /k/ in the word-final syllable of masculines or feminines may yield a /tʃ/ plural ending – at least in some varieties, e.g., *i turtš* ‘the Turks’; *i inetš(i)* > *i ineki* ‘the women’. Note that many feminine nouns – even [+HUM/ANIM] and certainly inanimates – are assigned the neuter plural article (see also adjectival declension, Section 3.1.2.1). According to Revithiadou & Spyropoulos (2021), the spread of the *-ðes* suffix to feminine nouns in *-að*, e.g., *mana* ‘mother’ > *mana-ðæs* ‘mothers’ triggers neuter concord, e.g., *ta kala ta manaðæs* ‘the good mothers’.

Neuter plural nouns with an (underlying) diachronic ending in *-ion* carry the ending *-ae* or *-e* (in some varieties), but interestingly also *-a*, e.g., *opsarae*, *opsara*, *opsarja* ‘fish.PL’. Whether /ae/ or /e/ is used depends, apart from the variety, on the word but is also subject to intra-speaker variation. In assignment of the correct plural form, word stress appears to be distinctive, i.e., nouns ending in stressed /i/ which receive the plural ending *-a* keep the vowel consequence /ia/, e.g., *peđi* + *-a* > *peđia.PL*; while words in unstressed /i/ + *-a* undergo the change /ia/ > /ae/ or /e/. In some varieties (RSür<sup>2</sup>) – and probably in speakers with exposure to SMG – also the ending *-ja* is used which features the underlying process of unstressed /i/ + *-a* > ia/ja. Turkish inanimate/neuter loans may receive the same endings; interestingly, even unstressed /ia/ occurs in the Turkish loan *doxtorja* ‘doctors’. Neuter nouns in *-on* receive the plural in *-a*; if the noun in the singular ends in a vowel, the plural suffix is *-ta*, e.g., *stroma* > *stromata* ‘beds’. The *-a* plural ending is also used for neuter loans, whereby it seems that loans with /i/ in the final syllable are assigned to the 3<sup>rd</sup> declension of nouns of all gender in *-i*, e.g., *lastig-a* ‘galoshes’, *demlig-a* ‘tea pots’; but it is not transparent according to which principle loan nouns end in *-e* and whether this can actually be taken as a sign of integration into the traditional declension class system (e.g., as part of neuter nouns with underlying *-ion*); phonology seems not to be the

<sup>158</sup> A special case is the Turkish loan *para* ‘money’ > *i paraðes* ‘the money.PL’, which has shifted in plural to a feminine declension class while being otherwise neuter, i.e., *to para* ‘the money’ (but cf. *i para* > *ta paraðes* ‘money.PL’ in PG, Drettas 1997: 129).

reason, cf. *fotorafe* ‘photographs’, *demliga* ‘tea pots’. Finally, nominalizations in (the singular in) *-ma* are very productive in Romeyka, e.g., *xorema* ‘the dancing’, but it is not clear whether these do receive a plural form and whether they would be grouped in the 3<sup>rd</sup> declensions class together with other nouns in *-ma* that receive the plural *-ta*.

In sum, the original declension classes have increasingly become difficult to identify (Drettas 1997) and the system of number assignment in Romeyka shows some traces of instability because of the following factors: (i) masculines and feminines have often come to share the same endings (Özkan 2013); (ii) there is the tendency to assign masculine and feminine endings only to animate nouns, whereas all inanimate nouns are, irrespective of their historical grammatical gender, now treated as neuter (Dawkins 1931; Drettas 1997; see also Karatsareas 2011 for neuterisation in Cappadocian Greek); (iii) a more general spread of plural neuter forms to masculine and feminine nouns; (iv) dynamics in the system that are related to changes in the productivity of plural endings and/or the gender system that lead to not few nouns having competing plural forms, e.g., *nife* > *nifaðe*, *nifæs* ‘brides’ (Özkan, n.d.), *i artši* vs. *t=argoðes.N.PL* ‘the bears’, *i manas* vs. *i manaðes* ‘the mothers’, *aurini* vs. *ayul*, *aurus*, *aurazis* ‘men’ (variation seems to be sensitive to definiteness, see Section 4.2.3). All this leads to the fact that for many nouns the plural ending cannot be easily deduced from the singular ending (Drettas 1997: 129). Janse (2002) suggests a possible areal influence of the neighbouring languages not exhibiting gender distinction on the instability of Romeyka nominal inflection.

The following examples show additional features of a potential instability of the number system. Firstly, there appear to be overgeneralizations of plural forms for some nouns where the plural form is more frequent (5), a phenomenon addressed in earlier literature as “local markedness” (Tiersma 1982).

- (5) *ebiya sa guguvaga ena guguvaka utš evra*  
 ‘I went to the mushrooms. I did not find a single mushroom.’ (04\_01072019F\_5; 38–39)

Second, number marking and number agreement of some Turkish loans is unclear (6, 7). In (8) the use of plural is unclear; possibly a large city can be semantically interpreted as plurality.

- (6) *ta is epidže*  
 ‘He made footprints.’ (04\_01072019F\_12; 24)
- (7) *efieme ta xismetin=ana*  
 do.1PL the.PL service=POSS.3PL  
 ‘We do their maintenance.’ (04\_01072019F\_1; 111)
- (8) *o džiri=muna sa istambola inedžikse*  
 ‘Our father got married in Istanbul.’ (02\_02022015F\_1; 121)

Thirdly, unlike Turkish, Romeyka uses the plural form of nouns in company with numerals (9) – even with loan nouns, e.g., *đio insane* ‘two people’ (08\_04072019M\_1; 086). However, in borrowed NPs, often denoting temporal or other units, the whole Turkish construction is borrowed including matter (i.e., the Turkish numeral and noun) and pattern (i.e., the noun in the singular form), e.g., *jüz lira* ‘hundred Lira’, *elli kiši* ‘fifty people’.

- (9) *đio nifaðe exo đio tše batsið exo*  
 ‘I have two daughters-in-law and I have two daughters.’ (07\_04072019F\_8; 15)



## 4.2.1.3 Case

Romeyka has morphological exponence of nominative, accusative, and genitive cases, in principle expressed on both articles and nouns. Adjectives used to have case agreement as well, but due to the shift to neuter declension, case inflection on adjectives is often not visible anymore (Section 3.1.2.1). Morphological case distinctions on nouns are entirely neutralized in the plural, yielding case-neutral plural endings (cf. Tables 20/21).<sup>159</sup> In the singular, there is only systematic case marking on nouns left in the nominative of masculine nouns in *-os/-as/-is* (10; although sensitive to definiteness, see Section 4.2.3), the accusative of these nouns (11, but cf. Fn. 161), and the feminine genitive of nouns of the 1<sup>st</sup> declension class (12); in genitives of other genders there is considerable variation (see also Section 4.2.2). Remnants exist of the AG vocative case, which is mostly identical with the nominative, except for masculine second and third declension nouns, e.g., *e bats* ‘hey girl’ (see Drettas 1997). (In SMG, the vocative function in proper names is taken over by the accusative.) With the articles, case distinctions are partially retained in the plural (not in neuter gender), but there is possibly syncretism of plural accusative and plural genitive forms, although there is dialectal variation here and this is less evident in the forms available at the Romeyka corpus (cf. Table 11 in Section 3.2.1.1).<sup>160</sup> It seems that although AG declension classes still determine gender assignment in Romeyka, number and case inflection may be overwritten by competing strategies, such as neuterization, which often point toward simplification. This is evidenced by competing nominal plural forms such as *artš-i* vs. *argo-ðes* ‘bears’ and syncretic case forms like *ta peðia*.GEN.PL instead of *ta peðiði*.GEN.PL ‘boy.GEN.PL’ (Özkan 2013).

<sup>159</sup> This applies not necessarily to all varieties of Romeyka; according to Mackridge (p.c., also 1987: 124), it is not the case for RO<sub>f</sub> as spoken in Saráchos (Uzungöl) where the nom./acc.pl. of feminines is in *-as*. Neocleous (2020: 38) lists masculine and neuter plural genitive forms in *-ion*, e.g., *ta yarðellion*.N.GEN.PL ‘the children’s’ (in line with Drettas 1997). Özkan (2013: 143) does not list a distinctive ending for masc./neut.gen.pl. for the nouns *andras* ‘husband’, *babugas* ‘grandfather’, *gomšis* ‘neighbour’ but notes a special masculine genitive plural for the noun *peðas* ‘boy’ > *da beðiji/beðiði* ‘the boy’s’.

<sup>160</sup> Interestingly, case agreement is in Romeyka pre-dominantly visible on the definite articles and not in nominal inflection; in Cappadocian, the definite articles are often dropped, so that for example the genitive is only visible on the noun, if at all (M. Janse, p.c.).

**Table 20:** Nominal case declension in Romeyka

		SG	PL (endings see Section 4.2.1.2)
M	NOM	-s	-ðæs/ðes, -i, -e
	ACC	-(n) <sup>161</sup>	-ðæs/ðes, -i, -e
	GEN	-u <sup>162</sup>	-ðæs/ðes, -i, -e
F	NOM		-ðæs, -(e)s/-i
	ACC	-(n) <sup>163</sup>	-ðæs, -(e)s/-i
	GEN	-s <sup>164</sup>	-ðæs, -(e)s/-i <sup>165</sup>
N	NOM		-æ/-e, -a, -ta
	ACC		-æ/-e, -a, -ta
	GEN	-(i) <sup>166</sup>	-æ/-e, -a, -ta

<sup>161</sup> A word-final /n/ originally occurring in the accusative singular of all genders (in nouns of certain declension classes) is subject to sub-variation and may appear in some nouns, e.g., *ton tširin* ‘the father.ACC’ (Mackridge 1987: 124). In the Romeyka corpus, the ending varies according to the speaker (e.g., *ton argo* (05\_03072019M\_4; 07) vs. *ton arkon* ‘the bear’ (02\_2906019F\_1; 05)) but is also often left out for phonological reasons, e.g., *do mustafa.ACC* (B1, C1). Furthermore, there may be even violation of case agreement between article and noun, whereby the article reflects the correct case (ii).

(ii) *esi to faji don aðelfos eðotšes*  
‘You gave the bread to the man.’ (B1, C1)

<sup>162</sup> The evidence is thin here as the corpus contains only very few examples in /u/, e.g., *to zu to ɣlitši* ‘the milk of the cow’ (A1). The questionnaire data only reflect a reduced uninflected form for a proper name, i.e., *t=ahme o daji* ‘Ahmet’s uncle’. Neocleous (2020: 37) notes a zero ending for the genitive singular of masculine nouns, while Özkan (2013: 143) notes -u for the genitive singular of the noun *andras* ‘husband’, but a reduced form without final /s/ for the nouns *babugas* ‘grandfather’ and *vuðias* ‘ox’, which is rare since all three nouns belong to the same class. The masc.gen.sg. in -u is well attested in other varieties of AMG, such as Cappadocian, though.

<sup>163</sup> Only attested in some feminine place names in the present corpus, e.g., *almanjan.ACC*, *tšaikaran.ACC*, *yoryoran.ACC*, see also Fn. 161 above, although Neocleous (2020: 37) generalizes it for the feminine singular accusative.

<sup>164</sup> Does not appear in all cases of the Romeyka corpus (cf. 02\_02022015F\_1; 051), especially, when a possessive suffix is attached (02\_02022015F\_1; 019, but cf. 07\_04072019F\_5; 12). Neocleous (2020: 38) suggests a zero ending here.

<sup>165</sup> But note the rare genitive plural form in *tsi*. GEN *mamikandus* ‘of the grandmothers’ (04\_01072019F\_2; 146).

<sup>166</sup> An -i ending for neuter genitive forms of the 2<sup>nd</sup> declension (i.e., in -on) is indicated by Mackridge (1987: 124; also Özkan 2013: 144 for *ospi* ‘house’), e.g., *tu provatı* ‘the sheep’s’ but does not clearly figure in the examples of the Romeyka corpus (but cf. 08\_04072019M\_3; 178). In Cappadocian, a neuter genitive singular ending in -u would be expected for 2<sup>nd</sup> declension masculines (M. Janse, p.c.).

**Table 21:** Examples of nominal declension in Romeyka<sup>167</sup>

		SG	PL	
M	NOM	<i>o argo / argos</i>	<i>i artši/ t=argoðes</i>	‘the bear’
	ACC	<i>ton argo(n)</i>	<i>t=argo-ðes</i>	
	GEN	<i>tu argo/(u)</i>	<i>tsi argo-ðes</i>	
F	NOM	<i>i mana</i>	<i>i mana-ðæs</i>	‘the mother’
	ACC	<i>ti mana</i>	<i>tsi mana-ðæs</i>	
	GEN	<i>tsi mana-s</i>	<i>tsi mana-ðæs</i>	
N	NOM	<i>t[o]=osbi(d(i(n)))</i> <sup>168</sup>	<i>t[a]=osbid-æ</i>	‘the house’
	ACC	<i>t[o]=osbi(d(i(n)))</i>	<i>t[a]=osbid-æ</i>	
	GEN	<i>tu=spidi</i>	<i>t[a]=osbid-æ</i>	

- (10) *o tšobanis da kedžia=and eðotše ena*  
 the.M.NOM shepherd.M.NOM the.PL goat.PL=POSS.3SG gave.3SG INDET.ACC  
*zengini auro*  
 rich.ACC man.ACC  
 ‘The sheperd sold his goats to a rich man.’ (C1)

- (11) *as eberum hado ton arkon*  
 OPT take.1PL DEM the.M.ACC bear.M.ACC  
 ‘Let us take this bear!’ (02\_2906019F\_1; 05)

- (12) *tsi patsi-s=im to ketši*  
 the.F.GEN daughter.F.GEN=POSS.1SG the.NOM goat.NOM  
 ‘my daughter’s goat’ (A1)

Nominative is the case assigned to grammatical subjects, which may have different semantic roles, including agent and experiencer. The accusative is used for both the direct and the indirect object (corresponding, among others, to the semantic roles of patient and recipient), unlike in other Greek dialects where the genitive is used to mark the indirect object (Mackridge 1987). Some elicited examples from the corpus suggest that the form of the definite article of direct and indirect object of one sentence is not the same to allow for a better distinction between both objects (see 13, 14); more research is required here. The genitive expresses mainly adnominal possessors within a NP. It is not evident that verbs govern certain cases but certain prepositions such as locative *s* ‘to, at’, *as* ‘from’ (15) and instrumental *me(d)* ‘with’ seem to govern accusative case which becomes evident at least in the article of masculine and feminine singular nouns (16). However, it is not straightforward to determine the conditions for the occurrence of the masculine and feminine accusative article with or without final /n/, i.e., *to(n)*, *ti(n)*. Probably phonological conditions play a role so that the form *tin* occurs [ \_V] but also [ \_C], and *ti* [ \_m]. For masculine nouns animacy may play a role, e.g., *son doxtorin* ‘to the doctor’ vs. *aso ðormo* ‘from the road’. Otherwise in masculine nouns starting in a vowel, the article occurs in the

<sup>167</sup> The data for the feminine forms stem from RSür as spoken in Beşkøy based on Özkan (2013: 143–144). The articles in this table reflect otherwise those of ROf as spoken in Caykara which appear in the present corpus. For the considerable micro-variation in definite articles cf. Section 3.2.1.1.

<sup>168</sup> For the variation in nominal endings see Section 2.3.5.2.

reduced form and merges with the preposition, e.g., *s=oros* ‘to the forest’ (Section 2.3.5.3, 3.2.5). Note that articles in other cases merge with the preposition as well, e.g., *su džur=um do mezar* ‘to my father’s grave’ (04\_01072019F\_17; 26). Another intriguing feature is word-final /n/ that occurs in the present corpus occasionally in accusative feminine place names after prepositions, while it is listed as general inflectional ending of feminine accusatives in Neocleous (2019: 37); see also Table 20, Fn. 163.

- (13) *i inetš to faji ton arđobo đokane*  
 ‘The women gave the bread to the man.’ (B1)
- (14) *ton para do mustafa bortš eđoka*  
 ‘I lend the money to Mustafa.’ (C1)
- (15) *ama ebebuka asi yoryoran ebebae udž en*  
 ‘But from lower Gorgoras, it is not the upper.’ (01\_04022016F\_1; 052)
- (16) *sin trabezunda me din araba bas*  
 ‘You go to Trabzon by car.’ (08\_04072019M\_1; 096)

While proper names underly principally the same inflectional paradigm, few particularities in case forms of proper names are discussed in Section 4.2.3.

#### 4.2.2 Possession

This section deals with adnominal possession (for pronominal possession see Section 3.2.2.3; for predicative possession see Section 5.2.1.5.5). Adnominal possessive marking in Romeyka is realized generally by (i) prenominal dependent-marking genitives of the type GenN (genitive preceding the head noun, Dryer 2013f). Some rare examples involving feminine weak possessive pronouns potentially suggest also a postnominal strategy (ii).

Romeyka does not have possessive classification and it is not distinguished between alienable/inalienable possession or other features such as animacy or referentiality.

##### (i) prenominal genitives

Genitives of the type GenN are dependent-marking, thus marking the possessor. The modifying noun occurs in prenominal position of the head. Both nouns require a definite article whereby the possessor is marked by the genitive article. Genitive inflection on the noun is the only case that shows different forms for the three genders, whereby there is a lot of (micro-)variation in the endings (Section 4.2.1.3); for examples see (17–28). The article of the possessee (head) shows agreement according to the function of the entire NP within the clause or PP to which it belongs (gender, number, case); as do additional modifiers that are part of the genitive construction to their individual extent (24). If a genitive construction is governed by a prepositional head, the article of the head noun is accusative (29, 30). The same applies to prenominal genitives that are direct objects (31). Proper nouns seem to be treated like lexical nouns, although the inflectional ending on the proper noun may differ from that of common nouns (18, 20, 21). As the periphrastic formulation in ex. (29) suggests, genitive constructions are possibly uncommon with indefinite NPs.

A possessed NP (i.e., with a possessive pronoun) can itself occur recursively as a preposed possessor, as in (17, 19, 22). Whether multiple possession and NP stacking is possible, is not clear, though (but note multiple possessors in 32 and multiple possesseees in 33). In nonreferential genitives that resemble compounds in English, the same prenominal genitive construction seems to apply (cf. 25).

- (17) *tsi patsis=im to ketši*  
 the.GEN daughter.GEN=POSS.1SG the.NOM goat.NOM  
 ‘my daughter’s goat’ (A1)
- (18) *tsi Aişes ta tria t= ađelfæ*  
 the.GEN Ayşe.GEN the.PL three.PL the.PL = sister.PL  
 ‘Ayşe’s three sisters’ (C1)
- (19) *tsi xala=m i nife*  
 ‘my aunt’s bride’ (02\_02022015F\_1; 019)
- (20) *du veysel i tejze*  
 the.M/N.GEN Veysel.M.GEN the.F.NOM aunt.F.NOM  
 ‘Veysel’s aunt’ (04\_01072019F\_5; 49)
- (21) *t=ahme o daji*  
 ‘Ahmet’s uncle’ (H1)
- (22) *du džur=um do meros*  
 ‘my father’s side’ (02\_02022015F\_1; 023)
- (23) *tu tşoban ta ketşia*  
 ‘the shepherd’s goats’ (C1)
- (24) *du dünja olon da tehlikeliya da dobe*  
 ‘the most dangerous place of the world’ (08\_04072019M\_2; 091)
- (25) *tu zu to ylitşi*  
 ‘the cow’s milk’ (A1 01\_06042017F\_0; 332)
- (26) *tu=spiti do duvari*  
 ‘the house’s wall’ (08\_04072019M\_3; 056–057)
- (27) *tu mileti i ðulia*  
 ‘the people’s work’ (08\_04072019M\_3; 015)
- (28) *du barxari da tşeire*  
 ‘the meadows of the pasture’ (08\_04072019M\_3; 092)
- (29) *asu spidi tin pendžere as enan*  
 from.the.M/N.GEN house.N.GEN the.F.ACC window.F.ACC from INDET.F.ACC  
*pendžere is teri*  
 window.F.ACC somebody look.3SG  
 ‘Out of the house’s window, out of a window, looks a person.’ (05\_03072019M\_3; 06–08)
- (30) *me di mim don beđa*  
 ‘with my uncle’s son’ (08\_04072019M\_1; 154)
- (31) *ahmedis du baxari do ðormo tin aişe erikse*  
 ‘Ahmet showed the pasture’s road [i.e., the road that leads to the pasture] to Ayşe.’ (C1)

- (32) *tsi mana du džuru do lakirdi*  
 ‘mother’s and father’s word’ (exact translation unclear; 02\_02022015F\_1; 051)
- (33) *t=aur i ineka tše ta yarðele*  
 ‘the man’s wife and children’ (C1)

Although genitives are by default prenominal, there seems to be some flexibility in word order and the possessor may be postposed (ex. 34a/b), although this is probably a marked construction resembling an afterthought. (Note also that the postposed possessor in (34b) carries in comparison with (34a) an additional possessive clitic which might be used to increase phonological weight in the post-posed genitive NP.<sup>169</sup>) Further research on genitives in Romeyka is worthwhile, especially in diachronic and synchronic comparison with postnominal genitives in SMG.

- (34) a. *s=aišes tria patsiðæs so mektebin bane* (H1)  
 b. *tria yarðeli so mektebin bane s=aišes=muna* (H1)  
 ‘Ayşe’s three sisters go to school.’

Note that there exist not few deviations from the general rules presented above: Often, the form of the feminine article of the possessor varies (35–37). The form of the article in a heritage speaker (38) is especially noticeable as it could be analysed as the neuter article. Moreover, the article of the possessee may be dropped [<sub>V</sub>] (36; cf. also the construction in 39 that resembles a Turkish compound construction, i.e., *mahalle baş-ı*.POSS.3SG).

- (35) *i fatimes to zo*  
 ‘Fatma’s cow [lit. ‘animal’] (B1)
- (36) *si nifes=muna ø andraðelfo*  
 ‘the brother-in-law of our bride’ (07\_04072019F\_5; 12)
- (37) *omon di hatidžes ti maiðia*  
 ‘like Hatice’s food’ (no. 115/16, 03\_30062019F\_11)
- (38) *t=Aišes ta ðia.. ta tria aðelfæ*  
 ‘Ayşe’s three sisters’ (H2)
- (39) *os tu mehelle ø baši*  
 ‘until the begin of the neighbourhood’ (08\_04072019M\_1; 150)

Moreover, nominalizations as arguments of volitionals require a genitive construction, reminiscent of the corresponding construction in Turkish (40; see Section 4.3.6.4).

- (40) *t=ali t=erθinimo u=θelo*  
 ‘I don’t want that Ali comes.’ (C1)

With non-specific indefinite pronominal possessors (41, 42) or possessives (43), the pronoun or indefinite article is treated as a noun and receives a definite article. However, note that for both constructions loan strategies involving Turkish indefinite pronouns (*kimi* ‘some’, *biri* ‘someone’) are applied as well (44, 45, also 04\_01072019F\_2; 058).

<sup>169</sup> In this example from a heritage speaker, the definite articles of the possessee NP are omitted.

- (41) *du ino do šere*  
 ‘of one the hand’ (04\_01072019F\_2; 059)
- (42) *t=ulinuna t=ospidæ mikra ine*  
 the.GEN=all the.NOM=houses small be.3PL  
 ‘Everybody’s houses are small.’ (A1)
- (43) *ta za to enan*  
 ‘one of the cows’ (08\_04072019M\_1; 110)
- (44) *kimi du kariði*  
 ‘of one the walnuts’ (04\_01072019F\_2; 282)
- (45) *biri d opsari so stoma=nat*  
 ‘one of the fish [is] in his mouth’ (04\_01072019F\_12; 53)

(ii) postnominal possession

A kind of postposed pronominal possessor is potentially possible as well, as shown in (12–14). Although there is only very weak evidence for this construction, this construction seems to consist of the 3SG feminine possessive weak pronoun *ates* together with the feminine accusative definite article; it is only attested for this clitic. Since the full possessive pronouns are historically derived from a merger of definite neuter articles and object pronouns (cf. *t=atines.POSS.3SG.F*), this construction with the accusative article would be remarkable. However, it needs to be mentioned that in part of the examples available (46, 47), the accusative article can also be analysed as unapocopated ending of the noun (see Section 2.3.5.2) in which case the construction would be a normal pronominal possessive with weak possessive pronoun (48, 49) as described in Section 3.2.2.3. The example in (50) remains intriguing; but more data would be required in order to confirm or reject the existence of such a postnominal possessive construction.

- (46) *t=ospi tin ates*  
 the.NOM=house.NOM the.ACC she.GEN  
 ‘her house’ (A1)
- (47) *do ibaded **din** ates utš eksero*  
 ‘I do not know her faith [lit. ‘worship’].’ (01\_28062019F\_2; 43)
- (48) *batsi-ka=tes*  
 ‘her little daughter’ (07\_04072019F\_5; 25)
- (49) *o tširis **ades***  
 ‘her father’ (C1, B1)
- (50) *do džiri **doin ades** s adapazari en*  
 ‘Her father is in Adapazari.’ (07\_04072019F\_8; 11)

There is a second rare candidate for a postnominal possessive for feminines, which might be also related to the post-posed weak possessive pronouns (Section 3.2.2.3), although this particular form does not figure so far. The form of the postposed element resembles at first sight the ablative preposition *as* + the ACC.FEM definite article *tin*. However in (51), *asin* follows the regular 2SG possessive clitic *-s*, while in the construction in (52) – which is probably related to

that in (51) although the form of the post-nominal element is *asi* – no other possessive is used and the meaning of the clause suggests a 3<sup>rd</sup> person reading. Further data are required to verify this construction.

- (51) *i mana=s asin d ađelfe=s asin tšok sever.. bola ayapana=natin*  
 ‘I loved your mother and your siblings much.’ lit. ‘The mother of you, the siblings of you, I loved them much.’ (01\_07072019F\_1; 12)
- (52) *adžaba andras asi layo en*  
 ‘How is actually her husband?’ (01\_28062019F\_2; 35)

### 4.2.3 Definiteness

Definiteness is a property of NPs in a specific context, dependent on certain information about structural constellations that hold in that context. It is widely assumed that definiteness reflects the degree of identifiability of the referent at a particular point in discourse, as assumed by the speaker. A NP can be considered ‘definite’, if the speaker assumes that its referent is identifiable or recoverable by the addressee (Lyons 1999). A NP may be interpretable as definite (identifiable) through either the preceding context (anaphoric definiteness), through the shared speech situation (situational or deictic definiteness), or through shared world knowledge of the interlocutors (inherent definiteness). Furthermore, definiteness can be inferred from the context, even if the particular referent has not yet been mentioned (so-called bridging or associative definiteness). Finally, a NP can be notionally definite by virtue of unique reference, generic reading, deictic reference signalled by demonstratives, proper names, possessives, or the presence of a relative clause.

Definiteness may be systematically marked through some dedicated morpheme (for example an article, or a definiteness affix), or it may be left morphologically unmarked, as in Russian. In Romyka, definiteness is generally assumed to be overtly marked through the set of so-called definite articles; other determiners such as demonstratives and certain quantifiers may also contribute to signalling definiteness, as can the presence of a possessor. The use of definite articles is affected by animacy, case, and specificity (cf. Janse 2004 for Cappadocian). Some nouns are considered inherently definite (e.g., proper names, nouns with unique referents such as ‘sun’). Notably, these also had definite articles in AG, and this remains characteristic for Romyka.

While situative definiteness (possibly including a gesture) is touched upon in Section 3.1.4.3, demonstratives in Section 3.2.2.2, possession in Section 3.2.2.3 (pronominal) and Section 4.2.2 (adnominal), and relative clauses in Section 5.3.4, the present section will focus on (i) masc. lexical nouns in *-os*, (ii) proper names, (iii) anaphoric definiteness, (iv) unique reference. Furthermore, non-definite NPs (or at least NPs that typically lack a definite article in Romyka) are considered: (v) count/mass nouns, (vi) generic nouns, and (vii) indefinites.

#### (i) animate/human masculine lexical nouns in *-os* (/*-as*/*-is*)

The declension of animate masculine nominative singular nouns of the 2<sup>nd</sup> declension in *-os* (and probably that of the 1<sup>st</sup> declension in *-as*/*-is*) appears to be sensitive to definiteness. Drettas (1997) notes the PG particularity of these nouns to use the accusative case ending for the nominative (also Dawkins 1937: 31). This phenomenon has been associated with definiteness: the original nominative ending in *-(o)s* becomes *-o(n)* when the noun is definite, e.g., *škilos*, *o škilo(n)* ‘(the) dog’ (1a; Mackridge 1987: 124; Dawkins 1931: 394, also for other Asia Minor Greek varieties). Karatsareas (2011: 113, esp. Fn. 13) specifies that “definiteness” is associated here with the overt expression of the definite article, i.e., the accusative ending occurs in subject heads only “when preceded by an overtly realised definite article”; if the definite article is zero-



realized or if the indefinite article occurs, the noun is in nominative. This PG particularity has been identified as differential subject marking (DSM) (Janse 2004: 25–26; see also Drettas 1997: 120, but cf. Revithiadou & Spyropoulos 2012). Janse (2004: 14) proposes for Cappadocian a similar development, namely that the (nominative) *-s* ending has been reanalysed as indefinite suffix and consequently led to a “definiteness split” whereby all indefinite NPs (subject and object) receive *-s* (but cf. Karatsareas 2011: 107–115 for a different analysis). According to Dawkins (1931), the use of the accusative ending for definite masc. nouns in *-os* in PG is caused by the merger of the second and third declension class whereby the nominative of the second declension in *-os* is used for indefinite and the nominative of the third declension in *-o(n)* for definite nouns. In this analysis, the phenomenon is caused by a syncretism in nominal inflection that became associated with definiteness. In the present Romeyka corpus, the masculine nouns in question, i.e., those of the 2<sup>nd</sup> declension in *-os* and also of the 1<sup>st</sup> declension in *-as* (especially humans/animates), appear indeed predominantly with the *-os* ending when either indefinite or definite with zero article (53a/b, 54a); however, there are also counter examples where the accusative ending occurs in definite context with zero-article (53c, 54b). If indeed the presence of the definite article should be the decisive factor here, one wonders, though, how a phenomenon that is to a large extent phonologically conditioned (i.e., omission of the def.masc. article *o* [ \_V]; see also Drettas 1997: 112–113) could be linked to definiteness.<sup>170</sup> Furthermore, the citation form of the masculine nouns in question as elicited in word lists seems to be to a certain point random, although the majority of those nouns is listed with the nominative in *-os*, but: *auros* ‘man’ but *andra* ‘husband’, *aðelfo* ‘brother’; *škilo* ‘dog’ but *likos* ‘wolf’, *argos* ‘bear’; *ilos* ‘sun’ but *fengo* ‘moon’.

- (53) a. *ena auros so šehere tšališeve*  
       ‘A man used to work in a city.’  
       b. *auros epije so xorio*  
       ‘The man went to the village.’  
       c. *auro iše so xorio ena ospidi*  
       ‘The man had a house at the village.’ (C1)
- (54) a. *as derume argos n=ard etši mi* (04\_01072019F\_12; 37)  
       b. *as derume argo n=ard etši mi* (04\_01072019F\_12; 34)  
       ‘Let’s see whether the bear will come.’

In sum, based on this weak evidence and the number of counter examples, it cannot be assumed here that Romeyka exploits differential subject marking although a connection to definiteness or specificity (and indirectly to animacy ~ declension class) cannot be ruled out. Further research on this topic is required.

## (ii) proper names

Proper names are definite NPs since they refer uniquely to specific individuals in the world (Schlucker & Ackermann 2017: 310). The general morphosyntactic properties of proper nouns in Romeyka are discussed in Section 3.1.1.2. These include in fact human proper names and toponyms (see Schlucker & Ackermann 2017). In short, the definite article is obligatory for proper names: there is a two-way gender distinction masc., fem. for human referents, and a three-way gender distinction (masc., fem., neut.) for place names; however, the neuter article may occasionally be used in elicited data with human referents as well (e.g., *t=ali ðio yarðelæ iše* ‘Ali had two children’, T1). For phonological reasons, the definite nominative article can be

<sup>170</sup> Note, however, that Janse (2008: 23) claims for Cappadocian that the association of the nominative with indefiniteness due to syncretism with indefinite animate object NPs led to the fact that even definite masc./fem. subject NPs do not take a definite article.

omitted [\_V]. This is possibly reinforced, at least in clause-initial position, by language contact with Turkish (in the corpus especially in feminine names, e.g., *xatitše faji tro tše stetš* ‘Hatice is eating lunch’, T1, elicited example; see also Section 3.1.1.2).<sup>171</sup> Like other nouns, proper names are sensitive to case following the case system described in Section 4.2.1.3. Despite the reduced paradigm of case inflection, case marking is still visible in masculine nouns of the *-is* declension class, to which many frequent masculine proper names (of Turkish/Arabic origin) belong, e.g., *ahmed-is*, *mehmed-is*, *ejüb-is*, *vahid-is*. This section will briefly investigate the patterning of case inflection in these masculine proper nouns in *-is* with special reference to the form of the nominative ending to explore whether proper names can add to the discussion described for lexical nouns in *-os* in (i) above given the fact that they are always definite.

Proper nouns denoting male persons are treated in principle like masculine singular nouns: they take the masculine definite article *o*, and the respective ending in *-s* for nominative, e.g., *o mehmed-is*.M.SG.NOM (see also 55); the definite nominative article is often omitted [\_V] (56; also cf. (i) above). In object function (as well as in accusatives governed by a preposition), we expect a proper noun to drop the nominative *-s* and take the accusative ending in *-o(n)* instead (57–59). Note that the form of the accusative ending is subject to strong variation (*-in*, *-Ø* or even deletion of /t/ in *ahme*), i.e., dialectal, idiolectal and even within the same speaker (60, 61). This variation is in line with that in common nominal inflection; cf. Fn. 161 in Section 4.2.1.3.

- (55) *o vahidis ješlanefde*  
‘Vahit has grown old.’ (04\_01072019F\_5; 37)

- (56) *alis tš=o mustafas epiyan so xorio*  
‘Alis and Mustafa went to the village.’ (C1)

- (57) *o hasanis do kedži=nat do mustafa ši buli*  
‘Hasan sold his goat to Mustafa.’ (B1)

- (58) *ton ali so đormo utš iđame*  
‘We did not see Ali on the road.’ (C1)

- (59) *i aiše mo don ali epije so pazari*  
‘Ayşe went with Ali to the market.’ (C1)

- (60) *ton ahme(t) pal elep tše kahete*  
‘He is seeing Ahmet.’ (A1)

- (61) *me don ali(n)*  
‘for Ali’ (H1, H3)

Interestingly, in (62), the *-s* ending of the nominative is dropped after the definite article. This is in line with what we would expect for masculine lexical nouns (see (i) above); however, compared to the patterning of other proper nouns (55–57), this is rare. Possibly, the omission of the *-s* ending here could be explained by the fact that in (62), the subject is not in subject-position (Schreiber 2018: 903). Papadopoulos (1955: 30, cited in Karatsareas 2011: 114) has suggested that information structure could play a role in the occurrence of the accusative for nominative subjects; however, according to him, the relevant subject NPs would only get the

<sup>171</sup> Note also the lack of the accusative article in an elicited example from a heritage speaker: *alin so pazarin utš iđam* ‘We didn’t see Ali at the market’ (H2). Again, Turkish might have an influence here.

accusative when found in subject position, which is the reverse situation as in (62). Thus, it remains unclear what causes the dropping of the nominative ending in (62).

- (62) *Eđotšen=eme(n) o Mehmet-Ø ato(n).*  
 gave.3SG=OPN.1SG the Mehmet-NOM OPN.3SG  
 ‘Mehmet gave me this/it.’ (Michelioudakis & Sitaridou 2012: 219, ex. 10a; glosses modified)

(iii) anaphoric definiteness

When a NP introduces a new, singular referent into the universe of discourse, it is generally accompanied by the indefinite article *ena* (ex. 63a, 64a/b/e/g in the two little elicited stories below), unless a generic reading is possible (see (v) indefinites, this section, below). For the inflectional marking on nominative masculine heads that has become associated with definiteness, see (i) above.

Discourse-recoverable referents are in principle marked by the definite article, although it can be deleted [ \_ V] as in 63b/c (also cf. Janse 2004: 12), or anaphorically referred to by a personal pronoun (63d); but also note the indefinite pronoun in (64f). Referents modified by a possessive construction are treated as definites, even if they are newly introduced to discourse (63c, 64d (conflicting with 64e)); referents with other modifiers are not necessarily definite, though (64g/h). If the existence of a referent is common knowledge, it is treated as definite (64c). Note that the oblique PPs in (63a/b) are also definite, although they are newly introduced to discourse; this is probably due to a fixed combination ‘in the city/village’ which includes the definite article; it can be considered a case of definiteness based on assumed common ground in the speakers’ world knowledge, i.e., the assumption that speakers are generally aware of the existence and location of any city that can be travelled to.

- (63) a. *ena auros so šehere tšališeve*  
 ‘A man used to work in a city.’  
 b. *auror epije so xorio*  
 ‘The man went to the village.’  
 c. *t=auror ineka tše ta yarðele so xorio jašaeve*  
 ‘The man’s wife and the children lived at the village.’  
 d. *t=aur i ineka tše ta yarðele bola eyapena atona*  
 ‘The man’s wife and the children loved him much.’ (C1)

- (64) a. *aðaha elebo ena ospit*  
 ‘Here, I see a house.’  
 b. *so spit embro ena ðromos*  
 ‘In front of the house, there is a road.’  
 c. *epera ine ta rašia*  
 ‘Opposite are the mountains.’  
 d. *asu spidi tin pendžere*  
 ‘From the house’s window...’  
 e. *as enan pendžere*  
 ‘... from a window...’  
 f. *is teri*  
 ‘... watches a person.’  
 g. *as ena alo pendžere*  
 ‘From an other window...’  
 h. *ðio nomade eterone*  
 ‘... two people watch out.’ (05\_03072019M\_3; 03–10)

Accusative definite masculine nouns in *-os* get the accusative article *to(n)* and, among others dependent on the variety, the accusative ending *-n* (65).<sup>172</sup> However, it seems that in inanimate definite objects with a potential generic reading, the definite article can be dropped (66; cf. 64f for indefinite subjects with a generic reading). Whether this can potentially point at some kind of differential object marking in Romeyka requires further scrutiny (although Cappadocian has developed DOM based on definiteness, Karatsareas 2013). Also, indefinite singular direct objects can appear without article probably subject to the availability of a generic reading (67). Plural nominatives and accusatives that are newly introduced to discourse can appear without article (68). Definite plurals get the plural article in the respective case (69).

- (65) *do milon droyo*  
‘I eat the apple’ (H2)
- (66) *oðen milon dros*  
‘Why are you eating the apple?’ (H1, H3)
- (67) *etšine bal dotš emena milo*  
3SG TOP gave.3SG OPN.1SG apple.ACC  
‘she also gave me an apple.’ (C1)
- (68) *ebarete mila*  
‘Do you buy apples!’ (C1)
- (69) *ta mila ðokan ti mana=tuna*  
‘They gave the apples to their mother.’ (C1)

(iv) unique reference

Nouns of unique reference like ‘sun’ or ‘moon’ differ from other nouns in that they do not pluralize. Otherwise, they (i) often appear with the definite article (70–72; note reanalysis of the article as part of the noun in 72), (ii) can appear without article with word-final *-s* (73), (iii) and even appear modified by a non-numeral quantifier like a mass noun (74).

- (70) *o ilo kru tše kahede*  
‘The sun is shining.’ (01\_06042017F\_0; 214)
- (71) *o fengo krui*  
‘The moon shines.’ (05\_03072019M\_3; 34)
- (72) *do ofengos ekseve*  
‘The moon rose.’ (04\_01072019F\_12; 39)
- (73) *aðadžeka ilos bal an jen [...]*  
‘If there is sun here [...].’ (08\_04072019M\_3; 130)
- (74) *ai bola ilos ebidžen*  
‘It made such a lot of sun.’ (03\_30062019F\_6; 47)

<sup>172</sup> Note that the examples (65–69) are not actually from a “context” but were rather separately elicited which is obviously less reliable for the purpose of this section, but the contextual stories did not cover (inanimate) objects.

## (v) Definiteness and the count/mass distinction

There seems to be no systematic formal distinction between mass and count nouns in Romeyka. Mass nouns referring to substances such as water, air, or wood can appear as definites (75–79); they can be modified by quantifiers and adverbs of degree (80 vs. 81). The noun *ksilo* ‘wood’ can be both a mass noun and a count noun, whereby the count noun refers to identifiable pieces of the mass. As such, it carries an indefinite (82) or definite (83) article. Mass nouns can pluralize (84). Like the last NP in ex. (84), nouns without article have an indefinite, non-specific reading, both in plural and singular (85).

- (75) *o havas temizi*  
‘The air is clean.’ (08\_04072019M\_1; 239)
- (76) *o havas vrað*  
‘The climate is humid.’ (08\_04072019M\_3; 127)
- (77) *[...] temiz havani ula beris=ata*  
‘[...], fresh air, you take them all.’ (08\_04072019M\_2; 119)
- (78) *avu do kalo do hava*  
‘This good air.’ (08\_04072019M\_3; 124)
- (79) *to nero=suna akev*  
‘Does your water flow?’ (03\_07072019F\_1; 48)
- (80) *valis eliyo nero*  
‘You put a bit of water.’ (03\_30062019F\_11; 101)
- (81) *med eliyo do nero*  
‘with little water’ (03\_30062019F\_11; 092)
- (82) *efidže so ðormo ban ena ksilo*  
‘He lost a (piece of) wood on the road.’ (04\_01072019F\_13; 41)
- (83) *taxelevdjēs so ksilo*  
‘She stumbled over the (piece of) wood.’ (04\_01072019F\_13; 26)
- (84) *ondan sora epejname sa ksila ta ksila=muna tširuðan epiname ksila*  
‘Thereafter, we went to the wood, our wood ran out, we made wood.’  
(04\_01072019F\_1; 174–175)
- (85) *do ylidži ejendune diri*  
‘The milk turned into cheese.’ (03\_30062019F\_1; 20)

## (vi) generic nouns

Nouns with an intended generic sense can carry the definite article (86, 89). The masc./fem. plural article is often omitted [<sub>V</sub>] (87, 88). These nouns can be either singular or plural; a noun in the singular can express a generic meaning, either with the definite article (89, 90) or without (91, 92a). The latter is like the Turkish pattern (92b); see also Section 4.3.1.

- (86) *ta batsiðæs edero džesit xizmet exune*  
‘Daughters do a different kind of service.’ (03\_30062019F\_7; 32)

- (87) *aurazis aso pazarin u=fasguntane*  
 ‘Men don’t go to the market.’ (02\_05072019F\_1; 28–29)
- (88) *ayul udž esan edži sade inetš*  
 ‘There were no men there, only women.’ (01\_06042017F\_4; 066)
- (89) *so rðend enivename ja so džörasi ja s=apið ja so milo enivename*  
 ‘We went up the tree. We went up either the cherry tree or the pear tree or the apple tree.’ (02\_02022015F\_1; 165–167)
- (90) *havudijega i fasulija kalo en*  
 here.side the.F.PL bean.SG good COP.3SG  
 ‘The bean(s) of the area are tasty.’ (08\_04072019M\_1; 132)
- (91) *ama har kaini za u=deri, provado u=deri, muskari u=deri, kosara u=deri*  
 ‘But today, nobody keeps animals, a sheep, a calf, a chicken.’ (08\_04072019M\_1; 220–222)
- (92) a. *jaralmasi ekserume*  
 b. *patates cikartıyorus*  
 ‘We take out potatoes.’ (06\_03072019M\_2; 42)

(vii) indefinites

In the singular, non-specific indefinite NPs usually do not carry an article (93–95). This is especially true for existential constructions (96, 97). Both singular and plural nouns with a sortal or generic meaning often lack an article – this is in line with the Turkish pattern (98). This extends to light verb constructions (99, 100), which are also common in Turkish (and other languages of the larger area).

- (93) *emorfo duvari eftes*  
 ‘You make a nice wall.’ (08\_04072019M\_3; 072)
- (94) *etšinos o zias šöferis edune*  
 ‘This Ziya was a chauffeur.’ (01\_04022016F\_1; 023)
- (95) *sahabis na inede*  
 ‘He will be a protector.’ (01\_04022016F\_1; 121)
- (96) *eši kalo mualimina omon esenan*  
 ‘She has a good teacher like you.’ (01\_04022016F\_1; 077)
- (97) *ade beði eš batsi eyapume n=efda*  
 ‘She has a son, we want her to make a daughter.’ (01\_04022016F\_1; 091)
- (98) *epiname ksila odun ederduk*  
 ‘We made firewood.’ (09\_04072019\_7; 02)
- (99) *epiname parakaθ*  
 ‘We chatted.’ (09\_04072019\_7; 20)

- (100) *epiname aryadia*  
 ‘We worked as hired hand.’ (09\_04072019\_7; 42)

Specific, i.e., referential, indefinite singular NPs carry the indefinite article. An NP is referential, for example, if it is modified (101), introduced as a referent to discourse (102, 103), or common knowledge, as in (104, 105), which are elicited by means of a picture story which is at the same time present for both interlocutors (but cf. the unclear function of the ancient numeral *is* ‘one’ in 106). Possibly, grammatical factors, as with the PPs in (107), apply as well. The referentiality of (108) is ambiguous but possibly the indefinite article is required to distinguish the subjects of the change-of-state predicate; cf. (109). The referentiality of exs. (110–112) is also not obvious.

- (101) *omon dž=ebiyam oneon dodže šimu na drano šimu*  
 ‘When we went, suddenly, it rained, a big rain.’ (03\_30062019F\_6; 41)
- (102) *ena peđas ebidže havus*  
 ‘A boy has made a pool.’ (04\_01072019F\_12; 18)
- (103) *eyo ixa ena arđobo*  
 ‘I had a husband.’ (03\_30062019F\_8; 07)
- (104) *ađaha elebo ena ospit*  
 ‘Here, I see a house.’ (05\_03072019M\_3; 03)
- (105) *s=irmani-džega ena inega dž=en auros*  
 ‘At the forest, a man and a woman.’ (04\_01072019F\_13; 03)
- (106) *ađaha is arthobos [...]*  
 ‘Here [is] a man [...].’ (05\_03072019M\_3; 27)
- (107) *ena s=en a sakul espringsam=a mono dže evalam=a s=en a betra bugo*  
 ‘It is in a bag, we squeezed it a bit and put it under a stone.’ (03\_30062019F\_1; 21)
- (108) *ondan sora ejendune ena yarđel*  
 ‘Therafter came a child.’ (01\_06042017F\_4; 123)
- (109) *ejendumune nife*  
 ‘I became a bride.’ (02\_02022015F\_1; 016)
- (110) *asa id eftayo na fai*  
 ‘I make a meal from them.’ (03\_30062019F\_6; 62)
- (111) *na pino ena tšai*  
 ‘I will drink a tea.’ (03\_30062019F\_7; 14)
- (112) *ebero tše bayo ena demligi*  
 ‘I take a teapot and go.’ (08\_04072019M\_1; 265)

Indefinite plural NPs appear without article (113), although they can be modified (e.g., 114).

- (113) *ebero domaditsæ ebero psomi ebero tšaji ebero šekeri*  
 ‘I take tomatoes, I take bread, I take tea, I take sugar.’ (04\_01072019F\_1; 037–040)
- (114) *sa rašia ine bola artši*  
 ‘At the mountains, there are many bears.’ (04\_01072019F\_12; 07)

#### 4.2.4 Diminution

Diminution is optionally expressed by two suffixes on the noun, *-opo(n)* (115–117) and *-ka* (118, 119, also lexicalized in *mamí-ka* ‘grandmother’). Both seem to be available for animate and inanimate heads, apparently even for the same nouns (115, 116 vs. 118); whether they have different distributions needs still to be investigated, although it seems that *-ka* might be preferred in possessives with enclitic pronouns (118). According to Drettas (1997: 168–169, Fn. 1), *-opo(n)* is used both for diminution and endearment. If *-opo(n)* follows a word-final vowel, this vowel is deleted, e.g., *orm-opo* ‘river’ (04\_01072019F\_7; 252) > *ormi* ‘river’, see also (115, 116). Diminution with *-opo(n)* alters the gender of the head noun to neuter (*batsi.F* > *bats-obo(n).N*). It is not clear whether diminution is available for plural as well (but cf. the unclear ex. in 119). Possessive clitics follow the diminutive (118).

- (115) *as ixa nan bats-obon omo hatena [...]*  
 ‘If I only had a little girl like her [...].’ (03\_07072019F\_1; 14)
- (116) *eyo ba exo ena mikro pats-obo*  
 ‘I also have a little daughter.’ (07\_04072019F\_5; 33)
- (117) *ena omorfo ospid-obo en etšis*  
 ‘There is a nice little house.’ (08\_04072019M\_3; 146)
- (118) *batsi-ka=tes ako mikro en*  
 ‘Her little daughter is still small.’ (07\_04072019F\_5; 25)
- (119) *kani o pardži-gas ebuga*  
 ‘I have’ some more’ fields below.’ (translation unclear; 03\_30062019F\_6; 58)

### 4.3 Verbal morphology

Verb stems in Romeyka are formed by the core lexical stem plus person suffixes, e.g., *điy-une.PRS.3PL* ‘they give’. Two verb classes exist based on the set of person/tense suffixes applied. Past tense is marked by the prefixed augment /e/ and often includes stem alternation, e.g., *e-đok-ane.AOR.3PL* ‘they gave’, which is suppletive in some cases, e.g., *ley-une.PRS.3PL* ‘they say’ vs. *ip-ane.AOR.3PL* ‘they said’ but *e-ley-ane.IPFV.3PL* ‘they used to say’. Romeyka finite verbs encode person, tense-aspect-mood (TAM), and voice. Evidentiality is not grammatically expressed and negation is realized on clause level only. Some TAM functions are expressed analytically, for example using particles such as the future marking particle *na*; others are expressed morphologically by affixes and/or stem alternation. Since a big part of Romeyka verbal morphology is expressed analytically and modality appears to be closely intertwined with, for example, negation, these topics are discussed in the present chapter, although they would traditionally belong rather to the syntax chapter. The verbal system of Romeyka contains many archaisms which are absent in other Modern Greek dialects. For example, the AG aorist infinitive is retained, which is a notorious topic in (Balkan) areal linguistic research (Mackridge 1987; Sitaridou 2014b).



### 4.3.1 Person agreement

Romeyka verbs are specified for person and number, conjugation is realized through personal suffixes agreeing with the subject. Subject-agree person marking is not syncretic. The clitic pronouns for objects form a distinct paradigm (Section 3.2.2.1); whether they should be considered “agreement” is an issue that is not taken up here. In what follows, it is exclusively to the subject-indexing suffixes referred to as agreement. The paradigms vary according to verb stems. Some verbs are irregular, i.e., they exhibit different/suppletive stems regarding tense. As examples of suppletive verbs, the paradigms of the copulas *ime* ‘be’ and *inume* ‘become’ as well as of *exo* ‘have’ are given in Table 22–24. Note that *ime* ‘be’ has no aorist form; the imperfective of the verb *inume* ‘become’ (< MedGr *jinoume/jenume* ‘become’, see also Fn. 177 below) is used as aorist of *ime* ‘be’ (Drettas 1997: 247, Fn. 1). Traditionally, the copula *ime* is considered to be a suffix when it appears in post-posed position (see also Section 5.2.1.1 on word order); but this issue is not taken up here further.

**Table 22:** Inflectional paradigm of *ime* ‘be’

	Present	Imperfect
1SG	<i>i-me</i>	<i>emun(e)</i>
2SG	<i>i-se</i>	<i>esin(e)/esun(e)</i>
3SG	<i>e(n(i))</i> <sup>173</sup>	<i>etun(e)</i>
1PL	<i>i-mis(t)</i> <sup>174</sup>	<i>emun(e)(t(in(e)))</i>
2PL	<i>i-st(e(n))/isti(ne)</i>	<i>esun(e(st(ine)))</i>
3PL	<i>i-ne</i> <sup>175</sup>	<i>esane</i>

<sup>173</sup> The 3<sup>rd</sup> person singular copula can be reduced to the vowel /e/, which appears enclitically attached to the NP, e.g., *kseris dohna=e* ‘do you know what it is’ (08\_04072019M\_2; 066). Enclitic forms of the copula (except for 2SG) are already reported for AG (van Emde Boas et al. 2019: 290). The form *eni*, which is also reported for PG by Janse (2020: 214, citing Andriotis 1974: 239) only occurs once in heritage data (H2); for a detailed discussion of 3SG *ine* in modern Greek dialects, especially Cappadocian, see Janse (2020: 214–216).

<sup>174</sup> Tursun (2019: 243) also lists the form *imes(t(ine))* ‘be.1PL’, which resembles the 1PL pronoun *emistine*. Neocleous (2020: 53) lists *imaste* for 1PL and *isaste* for 2PL.

<sup>175</sup> In SMG (in succession of AG, Holton et al. 2019: 1709), *ime* is ambiguous for 3SG/PL (Neocleous, 2017: 53; Giannakou & Sitaridou 2020; also Janse 2020 for Cappadocian). It seems, however, that the 3<sup>rd</sup> person forms are largely differentiated in Romeyka with *ine* for 3PL and *en(i)* for 3SG. But see the discussion at the end of this section.

**Table 23:** Inflectional paradigm of *inume* ‘become’

	Present	Aorist	Imperfective
1SG	<i>inume</i> <sup>176</sup>	<i>e-jend-umun(e)</i>	<i>inumune/enumune</i> <sup>177</sup>
2SG	<i>inese</i>	<i>e-jend-usun(e)</i>	<i>inusune/enusune</i>
3SG	<i>inete</i>	<i>e-jen-tun(e)</i>	<i>inudune</i>
1PL	<i>inumist(ine)</i>	<i>e-jend-umunest(ine)</i>	<i>inumunes(tine)/ enumunest(en)</i>
2PL	<i>ineste/isnestine/intestine</i> <sup>178</sup>	<i>ejendustun(e)</i>	<i>inus(t)unes(t)/enustine</i>
3PL	<i>intane</i>	<i>e-jen-tane</i>	<i>inusan(e)</i>

**Table 24:** Inflectional paradigm of *exo* ‘have’<sup>179</sup>

	Present	Imperfective
1SG	<i>ex-o</i>	<i>ix-a</i>
2SG	<i>eš-is/ej-is</i>	<i>iš-es</i>
3SG	<i>eš-(i)</i>	<i>iše</i>
1PL	<i>ex-um(e)</i>	<i>ixame</i>
2PL	<i>eš-ete</i>	<i>ixate</i>
3PL	<i>ex-un(e)</i>	<i>ixane</i>

The citation form of verbs used here is, in line with Greek linguistic practice, the first person singular present (see also Section 1.5.4). Verbs can be classified according to the endings of the 1<sup>st</sup> person singular present in two verbal inflectional classes: (a) verbs ending in *-o*, e.g., *tro(y)o* ‘eat’, and (b) verbs ending in *-me*, e.g., *tšimume* ‘sleep’ (Table 25 for full paradigms based on own fieldwork). Within these classes, verbs can be subcategorized according to ultimate and penultimate stress, the latter verb class forming the major group (Drettas 1997: 205). (a) Verbs in *-o* are traditionally active stems with the accent either on the ultimate syllable, e.g., *eyabó* ‘love’, or on the penultimate, e.g., *tróyo* ‘eat’, *elépo* ‘see’, whereby originally the accent lay always on the penultimate. (b) Verbs in *-me* are traditionally called pseudo- or medio-passives (or media tantum); these are passive verb stems with an active meaning, e.g., *érxume* ‘come’, *tšimúme* ‘sleep’. They can be differentiated according to the position of the accent in contract (i.e., antepenultimate accent) or non-contract (i.e., (pen-)ultimate accent) stems. Some verbs can belong to both classes, e.g., *xareno* vs. *xarenome* ‘rejoice’ (Tursun 2019: 542–543). In principle, active stems can become passive by adding person endings of the *-me* verb class in which case the accent moves from ultimate to penultimate accent, e.g., SMG *eyabome* ‘I am loved’.

<sup>176</sup> A speaker provides in the paradigm the (future) form *na ino* for 1SG.

<sup>177</sup> Drettas (1997: 247) lists only two forms of the imperfective *inumune* and *enumune* but no aorist. According to him, only the imperative forms starting with /e/ are used as aorist for *ime* ‘be’. In the paradigm in Table 23, the examples highlighted in grey were not found in the present corpus and were completed with data from Tursun (2019: 219), i.e., the imperfective forms in /i/, and Drettas (1997: 247), i.e., the imperfective forms in /e/. Note, however, that the 3<sup>rd</sup> person forms Drettas presents look more like the aorist forms than the imperfective.

<sup>178</sup> While the data from the present corpus show all three varieties, Tursun (2019: 219) lists the 2PL present form as *ineste(n)/inest(ine)* and Drettas (1997: 247) as *inustine*.

<sup>179</sup> The form highlighted in grey was not attested in the present corpus and is taken from Neocleous (2019: 56). Note that the future tense of *exo* ‘have’ was in the elicitations realized with the verb *ime* ‘be’ and the possessive pronoun in a type ‘of me will be X’ (C1, B1).

**Table 25:** Present tense inflectional paradigms of Romeyka verbs in *-o* and *-me*<sup>180</sup>

	Verbs in <i>-o</i>	Verbs in <i>-me</i>
	‘eat’	‘sleep’
1SG	<i>tró-(y)o</i>	<i>tšim-úme</i>
2SG	<i>tró-s</i>	<i>tšim-áse</i>
3SG	<i>tró-(i)</i>	<i>tšim-áte</i>
1PL	<i>tró-yme</i> <sup>181</sup>	<i>tšim-úmíst</i> <sup>182</sup>
2PL	<i>tró-te</i>	<i>tšim-ásten</i>
3PL	<i>tró-yune</i>	<i>tšim-úntane</i>
	‘do/make’	‘come’
1SG	<i>(e)ft<sup>(i)</sup>á-(y)o</i>	<i>érx-ume</i>
2SG	<i>eft<sup>(i)</sup>é-s</i>	<i>ér-se</i>
3SG	<i>eft<sup>(i)</sup>é-(i) / eftá</i>	<i>ér-te</i>
1PL	<i>eft<sup>(i)</sup>á-me</i>	<i>érx-umis(t(ine))</i>
2PL	<i>eftá-te</i>	<i>ér-sti(n(e))</i> <sup>183</sup>
3PL	<i>eft<sup>(i)</sup>á-ne / eftén</i>	<i>érx-untane</i>
	‘love’	
1SG	<i>eyab-ó</i>	
2SG	<i>eyab-ás</i>	
3SG	<i>eyab-á(i)</i>	
1PL	<i>eyab-úme</i>	
2PL	<i>eyab-áte</i>	
3PL	<i>eyab-áne</i>	

Irregular verbs have stems differing in phonologically unpredictable ways, or through suppletion. According to Drettas (1997: 205), they can be divided into two groups: (I) those with an imperfective stem used for present tense, and a deviating perfective (traditionally aorist) stem, e.g., *elepo* ‘see.PRS’, *elepa* ‘see.IPFV’, *iða* ‘see.AOR’ (see also Table 26; for further paradigms see the respective Sections 4.3.2.2, 4.3.2.3); this is the largest group; (II) those with a present stem different from the aorist and imperfect stem, e.g., *eftao* ‘do.PRS’, *epina* ‘do.IPFV’, *epika* ‘do.AOR’ (see also Table 27), which is cross-dialectally very rare and does not exist, for example, in Cappadocian (M. Janse, p.c.).

<sup>180</sup> In Table 25, as well as throughout Sections 4.3.1 and 4.3.2, verb forms of the paradigm that are highlighted in grey are reconstructed and not found in the present corpus. They await verification by actual language data.

<sup>181</sup> Albeit not attested for the verb *troyo*, the 1PL form of verbs of the same class can be reduced, e.g., *kruyom* vs. *krume* ‘we hit’ (05\_03072019M\_4; 06 vs. 08).

<sup>182</sup> Neocleous (2020: 48) lists different forms for the plural forms, i.e., *tšim-úmasten*.PRES.1PL, *tšim-úsaste*.PRES.2PL, *tšim-únde*.PRES.3PL.

<sup>183</sup> This form goes probably back to *erxustine*. Also note the form *eršestin* with is the progressive form with *eš* ‘have.3SG’ + *erstin* ‘come.2PL’ with r-metathesis.

**Table 26:** Tense paradigm of *tro(y)o* ‘eat’

	Present	Aorist	Imperfective
1SG	<i>tro-(y)o</i>	<i>e-fa-ya</i>	<i>e-tro-ya</i>
2SG	<i>tro-s</i>	<i>e-fa-es</i>	<i>e-tro-es</i>
3SG	<i>tro-(i)</i>	<i>e-fa-e(n)</i>	<i>e-trio</i>
1PL	<i>tro-yume</i>	<i>e-fa-yame</i>	<i>e-tro-yam(e)</i> <sup>184</sup>
2PL	<i>tro-te</i>	<i>e-fa-ete</i>	<i>e-tro-ete</i>
3PL	<i>tro-yune</i>	<i>e-fa-yane</i>	<i>e-tro-yan(e)</i>

**Table 27:** Tense paradigm of *pa(y)o* ‘go’

	Present	Aorist	Imperfective
1SG	<i>pa-(y)o</i>	<i>e-piy-a</i>	<i>e-pejn-a</i>
2SG	<i>pa-s</i>	<i>e-pij-es</i>	<i>e-pejn-es</i>
3SG	<i>pa-i</i>	<i>e-pij-e</i>	<i>e-pej(e)n-(e)</i>
1PL	<i>pa-me</i>	<i>e-piy-ame</i>	<i>e-pej(e)n-ame</i>
2PL	<i>pa-te</i>	<i>e-pij-ete</i>	<i>e-pejn-ete</i>
3PL	<i>pa-ne</i>	<i>e-piy-ane</i>	<i>e-pejn-ane</i>

Note that Turkish loan words sometimes do not agree in their number marking with the verbal inflection (120a vs. 120b). The singular form of the noun in (120b) is likely modelled on the Turkish use of generic or categorial nouns where quantity does not play a role and which lack the plural suffix (120c) (“transnumeral” in Göksel & Kerslake 2005: 151; although in the Turkish construction, the verb can be singular as well).

- (120) a. *erxundan da yarðelæ*  
‘The children come.’ (04\_01072019F\_1; 047)  
b. *erxundane misafiri*  
‘Guests come.’ (04\_01072019F\_1; 043)  
c. *misafir gelecek*  
‘A guest/guests will come.’

In (121), verbal and nominal agreement seems incongruent, both on the copula and the object clitic. Likely, the Turkish treatment of categorial nouns as singular has an influence here as well.

- (121) *fasulijas bote en erxume bero=na*  
‘When there are beans, I come and take them.’ (03\_30062019F\_11; 041)

Like in (121), violation of verbal person agreement is frequent with 3<sup>rd</sup> person forms of the copula *ime* ‘be’ (cf. also Fn. 175 above). Although *ine* is otherwise clearly stated as 3PL form in Romeyka, consider exs. (122)–(124) below. It is not clear whether this is continuance from AG (and in line with the situation in other modern Greek varieties where *en(i)/ine* occur for both 3<sup>rd</sup> person forms, Janse 2020: 214; Giannakou & Sitaridou 2020); whether the semantics of the subject suggests plurality (e.g., *tipo na troyume* ‘something to eat’ in 123), or whether

<sup>184</sup> Note regarding the apocope of verbal endings: Since all examples in these paradigms are from the present Romeyka corpus, the shortening of final vowels is only indicated where actual data evidence this so that it may look somewhat inconsistent in the paradigm. In general, apocope of verbal endings can occur for different persons and for all verbs (see also Section 2.3.5.3; also Drettas 1997: 248–249).

Turkish might have an influence (cf. the treatment of Turkish nouns that are modified by a numeral as singular).

- (122) *i dri en tu=spit i arthob*  
 ‘The three are the people of the house.’ (05\_03072019M\_4; 02)
- (123) *lege as dero ine mi dibo na troyume*  
 ‘He said, let me see, whether there is something to eat.’ (04\_01072019F\_12; 11)
- (124) *tema ta peðia=m ba kala en ta peðia=m abora utš ine ta peðia=m ba kala ine*  
 ‘My sons are also good, my sons are not bad, my sons are also good.’ (03\_30092019F\_7; 33–34)

### 4.3.2 Tense

Romeyka has three morphologically marked tenses: present, imperfect, and aorist. The past tenses, i.e., aorist (=perfect) and imperfect, manifest themselves for some verbs also lexically by means of suppletion. Other tenses such as the future are formed analytically (Payne 1997: 237). Remarkably, there is no periphrastic perfect tense like in other modern Greek varieties, e.g., with ‘have’ (Mackridge 1987: 127).<sup>185</sup> Note, however, periphrastic present progressives discussed in Section 4.3.3.1. Although aspect is (especially in the past tenses) not clearly distinguishable from tense, it is discussed separately in Section 4.3.3.

#### 4.3.2.1 Present

Present tense in Romeyka is formed on the basis of the imperfective stem. It expresses an imperfective aspect (125–128) and is also used for habitual actions (129). It is the default/unmarked present tense form irrespective of semantic word classes (130–133). When emphasis is put on actions taking place at the moment, the progressive is used. Note that the motion verb ‘come’ is in four of five questionnaire datasets presented in the progressive as the default present tense, whereas other verb paradigms occur in present tense. Present tense inflection is also used in analytic future marking with the particle *na* (Section 4.3.2.4), in volitional and potential constructions, and with (present tense) serial verbs. For the inflectional paradigm of present tense see Table 25 in Section 4.3.1.

- (125) *si yoryora kahede*  
 ‘She lives in Gorgoras.’ (01\_04022016F\_1; 061)
- (126) *abohendžeka stedžis*  
 ‘Where are you?’ (01\_15022015F\_1; 04)
- (127) *epera ine ta rašia*  
 ‘Opposite are the mountains.’ (05\_03072019M\_3; 05)
- (128) *to vakit ðevas*  
 ‘Time goes by.’ (03\_30062019F\_7; 10)

<sup>185</sup> See Sitaridou (2014a: 122) for the discussion of a complex predicate consisting of the verb ‘have’ in past tense (i.e., *ixa*) followed by an infinitive which resembles a past perfect but functions as counterfactual. It is not clear whether finite forms with ‘have’ in other forms than in 1SG exist. Romeyka has also no periphrastic pluperfect with *ixa* as auxiliary as in Cappadocian, e.g., *ixa iða* ‘I had seen’ (M. Janse, p.c.).

- (129) *her akšam ta za so mandri valo*  
‘Every evening I bring the cows to the stable.’ (C1)
- (130) *ađaha elebo ena ospit*  
‘Here, I see a house.’ (05\_03072019M\_3; 03)
- (131) *terune nunizune*  
‘They look and think.’ (05\_03072019M\_3; 44)
- (132) *permenune ton argo*  
‘They wait for the bear.’ (05\_03072019M\_4; 21)
- (133) *ama hidž uyanaxtume*  
‘We don’t get tired.’ (09\_04072019\_7; 30)

#### 4.3.2.2 Aorist

The aorist is formed on the basis of perfective aspect, i.e., the aorist stem with the ancient temporal augment /e/ (for exemplary paradigms see Table 28), its forms matching largely with the AG aorist endings (Sitaridou 2014b: 53). It is a specialty of PG, that the augment still exists in the whole paradigm, i.e., including the plural (cf. SMG *maθano*.PRS.1SG > *emaθa*.AOR.1SG but *maθame*.AOR.1PL, M. Janse, p.c.). In irregular verbs of group (I) (see Section. 4.3.1), the aorist stem differs significantly from the imperfective stem used for present (see Tables 29, 30).

**Table 28:** Aorist tense inflectional paradigms of Romeyka verbs in *-o* and *-me*

	Verbs in <i>-o</i>	Verbs in <i>-me</i>
	‘ate’	‘slept’
1SG	<i>e-fa-ya</i>	<i>e-tšim-éθa</i>
2SG	<i>e-fa-es</i>	<i>e-tšim-éθes</i>
3SG	<i>e-fa-e(n)</i>	<i>e-tšim-éθi</i>
1PL	<i>e-fa-yame</i>	<i>e-tšim-éθame</i>
2PL	<i>e-fa-ete</i>	<i>e-tšim-éθete</i>
3PL	<i>e-fa-yane</i>	<i>e-tšim-éθane</i>
	‘did/made’	‘came’
1SG	<i>e-pik-a</i>	<i>e-r-θa</i>
2SG	<i>e-pitš-es</i>	<i>e-r-θes</i> / <i>e-r-tes</i>
3SG	<i>e-pitš-e(n)</i>	<i>e-r-θe(n)</i> / <i>e-r-te</i>
1PL	<i>e-pik-ame</i>	<i>e-r-θame</i>
2PL	<i>e-pitš-ete</i>	<i>e-r-θete(n)</i>
3PL	<i>e-pik-ane</i>	<i>e-r-θane</i>

**Table 29:** Aorist tense inflectional paradigm of *elebo* ‘see’

	Present	Aorist
1SG	<i>eléb-o</i>	<i>íð-a</i>
2SG	<i>eléb-is</i>	<i>íð-es</i>
3SG	<i>eléb-(e)</i>	<i>íð-e/i</i>
1PL	<i>eléb-ome</i>	<i>íð-ame</i>
2PL	?	?
3PL	<i>eléb-une</i>	<i>íð-an(e)</i>

**Table 30:** Aorist inflectional paradigm of *diyo* ‘give’

	Present	Aorist
1SG	<i>ði-γo</i>	<i>e-ðók-a</i>
2SG	<i>ði-s</i>	<i>e-ðótš-es</i>
3SG	<i>ði-Ø</i>	<i>e-ðótš-e(n)</i>
1PL	<i>ði-γume</i>	<i>e-ðók-ame</i>
2PL	<i>ði-ti</i>	<i>e-ðótš-ete</i>
3PL	<i>ði-γune</i>	<i>e-ðók-ane</i>

The aorist is used for perfective actions in the past (134).

- (134) *ebije ekopse ksila*  
‘He went and chopped wood.’ (04\_01072019F\_13; 12)

Note that as a hapax legomenon the motion verb *erxume* ‘come’ uses its 2/3SG aorist forms together with the modal particle *na* for a variety of (temporal) functions. Remarkably, the 3SG aorist of *erxume* has two variants *erθe(n)* and *erte* whereby the latter combines with the particle *na* > *n=arde(n)* (135). The form *n=arθe/=arse*, however, is used for the 2<sup>nd</sup> person singular aorist (136, 137). Traditionally in MedGr, the aorist existed (a) as indicative (with the inflectional endings in *-a*, *-es*, *-e*, etc., as presented in Table 28 above) and (b) as subjunctive (with inflectional endings like in present tense: *-o*, *-is*, *-i*, etc.). In MedGr, the modal particle/future marker *na* always required the subjunctive. Whether the 3SG aorist form of *erxume* is *erθe* or *erte*.IND.AOR is diachronically subject to normal intra-speaker variation; the subjunctive form is *erθi*. Since Romeyka does not have a morphological aorist subjunctive (see Section 4.3.6.2), *na* takes the aorist indicative which, however, happens to have two variants for 3<sup>rd</sup> person singular whereby only one is used with *na*. But note also that Drettas (1997: 325, ex. 58) glosses *na érte* as PRT + 3SG present tense.

*N=arde/n=arse* is used for imminent actions happening in the near future (135–137) and as complement in ‘whether’-clauses (138) and volitionals (139). *Na* + imperfective is (not only in the case of the verb *erxume*) used for irrealis conditionals (140); see also Section 5.3.3.2.

- (135) a. *i nife=nades pote na=rde*  
‘When will her daughter-in-law come?’ (04\_01072019F\_5; 20)  
b. *sabale na=rde*  
‘She will come tomorrow.’ (04\_01072019F\_5; 22)
- (136) *pote na=rse*  
‘When will you come?’ (unpublished fieldnotes)

- (137) *na=rθe har [...]*  
 ‘When you come, [...].’ (06\_03072019M\_2; 06)
- (138) *na bermenune na erde argos*  
 ‘They will wait, whether the bear comes.’ (05\_03072019M\_4; 15)
- (139) a. *θelis mi n=arse*  
 ‘Do you want to come?’ (B1)  
 b. *θel n=arde*  
 ‘He wants to come.’ (B1)
- (140) *utše n=eborinam dž=efdename don pešgo*  
 ‘I could have not lit the stove.’ (04\_01072019F\_13; 50)

Furthermore, the aorist forms *erde.3SG/erse.2SG* are used without *na* in periphrastic progressives with *eš* ‘has’, i.e., *eš erse* ‘you are coming’ (C1) vs. *eš erde* ‘he is coming’ (C1; also cf. ex. 141); cf. also the negated 2SG imperative *m=erse* ‘Don’t come!’ (C1). Also consider ex. (142) with a future perfective reading and ex. (143) with a temporal subordinate clause and a strange (Turkic-type) nominalization. Although the diachrony of this rare use of an aorist form cannot be clearly established, it seems to be a hapax legomenon with *erxome*, which possibly points at some frozen grammatical form that became from a semantic view associated with imperfective mood. But note also that in SMG, irrealis mood can be formed by the future particle + both present and imperfective stem (with present endings), e.g., *na + erθo* ‘I will be coming’ (M. Janse, p.c.).

- (141) *na deri argos erd etši mi*  
 ‘He will watch, whether the bear came’. (04\_01072019F\_12; 40)
- (142) *ots erde o musafiris eš kliði*  
 ‘Whoever comes as your guest has a key.’ (08\_04072019M\_3; 157)
- (143) *o har je.. hare jeterun d=arde o zaman na bame eban so xorio*  
 ‘As soon as Yeter comes, we will go up to the village.’ (04\_01072019F\_17; 59–61)

#### 4.3.2.3 Imperfect

The imperfect expresses an action which happened continuously or habitually in the past (144–147); for past progressives see Section 4.3.3.1. The imperfect tense has the imperfective stem, i.e., the present stem, e.g., *mairev-o* ‘cook’, *e-mairev-a* ‘I used to cook.IPFV’ vs. *e-maireps-a* ‘I cooked.AOR. Like the aorist, it is (in most cases) formed by the vocalic augment /e/ (Bortone 2009) and has the ‘past’ endings (in *-a*, *-es*, etc.) instead of the ‘present’ endings (in *-o*, *-is*, etc.). As in the aorist, the inflectional paradigm depends on the verb class (Tables 31, 32; see also the Tables in Section 4.3.1; cf. Drettas 1997 for PG).

- (144) *embra dera mektep eleyame okul utš eleyame*  
 ‘In these times, we used to say ‘mektep’, we did not say ‘okul’.’ (02\_02022015F\_1; 043)
- (145) *d=eleyan edžin epiname*  
 ‘We used to do what they said.’ (02\_02022015F\_1; 098)



- (146) *erxusun s=ema*  
 ‘You used to come to us.’ (01\_070702019F\_1; 05)
- (147) *ipe ki eleya ki ej bejuk allah-um emena sin almanjan nesib bisun*  
 ‘She said, “I used to say, oh great God, destine me to Germany.”’ (01\_04022016F\_1; 118–119)

**Table 31:** Imperfective endings in ROF

	Verbs in <i>-o</i>	Verbs in <i>-me</i>
	‘ate’	‘came’
1SG	<i>e-tro-ya</i>	<i>erxu-mune</i>
2SG	<i>e-tro-es</i>	<i>erxu-sun(i)</i>
3SG	<i>e-trio</i> <sup>186</sup>	<i>erxu-tune</i>
1PL	<i>e-tro-yam(e)</i>	<i>erxu-munestine</i>
2PL	<i>e-tro-ete</i>	<i>erθ-esint</i> <sup>187</sup>
3PL	<i>e-tro-yan(e)</i>	<i>erxu-sane</i>
	‘did/made’	
1SG	<i>e-pin-a</i>	
2SG	<i>e-pin-is</i>	
3SG	<i>e-pin-e</i>	
1PL	<i>e-pin-am(e)</i>	
2PL	<i>e-pin-ete</i>	
3PL	<i>e-pin-ane</i>	

**Table 32:** Inflectional paradigm of the verb *kahume* ‘sit’

	Present	Aorist	Imperfective
1SG	<i>kahu-me</i>	<i>e-kats-a</i>	<i>e-kahu-mune</i>
2SG	<i>kahe-se</i>	<i>e-kats-es</i>	<i>e-kahu-sune</i>
3SG	<i>kahe-te</i>	<i>e-kats-e</i>	<i>e-kahu-tune</i>
1PL	<i>kah-umestine</i>	<i>e-kats-ame</i>	<i>e-kahu-munestine</i>
2PL	<i>kah-hustine/kahistine</i>	<i>e-kats-ete</i>	<i>e-kahu-stine</i> <sup>188</sup>
3PL	<i>kah-untane</i>	<i>e-kats-ane</i>	<i>e-kahu-sane</i>

Like the aorist, the imperfective can occur with the modal particle *na* for a variety of modal functions, that is, the subjunctive (ex. 148, 149b for stronger ‘have to’; ex. 150 for weaker ‘should’) or potential (151). In the complement of the volitional in (152), the use of *na* is for phonological reasons not clearly discernible.

- (148) *üsteluk n=ebejna s=oros n=epina xortare*  
 ‘On top of that I had to go to the forest and had to make hay.’ (04\_01072019F\_2; 140)

<sup>186</sup> This form would be expected to be *e-tro-en*, the form *e-trio* results probably from metathesis of *e-tro-i*.

<sup>187</sup> According to Neocleous (2020: 49), the imperfective 2PL ending is *-saste(n)*. In analogy to other imperative forms in the present corpus, however, in the present variety, the full form should be *-sunest(ine)*. The 2PL forms in Table 31 and 32 show some of the considerable variation in the imperfect 2PL forms of verbs in *-me*. (But note that Drettas (1997: 241) also only lists the reduced form in *-(u)stine*.)

<sup>188</sup> Probably a contracted form of *e-kahu-sunestine* (see also Fn. 187 above).

- (149) a. *ta patsiðaes utš ejollaevane n=eðuilevame*  
 ‘They did not send the girls, we had to work.’ (02\_02022015F\_1; 044)  
 b. *ti mana muna jardim n=epiname*  
 ‘We had to help our mother.’ (02\_02022015F\_1; 045)
- (150) *ama aslinda sa komata n=ebejn edži*  
 ‘But actually, it should come till the open spaces.’ (08\_04072019M\_3; 142)
- (151) *tevekela tib=utš eborenam n=epiname*  
 ‘Without asking, we could not do anything.’ (02\_02022015F\_1; 089)
- (152) *utš eθelene buθen (n=)ebejene*  
 ‘She did not want to go somewhere.’ (01\_04022016F\_1; 105)

Note that in the speech of an Istanbulite ROF consultant, imperfect forms were rarely attested. However, unlike stated in Schreiber (2019: 913), the imperfect occurs in the correct environment in ex. (153). Still, in the Romeyka corpus the number of imperfect forms is only approx. 25% of all past tense forms. However, it cannot be concluded solely on the basis of this fact that the imperfect is in decline in Turkish-dominant speakers; its particular environment is probably simply more restricted than that of the aorist. Furthermore, it should be noted that the semantic distribution of aorist and imperfective is also not always clearly made in SMG, since some adverbs allow for both past tenses (M. Janse, p.c.). Further research needs to determine whether the imperfect forms in Romeyka are possibly restricted to some frequent verbs, directing at “frozen” remnants of the imperfect category.

- (153) *mikrina anda emunesten emist pal so mektep epiyame*  
 small when be.IPFV.1PL we TOP to.the school go.AOR.1PL  
 ‘When we were small, we used to go to school.’ (A1)

#### 4.3.2.4 Future

There is no morphological future tense in Romeyka. Future is expressed periphrastically by means of the multi-purpose modal particle *na* (154–156) and the present stem, like in Medieval Greek and unlike in SMG, which uses the particle *θa* (also PG, Drettas 1997: 298). In Romeyka, *na* has, apart from its modal functions extended its use to the marking of future tense (see Section 3.2.4.4 for an overview of all functions of *na*). For RSür as spoken in Beşkøy, Özkan (2013: 147) also mentions a preverbal particle equivalent to *na*, *ha*, used in negated sentences and interrogatives. According to Tursun (2019: 371), *ha* is a variant of *na* with the same functions in ROF as spoken in Çoruk (see also Tursun 2019: 242–243). While *θa* is a grammaticalization going back to *thelo* ‘want’, i.e., *θéli*.PRS.3SG *na* > *θé na* > *θa*, Neocleous (2020: 70) suggests for *na* an etymology going back to *íme* ‘be’, i.e., *én*.PRS.3SG *na* > *na*. If negated, the negation particle precedes the future particle (154b). For counterfactual future in conditionals formed with *na*, see Section 5.3.3.2. Tursun (2019: 371) also exemplifies a periphrastic perfective future formed with *exo* ‘have’ as auxiliary and a participle (157).

- (154) a. *sabahle na pame so düyun*  
 tomorrow PRT go.1PL to.the wedding  
 ‘Tomorrow we will go to the wedding.’ (A1)  
 b. *utše na pao*  
 not PRT go.1SG  
 ‘I will not go.’ (A1)

- (155) *ifedi na kofdum=ada*  
 ‘This year, we will cut them.’ (08\_04072019M\_3; 096)
- (156) *jani gözüz na en tomara na başlaeps so kseraθinimo adode-džes*  
 ‘When it will be autumn, we’ start the drying all together around that time.’  
 (03\_07072019F\_1; 44)
- (157) *an ine n=arse ipe=m=a jia na exo psemeno to faji*  
 ‘If you will come, tell me so that I will have cooked a meal.’ (Tursun 2019: 371)

### 4.3.3 Aspect

Perfective and imperfective aspect are in Romeyka morphologically realized in the past tense indicative forms aorist (Section 4.3.2.2) and imperfect (Section 4.3.2.3) (Mackridge 1987; Sitaridou 2014a). Present and future have no morphological aspectual distinctions (cf. 158). In present tense as well as in past tense, an analytic progressive exists (Section 4.3.3.3 below). Drettas (1997: 336) highlights the role of motion verbs in expressing periphrastic ‘Aktionsart’. Romeyka seems not to have a periphrastic perfect of the Standard Average European (SAE)-type with auxiliary ‘have’ (Haspelmath 2001: 1495) but cf. the use of the auxiliary *ixa* ‘I had’ in conditionals (Section 5.3.3.2). Further research is required to determine how pluperfect (i.e., perfect aspects in past tense) is expressed in Romeyka. Note, however, that a pluperfect with progressive aspect consisting of a finite verb in aorist (or imperfect) + 3<sup>rd</sup> person aorist of *ime* (=edon), that is described by Drettas (1997: 334–335) and also attested in Cappadocian (M. Janse, p.c.), is not clearly attested in the present data (but see Sections 4.3.3.2 and 4.3.5; see also Neocleous & Sitaridou 2022: 13).

Due to the lack of morphological aspect distinctions in present (and future) tense, in the following sections, a focus is laid (in first line) on different periphrastic constructions to express progressive or continuative aspect.

- (158) *t=ospi      tamir      na      inete*  
 the=house   repair   PRT   become.3SG  
 ‘The house will have been repaired.’ (RSür; Özkan, n.d.; glosses added)

#### 4.3.3.1 Phasal aspect

Phasal aspects of inceptive (and likely also completive) are realized periphrastically by the locative preposition + nominalization as complement of a finite predicate like ‘start’ (159–161).<sup>189</sup> Interestingly, locative PPs do not occur with progressives (Section 4.3.3.3) as is cross-linguistically otherwise often the case (cf. Payne 1997: 243; see also the ‘Rhenish’ progressive in German of the type ‘to be at X-ing’). Payne (1997: 244) notes that cross-linguistically there is a frequent connection between aspect and locative/direction marking. This appears to hold true also for Romeyka, although according to Payne (1997: 249), Greek does not show “verbal operations expressing spatial grounding”.

- (159) *ebašlaepse so borbatima*  
 ‘She started walking.’ (04\_01072019F\_13; 30)
- (160) *tšebedži bašlejevum do dovari so piθenimo*  
 ‘Thereafter, we will start making the wall.’ (08\_04072019M\_3; 049)

<sup>189</sup> Note that Sitaridou (2014b: 45, ex. 39) also considers the periphrastic strategy of finite verb + CONJ + finite form of *steko* ‘stand’ as inchoative, rather than expressing durative aspect (see Section 4.3.3.3).

- (161) *gözüz na en tomara na başlaeps so kseraθinimo adode-džes*  
 ‘When it will be autumn, we’ start the drying all together around that time.’  
 (03\_07072019F\_1; 44)

Note, that the same type of analytic construction with locative + nominalization is used for complements of ‘explain’ (162a/b), while nominalizations as complementation strategy otherwise do not include a preposition (cf. Section 5.3.2.2.2).

- (162) a. *i aiše tin batsi=nates ebolise diri so piθenimo* (C1)  
 b. *i aiše tin batsi=nades diri so biθimo erlaeps* (B1)  
 ‘Ayşe explains her daughter how to make cheese.’

#### 4.3.3.2 Lexical aspect

In the Romeyka corpus, several verb forms showing resultative predication were attested. Lexical aspect (or ‘Aktionsart’) is partly marked by morphological means. For example, the resultative (past) participle *-menos* (see Section 4.3.7.3; also Section 4.1.1.2) marks a final state (163, 164), e.g., *pinasmenos* ‘hungry’ > *epinasame* ‘we have become hungry’ (02\_02022015F\_1; 114). It can also function as a predicate adjective (Section 5.2.1.5.2) preceding an auxiliary (165, 166).

- (163) *evridžes=a epsemeno*  
 ‘You find it cooked.’ (03\_30062019F\_11; 113)
- (164) *xamenos evra=na*  
 ‘I found him dead.’ (01\_06042017F\_4; 128)
- (165) *to šerim ekopsa kosmeno en*  
 ‘I cut my hand, it is cut.’ (A1)
- (166) *dolabi taxemeno dune*  
 ‘The cupboard was locked.’ (04\_01072019F\_5; 62)

Like in (165) and (166), predicate adjectives with resultative aspect are formed by the participle + auxiliary *ime* ‘be’; however, some finite verbforms already include aspectual information (167, 168). In (169), it is not clear whether it features a finite passive aorist form of the verb *xalayume* ‘collapse’ (Drettas 1997: 244; see also *xalano* ‘demolish’ in Tursun 2019: 537) or a predicate adjective *xalaen* ‘collapsed’.

- (167) *ta xorafe=muna ula exalayane*  
 the.PL.NOM field.PL.NOM=POSS.1PL all collapse.AOR.3PL  
 ‘Our fields have all collapsed.’ (02\_05072019F\_1; 16–18)
- (168) *o vejselis edžimadune*  
 the.M.NOM Veysel.M.NOM sleep.IPFV.3SG  
 ‘Veysel was sleeping.’ (04\_01072019F\_5; 41)
- (169) *etšiega tu spidi do duvari xala én*  
 ‘The wall of the house there has collapsed.’ (08\_04072019M\_3; 056–057)

The morphology of other verbs involving resultative aspect both transitive (170) and intransitive (171) is still to be investigated. A rather unclear aspectual form of an intransitive verb is also shown in (172).

- (170) *exasu=na*  
‘I lost him.’ (translation unclear, cf. *xano* ‘loose’ > *exasa* ‘I lost’; 03\_30092019F\_8; 29)
- (171) *eskidefidefkete*  
‘She got up/she erected.’ (3\_30092019F\_7; 03)
- (172) *kati ta malia=s elaiskusanie etrežes*  
‘Somehow<sup>?</sup>, your hair waived while you were running.’ (03\_07072019F\_1; 15)

#### 4.3.3.3 Periphrastic progressive and continuative aspect

The progressive reflects a continuous action taking place either in the present or past (i.e., imperfective aspect). Progressive aspect is realized by two strategies: (i) invariant 3SG auxiliary *eš(i)* ‘has’ plus a finite verb in present or imperfect tense; (ii) two finite verbs coordinated by *tše* ‘and’ whereby the second verb has durative meaning and is either *steko* ‘stay’ or *kahume* ‘sit’.

(i) The progressive formed by the auxiliary *exo* ‘have’ in the frozen grammaticalized invariant 3SG form *eš(i)* ‘has’ (see also Drettas 1997: 334)<sup>190</sup> and a finite verb is used in present and past tense. Only the finite verb inflects for tense, person and number in agreement with the subject; for present tense, see exs. (173)–(175); for imperfective see exs. (176), (177). Note that progressive past tense can only be expressed with the imperfective, not with the aorist, as the correction of the speaker in (177) shows (see also Özkan 2013). Future tense in this progressive construction becomes evident only from the context or temporal adverbs and not from future tense marking with *na* (178, 179). Finally for discourse reasons, past actions can be rendered present tense (as is typologically often the case, cf. Payne 1997: 236); see ex. (180).

Progressive aspect with *eš(i)* ‘has’ is frequently applied to motion verbs (e.g., *erxume* ‘come’,<sup>191</sup> *pao* ‘go’, *eyveno* ‘leave’; although the first two also occur in the second progressive form described below) but is also attested for *ðiyo* ‘give’, *eftayo* ‘do’, *inetšizo* ‘marry’, and *xame* ‘die’ (see for examples 173–180).

- (173) *eši ðiyo tin aiše*  
‘I am giving you Ayşe.’ [on the phone] (01\_04022016F\_1; 147)
- (174) *itš=im eš eyven*  
‘My inside is coming out.’ (04\_01072019F\_2, 152)
- (175) *eš eftayo=se faji na tros mi*  
‘I am making food for you, will you eat it?’ (A1)
- (176) *inega=nad a eš epein eroise eruvie dirie so nero son.. s=ormi abes*  
‘His wife was going and fell, she tumbled directly into the water, into the river.’  
(04\_01072019F\_13; 43–44)

<sup>190</sup> But cf. Neocleous (2020: 57) suggesting a relation to the optative particle *as* instead of a form of *exo* ‘have’.

<sup>191</sup> Note that in the elicitation of the verbal paradigm for *erxume* ‘come’, all speakers (except for one heritage speaker) provided the progressive form with *eš* as default present tense.

- (177) *i batsi=m eš epitše [...] eš ebine nife*  
 ‘My daughter was going to marry.’ (01\_06042017F\_4; 055)
- (178) *iki gün sora aso spid=nades eši ba*  
 ‘Two days later, she will be leaving her home.’ (01\_06042017F\_4; 023)
- (179) *to peði=m bal eš inedžis*  
 ‘My son is also going to marry.’ (01\_06042017F\_4; 056)
- (180) *eši bame su mavrea son barxar*  
 ‘We are going to Mavrea’ to the pasture.’ [at the begin of a story] (03\_30062019F\_1; 12)

This kind of progressive with *eš(i)* is cross-dialectally very rare and does not exist in Cappadocian. But note that in PG, *eš* + *k* ‘and’ (*tše* in Romeyka) + present/imperfect is used to express explicit processes which are near to completion (Drettas 1997: 334); a phenomenon that is likely related to the progressive type of Romeyka.

(ii) The second periphrastic progressive construction consists of a finite verb + coordinative *tše* ‘and’ + finite *steko/stekume* ‘stand’<sup>192</sup> or *kahome* ‘sit’ (ex. 181, 182; see also Dawkins 1916: 199; Özkan 2013: 148). Both verbs agree in person and number (for the inflectional paradigms see Table 33). Note that the person ending of the main verb can be reduced to avoid phonologically too heavy VPs (see 3PL forms in Table 33 vs. Table 25). In the past, both verbs are in the imperfective stem (Özkan 2013): The present corpus exhibits rarely examples of this progressive construction in the past, though, but see (183). Compare also an example of Istanbulite ROF in (184) which uses the aorist with the present form of the auxiliary *steko*. The construction is also attested for negative contexts (185).

- (181) *har harekas to milo droyo tše steko*  
 ‘I am eating the apple right now.’ (C1)
- (182) *ate ti ðulia=tis epitše tši hare tšimate tše kahete*  
 she the work=POSS.3SG did.3SG and now sleep.3SG and sit.3SG  
 ‘She has done her work and now she is sleeping.’ (A1)
- (183) *enan imera ipe evreše dže sdedže*  
 one day said.3SG rain.IPFV and stay.3SG  
 ‘“One day”, she said, “it kept raining.”’ (01\_04022016F\_1; 123)
- (184) *to vakit erθe (tše stitš) n=efta to faji*  
 the time come.AOR.3SG (and stand.3SG) PRT=make.3SG the food  
 ‘The time has come, she will prepare the food.’ (A1)
- (185) *u=tšimume tše kahome*  
 ‘I am not sleeping.’ (A1)

<sup>192</sup> Both verb stems are attested for *steko* ‘stand’: a verb stem in *-o*, *steko*, and one in *-me*, *stekume*. They even seem to constitute a mixed paradigm in themselves since Table 33 has been elicited from a single speaker (A1). Further research on this is in order; note that Tursun (2019: 455) only lists the verb stem in *-o*.

**Table 33:** Periphrastic present progressive forms with *kahome* ‘sit’ and *steko/stekume* ‘stand’ on the example of *tšimume* ‘sleep’<sup>193</sup>

	<i>kahome</i> ‘sit’	<i>steko/stekume</i> ‘stand’
	‘sleep’ + ‘and’ + ‘sit’	‘sleep’ + ‘and’ + ‘stand’
1SG	<i>tšimume tše kahome</i>	<i>tšimume tše steko/stekume</i>
2SG	<i>tšimase tše kahese</i>	<i>tšimase tše stetšis</i>
3SG	<i>tšimate tše kahete</i>	<i>tšimate tše stitš</i>
1PL	<i>tšimumist tše kahomist</i>	?
2PL	<i>tšimaste tše kahesten</i>	?
3PL	<i>tšimun tše kahontane</i>	<i>tšimun tše stekune/stekontane</i>

Furthermore, some examples of an Istanbulite speaker feature a prolonged form of the coordinating conjunction *tše-ste* ‘and’ (186, 187); the analysis of this form is unclear, though.

(186) *to miam poni tše (ste) kahete*  
 ‘The stomach hurts.’ (A1)

(187) *Ate-han d=efte tše ste kahete*  
 ‘What is he doing?’ (A1)

Word order in this progressive construction varies between pre-dominantly verb-final (188, 189) and verb-second (190; cf. Section 5.2.1.4).

(188) *ade kadi tro tše stitš*  
 ‘She is eating something.’ (B1)

(189) *tin para do mustafa odüntš to ðothimo düšünefkume dže kahume*  
 ‘I am thinking about lending the money to Mustafa.’ (B1)

(190) *hatinin troyun tše stekune faji*  
 ‘They are eating.’ (C1)

Object clitics are attached to the main verb (191, see also ex. 193 below).

(191) *hatšin efjen=a tše stekune*  
 they do.3PL=OPN.3SG and stay.3PL  
 ‘They are doing it.’ (C1)

In free speech, this construction occurs less frequently than the one with *eš(i)* ‘has’ but in one elicited dataset of an Istanbulite speaker it is very prominent. While the other progressive construction is frequent with motion verbs, namely *pao* ‘go’ and *erxome* ‘come’, this construction occurs with a variety of verbs from different classes such as action verbs (*tšališevo* ‘work’, *pleko* ‘weave’, *fero* ‘bring’, *troyo* ‘eat’, *kruyo* ‘hit’, *elepo* ‘see’, *leyo* ‘say’), weather verbs (*šonizo* ‘snow’, *akevo* ‘flow’), and notably verbs of cognition (*düşünefkume* ‘think’), mental state (*tšimume* ‘sleep’) and emotion (*eyabo* ‘love’); for examples see (192), (193).<sup>194</sup> No examples are attested for stative verbs like ‘sit’ or ‘stand’. For some verbs, both auxiliaries of

<sup>193</sup> For the forms of *tše steko/stekume* marked by a question mark no information was available in my data; the form highlighted in grey is hypothetical (based on a partial elicitation).

<sup>194</sup> Note that Sitaridou (2014b: 44, Fn. 17) analyses this construction as inceptive construction with inchoative aspect equivalent to ‘start’ but not as progressive.

this construction seem to be used interchangeably (namely, *pa(y)o* ‘go’, *elepo* ‘see’, *tšimume* ‘sleep’, *tro(y)o* ‘eat’, *pono* ‘hurt’); others occur in the present corpus primarily with one auxiliary (194); overall, the auxiliary *steko* ‘stand’ is slightly used more frequent than *kahume* ‘sit’. Note that in Cappadocian, also *tšime* ‘lie’ is used as second finite verb (Dawkins 1916: 199).

- (192) *der d=efte tše stitš eši bai*  
 ‘Look what he is doing, he is going.’ (04\_01072019F\_12; 53)

- (193) *ae kolaieve=da tše steš ama eš erde argo*  
 ‘He is shielding them but the bear is coming.’ (04\_01072019F\_12; 19–20)

- (194) Finite verbs attested only with *steko* ‘stand’:  
*eyabo* ‘like’, *kruyo* ‘hit’, *kollaievo* ‘overwatch’, *eftayo* ‘do’, *erxume* ‘come’, *tašliševo* ‘work’, *leyo* ‘say’, *akevo* ‘flow’

Finite verbs attested only with *kahume* ‘sit’:  
*fero* ‘put’, *pleko* ‘weave’, *düşünefkume* ‘think’, *šonizo* ‘snow’

According to Özkan (2013: 148), this progressive construction is used especially for “longer lasting” progressive actions (see ex. 195), which holds true for many of the attested verbs in the corpus (see, e.g., 196) but not all (<sup>?</sup>*elepo* ‘see’, <sup>?</sup>*leyo* ‘say’).

- (195) *so mutfak tšališevo tše steko*  
 in.the kitchen work.1SG and stand.1SG  
 ‘I am working in the kitchen.’ (A1)

- (196) *šone tše kahete*  
 ‘It is snowing.’ (A1)

This analytic progressive construction possibly replicates a Turkish periphrastic progressive *-(y)Ip + dur-* ‘stay’, like in *çalış-ıp dur-uyor-um* ‘I keep working’ (see also Dawkins 1916: 199 who reports the same construction with ‘stop’ and ‘lie’ in other AMG varieties). In general, verb serialisation of the type finite verb + coordinating conjunction *tše* ‘and’ + finite verb is very frequent in Romeyka (ex. 197a; see Section 5.3.1). This construction is likely a case of pattern borrowing from the Turkish *-(y)Ip* construction used in double verb constructions (197b).

- (197) a. *ebero tše bayo ena demligi*  
 b. *Bir demlik gidip götürüyorum.*  
 ‘I bring a tea pot.’ lit. ‘I go and take a tea pot.’ (08\_04072019M\_1; 265)

Ex. (198) features a less grammaticalized reading of the progressive construction with the verb *exo* ‘have’.

- (198) *hae ul hae ospid exune tše stekune*  
 ‘Like this, they all have houses like this and they stay (at home).’ (03\_30062019F\_7; 48)



#### 4.3.4 Mood

Romeyka exhibits four moods: Indicative as realis mood (*tros* ‘you eat’) and imperative as irrealis mood (*fa* ‘Eat!’) are morphologically marked on the verb whereas subjunctive (*na troyo* ‘I should eat’) and optative (*as troyo* ‘let me eat’) are formed by the particles *na* and *as*, respectively, followed by the indicative form of the verb (Sitaridou 2014a). In the following, imperative morphological mood will be described. For the syntax of imperative clauses see Section 5.2.3.1. Subjunctive and optative mood are treated in the Sections 4.3.6.2 and 4.3.6.3, respectively. Other irrealis modes like potentials, expressions of wish/desire, and deontic expressions are also formed analytically by particles and are therefore outlined in respective sub-sections in Section 4.3.6 as well. For conditional clauses see Section 5.3.3.2.

The imperative covers the second persons and differs according to verb class; for imperative forms of verbs in *-o* and *-me* see Table 34 (see also Drettas 1997: 227–232). Some verbs have suppletive imperative stems like *ayome* ‘go.IMP.2SG’ of the verb *pao* ‘go’ (see also Janse 2020 who states that the form *ayome(n)* is a reanalysed late medieval Greek imperative form), *fa* ‘eat.IMP.2SG’ of the verb *troyo* ‘eat’, *ela* ‘come.IMP.2SG’ of the verb *erxume* ‘come’. Some verbs in Romeyka retain the ancient imperative ending *-(s)on*. These are mostly verbs in *-o*, e.g., *maθison* ‘teach.IMP.2SG’ of the verb *maθizo* ‘teach’, *ilaxsu* ‘bark.IMP.2SG’ (*ilakson*, Tursun 2019) of the verb *ilazo* ‘bark’, suppletive *pisun* ‘do.IMP.2SG’ of the verb *eftayo* ‘do’ and likely also *feron* ‘bring.IMP.2SG’ (*feride.IMP.2PL*) of the verb *fero* ‘bring’, but also verbs in *-me*, e.g., *xasun* ‘remove.IMP.2SG’ of the verb *xame* ‘die, remove, loose’ (see also Bortone 2009: 84). Interestingly, also Turkish loan words ending in *-evo* have adopted the archaic imperative ending, e.g., *jolaepson* ‘send.IMP.2SG’ of the verb *jolaevo* ‘send’ and *arajepson* ‘call.IMP.2SG’ of the verb *arajevo* ‘call, seek’. Other archaic imperatives like *ipé* ‘say.IMP.2SG’ of the verb *leyo* ‘say’ are retained as well (Mackridge 1987: 125).<sup>195</sup>

**Table 34:** Morphological imperatives in Romeyka

	2SG	2PL	NEG 2SG	NEG 2PL
Verbs in <i>-o</i> :				
<i>ðiyo</i> ‘give’	<i>ðos</i>	<i>ðote</i>	<i>mi ðis</i>	
<i>troyo</i> ‘eat’	<i>fa</i>	<i>fate</i>	<i>mi dros</i>	<i>mi drote</i>
<i>pa(y)o</i> ‘go’	<i>ayome</i>	<i>a(yo)miti</i>	<i>mi pas</i>	<i>mi pate</i>
<i>tero</i> ‘look’	<i>de(r)</i>	<i>deride</i>		
<i>epero</i> ‘take’	<i>ebare(n)</i>	<i>ebarete</i>		
Verbs in <i>-me</i> :				
<i>tšimume</i> ‘sleep’	<i>tšimeθ</i>	<i>tšimeθisten</i>	<i>mi tšimaθe</i>	<i>mi tšimasten</i>
<i>kahume</i> ‘sit’	<i>kaθ</i>	<i>kahistine</i>	<i>mi kahese</i>	
<i>erxume</i> ‘come’	<i>ela</i>	<i>elate</i>	<i>m=erse</i>	

Negation is formed by the negation particle *mi* (see Table 34) and – at least in Istanbulite ROF – the present indicative of the second persons. Thus, ROF neutralizes the imperative/indicative distinction under negation.

<sup>195</sup> In this case, word stress is distinctive between *ibe(n).AOR.3SG* ‘she said’ and *ibé.IMP.2SG*. According to Drettas (1997: 229), in PG the imperative form has been reduced to *pé*. This process is also discernible in Romeyka when *ipé* merges with previous elements (e.g., 04\_01072019F\_1; 066; 04\_01072019F\_1; 090).

### 4.3.5 Voice

Active voice is the default voice in which the most controlling and agentive argument (“agent”) is assigned the subject role. In passive voice of transitive verbs, the patient is the subject and the agent an oblique. When the subject is both controller and experiencer of a predicate, this is called middle voice. Middle voice also includes some verbs that do not have a corresponding active form. Drettas (1997) distinguishes for PG “active” from “middle” voice (the latter including “passive”), while others would distinguish “active” from “medio-passive”. According to Janse (p.c.), in AG there was an active, middle and passive aorist, but in PG, there is only an active ~ middle present/imperfect and an active ~ passive aorist left; some verbs are “deponent” or “medium tantum” (only middle forms) and some verbs are mixed, e.g., *jínome* ‘become’, *éjina*.AOR.1SG (active ending), but *ejénton*.AOR.3SG (middle ending). Passive and middle voice in Romeyka (or remnants thereof) are briefly sketched below (for some information on the syntax of passives, see Section 5.2.2.2). Causative voice is not expressed morphologically and is therefore discussed in Section 5.2.2.1. For reflexive voice see Section 3.2.2.4; for reciprocal voice see Section 3.2.2.5.

The details of passive morphology in contemporary Romeyka are yet to be established. Historically, the categorization of verbs in two classes approximately corresponds to a voice distinction: verbs in *-o* are called actives whereas verbs in *-me* are called (medio-)passives (Drettas 1997: 205). Whereas ROF seems to lack passives in general (see Section 5.2.2.2), some passives are attested for RSür (Michelioudakis & Sitaridou 2012: 236). The syntax of datives under passivization in RSür has been investigated by Michelioudakis & Sitaridou (2012); see Section 5.2.2.2.

Morphological passives are attested in ROF merely in the form of some passive verb stems like *skotume* ‘I was killed’ vs. *skotono* ‘kill’ (Tursun 2019; see 199) and *kolizo* ‘burn.CAUS’ vs. *koliyo* ‘burn’ (200), also *ðevazo* ‘spend’ vs. *ðeveno* ‘pass’ (201), although they might not be considered actual passives in the typological sense. Following Greek linguistic tradition, it has to be distinguished between productive inflectional means to mark active (*-o*) and medio-passive voice (*-o/ume*), derivational means (i.e., *-izo* for the formation of causatives) and loan verbs in *-évo* which may include a calque of the Turkish causative suffix *-Dlr* (M. Janse, p.c.). In other cases, active stems seem to be used to express passive meaning, e.g., the active stem *θiko* ‘put’ in ex. 202 (also note the reduced verb ending suggesting violation of verb-noun agreement), while others seem not to allow passivization, e.g., *eksengame eksumer* ‘we took them out’ (03\_30062019F\_2; 28).

- (199) *esgodoθame*  
‘We were killed.’ (03\_30062019F\_6; 45)
- (200) *ta xorafe=muna koliyane*  
‘Our fields burned.’ (03\_30062019F\_6; 45)
- (201) *da imeras ðevazum tše pame*  
‘We spend the days [and go].’ (07\_04072019F\_6; 61)
- (202) *ta xorafe ula ha eθedž ega*  
‘The fields were all put down.’ (03\_30062019F\_6; 54)

Another construction whose analysis is not quite clear is the unaccusative-like use of some verbs in ROF (impersonal passive). Unaccusatives are intransitive predicates with a patient-like argument (Payne 1997: 145). Lexical passives are verbs that are inherently passive and include the presence of a causing agent (e.g., ‘break’); cf. impersonal passives where the agent is omitted, e.g., *har ekabaniefte* ‘now it is closed’ (08\_04072019M\_2; 020). In exs. (203) and

(204), active verbs seem to be combined with the auxiliary *edun* ‘was’ probably to express passivization. Note, however, other occurrences of participles with post-posed *edun* in Section 4.3.3.2, for convenience repeated here below (205, 206). All forms with finite verb in aorist (or imperfect) + post-posed *edon* resemble a kind of pluperfect also reported for Cappadocian (M. Janse, p.c.) and described by Drettas (1997: 334–335) as expressing simultaneity at a referential point in the past.

(203) *du ino do tšefali ški odun* (< *škizo* ‘split.ACTIVE’)  
‘Of one the head was split.’ (04\_01072019F\_2; 058)

(204) *du ino do šere tsakotun* (< *tsakono* ‘break.ACTIVE’)  
‘Of one the hand broke.’ (04\_01072019F\_2; 059)

(205) *dolabi taxemeno dune*  
‘The cupboard was locked.’ (04\_01072019F\_5; 62)

(206) *o vejselis edžimadune*  
the Veysel sleep.IPFV  
‘Veysel was sleeping.’ (04\_01072019F\_5; 41)

#### 4.3.6 Negation and (irrealis) modality

Negation and modality in Romeyka are closely intertwined (see, e.g., Sitaridou 2014a: 138–143) and therefore discussed in a common section. This section is placed on the fringe between morphology and syntax since both negation and moods that are on the irrealis end of the realis-irrealis continuum like subjunctive and optative, but also potentials, expressions of wish/desire and deontic expressions (Payne 1997: 245) are realized analytically using particles.

##### 4.3.6.1 Negation

Romeyka expresses negation analytically on clausal level using negators with rich allomorphic and cross-dialectal variation, whereby the allomorphic distribution of the negators correspond to that of AG (Sitaridou 2014b). The negation particle varies according to clause-type, i.e., declarative clauses take a different particle than imperatives which is cross-linguistically a quite common phenomenon (Payne 1997: 285). While generally invariant, the form of the negation particle used in declarative clauses *utš* varies according to phonological conditions; under some phonological circumstances it can cliticize with the finite verb. The issue of whether the negation particle (or some variants of it) is a proclitic or prefix is not taken up here, though, rather it is followed typological tradition in treating it as a particle. Generally, the negation particle occurs pre-verbally, though a modal particle like *na* can intervene between it and the verb (but not, for example, proclitic object pronouns like in other Greek varieties). The particle may undergo phonological assimilation with the verb, but generally, it does not affect the stress placement in the negated verb. In existential negation, the Turkish particle for existential negation *yok* is used. For an overview of the negation particles see Section 3.2.4, under (i); for the responsive particles Section 3.4.2 (ii). Indefinite pronouns (and spatial adverbs) like *ka(i)nis* ‘some-/nobody’, *tipu* ‘some-/nothing’ and *buthen* ‘some-/nowhere’, that are used in affirmative and negative contexts alike, are described in Section 3.2.2.7. Alongside clausal negation, the use of the Turkish negative adverb *hitš* ‘never/not (at all)’ will be addressed below. NP negation, however, is not attested in the corpus and remains an issue for further research.

The negative particle *utš* is used for sentential negation in indicative sentences (but see discussion on subjunctives below). According to Mackridge (1987: 129), the clausal negator *ki* is used for indicative sentences in ROF; however, it seems *ki* is used as clausal negator only in Romeyka of Tonya (Özkan 2013: 147, Fn. 29) and in PG (Drettas 1997), i.e., only in those dialects that do not have palatalization of /k/ [ \_i]. Note, however, that the palatalized variant *tši* occurs indeed in ROF (see Table 32; see also Tursun 2019: 394). *Utš* (as well as probably *ki*) is derived from the ancient pre-verbal negative particle *ούκι* (involving palatalization). It shows considerable variation of the form *(u)(tš(e))*, dependent on the phonological environment (Table 35; note that most variants can occur as well in clause-initial position as following vowels and consonants, so only the subsequent environment is specified in the table; for a detailed description of the environments in which each variant occurs see below).

**Table 35:** Negation of indicative clauses in Romeyka

Negator		Conditions	Example
<i>utš</i>	<i>u</i>	[ _C]	<i>u θelo</i> ‘I don’t want’
	<i>utš</i>	[ _V]	<i>utš eyabo</i> ‘I don’t like’
	<i>utše</i>	[ _n]	<i>utšé na troyo</i> ‘I wouldn’t eat’
	<i>tše</i>	[ _n]	<i>tše na troyune</i> ‘they will not eat’
	<i>tši</i>	[ _C]	<i>an tši θelide</i> ‘if you don’t like’
	<i>tš</i>	[ _V]	<i>tš alo</i> ‘not more’
<i>xe</i> <sup>196</sup>		subjunctives	(see Fn. 196)
<i>mutš</i>		counterfactuals	<i>na mutš ixa xasini</i> ‘If I had not lost’ <sup>197</sup>

/u/ is used before consonants, e.g., in potentials before *boro* ‘can’ (207). According to Sitaridou (2014a: 121), it is used only after consonants, but the Romeyka corpus features more frequently occurrences after vowels (208) and in clause-initial position (209). According to Tursun (2019: 394), /u/ is used [ \_C] interchangeably with /utši/, e.g., *u=θelo* vs. *utši θelo* ‘I don’t want’. The latter is, however, not attested in the present data (see also next paragraph).

(207) *e manaxos u=boro na tšališevo*  
 ‘I cannot work alone.’ (08\_04072019M\_3; 052)

(208) *ate u=xa*  
 ‘She does not remove [them].’ (01\_28062019F\_3; 04)

<sup>196</sup> *Xe* is listed as negator for subjunctives in Sitaridou (2014a: 123, see ex. iii). Neocleous (2020: 64–69) lists examples for the subjunctive with a negator [ç] + the invariant 3SG copula *en* + the *na*-clause (iv; see also Dawkins 1931: 106). However, the construction in (iv) would be probably more likely the common negator *utš* with the phonologically conditioned variant used [ \_n], i.e., *tše* (whether /n/ actually occurs in the coda of the negator or is rather only present in the onset of *na*, can be only determined from the sound recordings).

(iii) *esi θelis eyo xe na troyo*  
 you.NOM want.2SG I.NOM not PRT eat.1SG  
 ‘You don’t want me to eat.’ (Sitaridou 2014a: 123, ex. 14a; glosses modified)

(iv) *θaró alís č en na faízi=mas*  
 think.1SG Alis.NOM not be.3SG PRT eat.3SG=OPN.1PL  
 ‘I think that Alis would not feed us.’ (Neocleous 2020: 66, ex. 76; glosses modified)

<sup>197</sup> Example from Chatzopoulou & Sitaridou (2014, ex. 7).

- (209) *u=ksero*  
 ‘I don’t know.’ (01\_04022016F\_1; 035)

/utš/ occurs before vowels, e.g., in existentials before *en* ‘be’ (210) and in predicative possession before *exo* ‘have’ (211). According to Sitaridou (2014a: 121) it occurs clause-initially and after consonants, however, in the Romeyka corpus it occurs very frequently after vowels (see 210). From the variants appearing before vowels, *utš* is the most frequent.

- (210) *har ula šenlige utš en*  
 ‘Now they are unoccupied.’ (08\_04072019M\_2; 034)

- (211) *# utš ixame*  
 ‘We didn’t have [bread].’ [after a long pause] (07\_04072019F\_6; 53)

/utše/ is used before /n/, i.e., the modal particle *na*. Possibly, its selection depends on metrics, distinguishing it from monosyllabic /u/, which also occurs [\_C], but not before /n/, though. Rather, it is more likely that the long form (*u*)*tše* used with subjunctives is reserved for this domain. The fact that /utše/ or /tše/ occurs before *na*-clauses only suggests that (*u*)*tše* is the variant of the negator *utš* used with subjunctives; instead of *xe* as suggested by Sitaridou (2014a: 121) and instead of the construction [ç] + *en*.PRS.3SG ‘is’ + *na*-clause suggested by Neocleous (2020: 66). In this case, however, subjunctives would take the ‘default’ clausal negator *utš* and not a particular negator like in SMG where subjunctives get *na* + *min*.<sup>198</sup> The form *utše* appears clause-initially (212) and after both vowels (213) and consonants (214; cf. Sitaridou 2014a: 121).

- (212) *utše ne borum tš=erxumesini*  
 ‘We won’t be able to come.’ (07\_04072019F\_5; 16)

- (213) *eyo udže na bayo*  
 ‘I will not go.’ (01\_04022016F\_1; 112)

- (214) *bola xastas en utše ne borum tš=erxumesini*  
 ‘He is very sick, we won’t be able to come.’ (07\_04072019F\_5; 16)

/tše/ is used in similar environments as *utše*, albeit after a vowel (215) or clause-initially (216).

- (215) *i faji tše na troyune u=θelunje*  
 ‘They will not eat the food, they don’t want.’ (08\_04072019M\_2; 163)

- (216) *tše ne borum tš=erxumesine*  
 ‘We cannot come.’ (07\_04072019F\_5; 17)

It should be noted that there is probably a form of the negator, /he/, used with *na*-clauses (see 217 and possibly also 218). Tursun (2019: 231) also lists this negator in the form /the/ or /ðe/ for RO<sub>f</sub> as spoken in Saráchos (Uzungöl) (219) and in the form /he/ in RO<sub>f</sub> as spoken in Çaykara and Karaçam (320); according to the lemma used for the dictionary entry, he links it to SMG *ðe(n)*. Although this is very weak evidence here, it resembles the form of the modal

<sup>198</sup> This is not in line, though, with what Mackridge (1987: 128–129) notes, namely that RO<sub>f</sub> distinguishes indicative and subjunctive in terms of negation by the use of the SMG form *na/as* + *min*. This kind of negation is, however, not attested in the present data (nor mentioned by Neocleous 2020).

particle *na*, *ha*, which Özkan (2013: 147–148) mentions for RSür in negated clauses and interrogatives (321). Future research has to determine whether there is a point in here.

- (217) *eyo leyo=se he na troyume*  
 I say.1SG=OPN.2SG not PRT eat.1PL  
 ‘I am telling you that we shouldn’t eat.’ (Sitaridou 2014b: 41, ex. 25b; glosses modified)
- (218) *pudžeka istin he na stin so inegölä joksa so xorio mi na stjetšidi panda*  
 ‘Where will you stay? Will you (not?) come to Inegöl or will you be the whole time at the village?’ (03\_07072019F\_1; 45)
- (219) *tše na troyo*  
 ‘I will not eat.’ (Tursun 2019: 231)
- (220) *he na fori*  
 ‘He will not dress.’ (Tursun 2019: 231)
- (221) *udže ha fdæo ado*  
 ‘I will not do it.’ (Özkan 2013: 147)

Sitaridou (2014a: 121) lists /tši/ (instead of *tše*) which occurs according to her [V\_C]. In the Romeyka corpus, there is only a single token of /dži/ following the conditional particle *an* (222), thus not in indicative mode (see also Drettas 1997: 326). Tursun (2019: 394) lists /tši/ also in examples like *tši pao* ‘I don’t go’ and *tši leyo* ‘I don’t say’. Interestingly, he translates these occurrences of /tši/ by the Turkish aorist *git-me-m* ‘I don’t go’, *söyle-me-m* ‘I don’t say’, which combines tense and mode indicating as well actions without clear temporal reference as hypothetical/potential assertions, instead of the present. Note, however, that the finite verb Sitaridou (2014a: 130) attested with *tši* (after a vowel, ex. 223) is realized in the Romeyka corpus with *u* (ex. 224, following a consonant, though). Probably, there is a difference in the reading of these examples with regard to mode and /tši/ is used predominantly for nonveridical possible assertions, i.e., those that are expressed in Turkish by the aorist present.<sup>199</sup>

- (222) *an dži θelide zamani andž üdž ešide..*  
 ‘If you don’t like, if you don’t have time..’ (08\_04072019M\_4; 05–06)
- (223) *[...] eyo pa tši payo*  
 ‘[...] I do not go.’ (Sitaridou 2014a: 130, ex. 44)
- (224) *jane so piknik etši-mer u=bayo*  
 ‘I don’t go there for a picnic.’ (08\_04072019M\_3; 118)

/tš/ is used before vowels, most frequently the mid-front vowel /e/. It occurs clause-initially (225) and following both consonants (226) and vowels (227). It also occurs after conditional *an* (228), thus being equivalent to *tši* – albeit preceding a vowel. Note that Chatzopoulou & Sitaridou (2014) also report the use of indicative negation particle *tše(n)* with conditionals (229). This is probably in line with the use of *tš(i)* in conditionals found in the Romeyka corpus in exs. (222) and (228).

<sup>199</sup> This assumption is not ruled out by considering the other examples with /tši/ in Sitaridou (2014a), which include potentials and volitionals.

- (225) *dž=eberenefdena=da*  
‘I didn’t like them.’ (03\_30062019F\_11; 056)
- (226) *har kainis tš=en*  
‘Now, nobody is (there).’ (08\_04072019M\_2; 020)
- (227) *đormo tš=edon*  
‘There was no road.’ (08\_04072019M\_2; 063)
- (228) *an dž=eyabenis=a [...]*  
‘If you don’t like it [...].’ (08\_04072019M\_2; 124)
- (229) *an tšen potiziz ta za na psofun*  
‘If you don’t water the animals, they will die.’ (Chatzopoulou & Sitaridou 2014: 32, ex. 23)

Comparing the environments, it seems that the forms of the pattern [(V)CV] (i.e., (*u*)*tše*, *tši*) appear predominantly in aspectual and modal contexts, while *u* is used as ‘default’ indicative negator preceding consonants. This hypothesis, however, may be simply caused by syllabic grounds. But also consider the fact that Chatzopoulou & Sitaridou (2014) report the use of the indicative negation particle *tše(n)* for nonveridical contexts (i.e., directives, also conditionals). The distribution of *utš* and *tš* before vowels seems to be clearly phonologically conditioned, though.

In non-veridical contexts, a variety of different negators is used. The negative particle used for imperatives (230–233) and wishes (234, 235) is *mi(n)* (Sitaridou 2014a: 121). Before vowels, the particle is reduced to *m* (231, 235, 236).<sup>200</sup> In the corpus, it also occurs in counterfactuals with the auxiliary *ixa* ‘I had’ (237), which, according to Sitaridou (2014: 121), are negated by the particle *mutš*, though (see below).

- (230) *mi ranaxtarde*  
‘Don’t become tired!’ (04\_01072019F\_2; 212)
- (231) *iben adona m=eftjes adona hajes*  
‘She said to him, “Don’t do this like this.”’ (04\_01072019F\_13; 21–22)
- (232) *raša exo ahadaha esi mi binis edžeha*  
‘I have a back here, but don’t mount it.’ [saying] (01\_28062019F\_4; 67–68)
- (233) *esi ešis i đio mayula, mi ges=ada don andras*  
‘You have two cheeks, don’t give them to your husband.’ [saying] (01\_28062019F\_4; 85–88)
- (234) *ama laya foyumunesine kanis me lepe=masine*  
but how feared.1PL nobody NEG see.3SG=OPN.1PL  
‘But how we were afraid that nobody should see us.’ (04\_01072019F\_2; 288)
- (235) *iba na m=eftayo šamada apes*  
‘I said, “In order not to make noise inside.”’ (03\_30062019F\_6; 31)

<sup>200</sup> Neocleous (2020: 64) also lists *me(n)* as a variant of *mi(n)*.

- (236) *jane hab=ado zori m=ešide*  
 ‘It should not be to your disadvantage.’ (08\_04072019M\_4; 35)

- (237) *eyo na mi ixa škisenda da ksila [...]*  
 ‘If I had not chopped the wood, [...].’ (04\_01072019F\_13; 53)

*Mutš* is according to Sitaridou (2014a: 121) used to negate counterfactuals. Nevertheless, it does not co-occur with the conditional particle *an* but rather the future particle *na* (ex. 238, Chazopoulou & Sitaridou 2014). *Mutš* is a fossilized form of the MedGr negators *mi + ouk* > *mutš* (with palatalization) (Chazopoulou & Sitaridou 2014). In the Romeyka corpus, there is a single token of *mutš* without particle but rather co-occurring with the Turkish conditional *yok-sa* ‘if not’ (239).

- (238) *na mutš ixa xasini ton paran [...]*  
 ‘If I had not lost the money [...].’ (Chatzopoulou & Sitaridou 2014: 17, ex. 7)

- (239) *[...] joksa mutš eyriga*  
 ‘[...] Or did I not understand?’ (01\_28062019F\_3; 33)

Sitaridou (2014a: 121) mentions *xe* as negator for non-veridical embedded predicates, namely subjunctives (240). The Romeyka corpus contains no data for non-veridical embedded predicates (thus no occurrence of *xe*); anti-veridical volitionals with embedded predicates are negated by *u(tš)* together with either a nominalization or a *na*-clause in the embedded phrase (see 241, 242a/b and ex. 209 in Chp. 4 for different realizations); see also Fn. 196 above.

- (240) *Esi            thelis       eyo       xe       na       troyo.*  
 you.NOM   want.2SG   I.NOM   not   prt   eat.1SG  
 ‘You don’t want me to eat.’ (Sitaridou 14a: 123, ex. 14a; glosses modified)

- (241) *eyo utš eyabo esi n=eftas maiðia*  
 ‘I don’t want you to cook.’ (A1)

- (242) a. *sade jad emena faji utš eðelisa pseθinimo* (C1)  
 b. *sadedže jad emena faji d=opsesme utš eðelisa* (B1)  
 ‘I did not want to cook just for myself.’

Chatzopoulou & Sitaridou (2014) also mention for non-veridical clauses the negator *(a)miðen* in possibility conditionals (243), for which no data exist in the Romeyka corpus.

- (243) ***amiðen*** *paθanis   u       manθanis*  
 if.NEG   suffer.2SG   NEG   learn.2SG  
 ‘If you don’t suffer, you can’t learn.’ (Chatzopoulou & Sitaridou 2014: 17, ex. 6; glosses modified)

Finally, Chatzopoulou & Sitaridou (2014) mention the exhortative negator *tšas* in negative demands (244), which is probably etymologically linked to the optative particle *as*, i.e., *(u)tš + as*. No data exist for this form in the present corpus.

- (244) *Alis tšas erte*  
 ‘Ali should not come.’ (Chatzopoulou & Sitaridou 2014: 29, ex. 19)



In the negation of modal clauses involving a modal auxiliary, variation or contact-influence is more likely. For example, the Romeyka corpus contains an example where the Turkish negative copular construction *degil* ‘not’ + POSS.1SG > *degil-im* ‘I am not’ is borrowed in a deontic construction (245).

- (245) *esi medžburen na pas eyo medžbur **deulum** mila so banemo*  
 ‘You will have to go, I don’t need to go.’ (03\_30062019F\_6; 06)

Sitaridou (2014a: 139) reports that in potentials with *boro* ‘can’, the negator can be omitted and *boro* itself has the meaning of ‘I cannot’, thus, functioning as a negative polarity item when (i) following a vowel at the end of the preceding word (246); (ii) the infinitival complement of *boro* can be left out (for the lack of ‘positive’ potentials with *boro* see Section 4.3.6.5). In the Romeyka corpus, *boro* appears mostly together with the negator (247), even following a vowel (248). However, in a few cases it is omitted (249, condition (ii)), especially in a repeated clause and following a vowel (248, 250, condition (i)).

- (246) *Monaxesa poro na payo.*  
 alone cannot PRT go.1SG  
 ‘I cannot go on my own.’ (Sitaridou 2014a: 139, ex. 83a; glosses modified)

- (247) *manaxos **u=boro** na tšališevu*  
 ‘I cannot work alone.’ (08\_04072019M\_3; 052)

- (248) *so stroma **u=boro** na tšimu **boru** na tšimume*  
 ‘In bed I cannot sleep, I cannot sleep.’ (03\_30062019F\_6; 35)

- (249) ***bori=nata** jani*  
 cannot=OPN.3PL well  
 ‘Well, she cannot do it [lit. ‘them’].’ (07\_04072019F\_5; 28)

- (250) ***u=boro** na steko bola **boro** na steko*  
 ‘I cannot stay much, I cannot stay.’ (03\_30062019F\_6; 02–03)

In ‘neither/nor’ expressions, the Turkish construction with negator *ne* ‘neither/nor’ is borrowed (251a). It seems like in Turkish, no additional negator is required. For the use of *tše* ‘and’ as conjunction (although the element could be possibly also analysed as the negator *utš?*) in (251b) instead of *na* in potential constructions, see Section 4.3.6.5.

- (251) a. *ne ben gelirim, ne sen*  
 ‘Neither I can come, nor you.’  
 b. *ne boro tš=erxume ne esi*  
 ‘You are very far, neither I can come nor you.’ (01\_15022015F\_1; 23–24)

Clausal negation of existentials (Section 5.2.1.3.1) is expressed by the Turkish predicative negator *yok* (252–254). At the same time, the copula is left out, thus the construction being a full copy of the Turkish strategy. Apparently (ex. 254), the Turkish *yok* construction in present tense is also applied to past contexts in Romeyka, although the Turkish past tense marker *idi* can be used in code-switching as well (255). An inherited Greek strategy with a negated copula (in present and past) exists alongside the Turkish copy (254–257).

- (252) *jane hađa so mehele jok*  
 ‘There is none [=school] in this neighbourhood.’ (08\_04072019M\_1; 014)
- (253) *kanis jok*  
 ‘There is nobody.’ (01\_28062019F\_2; 21)
- (254) *alo jok ali tibu dž=esuni*  
 ‘There was nothing else, there was nothing else.’ (04\_01072019F\_17; 15)
- (255) *edotes kilo iši pek jox idi tš=eberename*  
 ‘In earlier times, there was no kilo counting, we didn’t take it.’ (08\_04072019M\_2; 156)
- (256) *okul utš edon*  
 ‘There was no school.’ (08\_04072019M\_2; 003)
- (257) *geniše đomo utš en*  
 ‘There is no broad road.’ (08\_04072019M\_1; 077)

In inverse questions, Tr. *yok* is used as question tag only marked by intonation and not by means of the question particle *mI*, like in Turkish (258).

- (258) *heralda n=arde jok*  
 ‘He will most likely come, won’t he?’ (04\_01072019F\_17; 64)

The Turkish negative adverb *hitš* ‘never/not (at all)’ is borrowed into Romeyka. Like in Turkish (259a), the adverb does not affect clausal negation, which is expressed by the appropriate negator (259–262). However, in a heritage speaker in Germany, *hitš* is used as a negative polarity item without any other negator (263a), probably being associated with the German clausal negator *nicht* (263b). In terms of word order, adverbials like *hitš* (the same holds true for *panda* ‘always’) are usually in pre-verbal position, although they can be postposed (262). If an object pronoun is inserted, it goes to the immediate pre-verbal position while the adverb is clause-initial (261).

- (259) a. *hiç bırak-ma*  
 b. *hiš ma afinis*  
 ‘Never leave it.’ (01\_15022015F\_1; 25)
- (260) *hitš utš enumunesten andalo*  
 ‘We were never together.’ (02\_02022015F\_1; 075)
- (261) *hitš emena utš emnazen*  
 ‘He didn’t resemble me at all.’ (03\_30062019F\_8; 08)
- (262) *jauzluslu du đulias utš epiname hitš*  
 ‘We never did anything on our own.’ (02\_02022015F\_1; 094)
- (263) a. *t=ali d=erθinimo hitš eyabo* (H3)  
 b. *Ich möchte nicht, dass Ali kommt.*  
 ‘I don’t want Ali to come.’

## 4.3.6.2 Subjunctive mood

The periphrastic subjunctive formed by the particle *na* and a finite verb in the present stem (or the imperfect in a *na*-clause governed by an aorist verb, Sitaridou 2014a) is used for a variety of constructions, such as volitionals, conditionals, in complementation and subordination (see also Section 3.2.4.4).<sup>201</sup> There is no aorist subjunctive (but cf. Dawkins 1931: 27 who claims the aorist subjunctive survived in the PG of Christians from Samsun). It is a striking phenomenon that in Romeyka the periphrastic subjunctive with *na* is morphologically simply formed by the present tense indicative and not the aorist form (unlike in SMG) and that a distinction between perfective and imperfective aspect is only realized in the aorist indicative (Sitaridou 2014a: 121); the same holds for other modes, e.g., the optative with *as* + present tense (Section 4.3.6.3).<sup>202</sup> Note that the subjunctive seems to be used in counterfactuals interchangeably with the optative particle *as* (Tursun 2019: 243, 371; also Neocleous 2020: 52). For some examples of the subjunctive with *na*, see (264–269) below. In negative contexts, *na*-clauses are negated by the clausal negator (*u*)*tše* immediately preceding the *na*-clause (for a discussion of alternative analyses, see Section 4.3.6.1). At times, it is difficult to discern whether *na* is used as a future marker or whether it has a modal meaning (270–272). Since the subjunctive is not morphologically formed and the term “subjunctive” is in principle used synonymously with the term “*na*-clause” in complementation (but cf. exs. 265, 266), it is questionable whether the term “subjunctive” is appropriate at all or can be just subsumed under the modal functions of *na*.

- (264) *öretmenis elej masine türki na lede*  
 ‘The teacher told us to sing a folksong.’ (04\_01072019F\_2; 040–041)
- (265) *ta patsiðes utš ejollaevane n=eðuilevame*  
 ‘They did not send the girls, we should work.’ (02\_02022015F\_1; 044)
- (266) *ti mana=muna jardim n=epiname*  
 ‘We should help our mother.’ (02\_02022015F\_1; 045)
- (267) *eyabo n=elebo=se*  
 ‘I want to see you.’ (01\_15022015F\_1; 12)
- (268) *utš eθelena n=andriza*  
 ‘I didn’t want to marry.’ (02\_02022015F\_1; 014)
- (269) *panda si ðulia=muna epejname ti ðulian n=eftjame deine*  
 ‘We always went to work in order to get the work done.’ (09\_04072019\_7; 25)
- (270) *ðo=me                      nero      na      pino.*  
 give.IMP.2SG=1SG    water    PRT    drink.1SG  
 ‘Give me water, (that) I may drink.’ (A1)
- (271) *dohna na leyo alom*  
 ‘What else will/shall I say?’ (03\_07072019F\_1; 21)

<sup>201</sup> Cf. also Revithiadou & Spyropoulos (2021), who assume the existence of different *na* morphemes according to their different functions.

<sup>202</sup> But note the rare form of *na* + come.AOR.2/3SG, i.e., *n=arde*, *n=arse* ‘he/you will come’ (see Section 4.3.2.2).

- (272) *nde na ine*  
 ‘What should it be?’ (03\_30062019F\_11; 061)

#### 4.3.6.3 Optative mood

Optative mood<sup>203</sup> is formed non-morphologically by the particle *as* and the verb in present or aorist tense. It is available for all persons (see Dobrusina, van der Auwera & Goussev 2013 for a typological view). It is used in voluntatives of the ‘let us’-type (273), in wishes of the ‘if only’-type (274), in weak deontic assertions including ability, including questions (275), and counterfactuals (276; cf. also 277, 278). For a more detailed description of the functions of *as* see Section 3.2.4.4. According to Neocleous (2020: 52), optatives can be formed by both *as* and *na*-clauses (also Tursun 2019: 243); see (276).

- (273) *as peftum eka as tšimumist*  
 ‘Let us lay down and sleep.’ (A1)
- (274) *as elebame denandalo*  
 ‘If only we could see each other.’ (01\_15022015F\_1; 22)
- (275) *as erxumi mi*  
 ‘Shall/may I come?’ (H1)
- (276) *na / as emune etši kalo n=etone*  
 ‘If I was there, it would have been good.’ (Tursun 2019: 243)
- (277) *i sevda=m an evriški as efte=a jakaluk*  
 ‘If my love finds it, she will make it a collar.’ [folksong] (04\_01072019F\_1; 092)
- (278) *[...] as đesondani=me psomi*  
 ‘[...] in order that they may give me bread.’ (07\_04072019F\_6; 43)

According to Sitaridou (2014b), counterfactual wishes and exclamatives of the ‘if only’-type can be also formed by *as* or *na* + *ixa* ‘had’ + infinitive (279, 280).

- (279) *as išen porpatesini sa rašiæ*  
 OPT have.IPFV.3SG walk.INF to.the.ACC mountains.ACC  
 ‘S/he should have walked in the mountains.’ (Sitaridou 2014a: 136, ex. 72a; glosses modified)
- (280) *na ixame panini xtisini t=ospít so parxar*  
 PRT have.IPFV.3SG go.INF build.INF the.ACC=house.ACC in.the.ACC pastures.ACC  
 ‘I wish we had gone to build the house in the highland pastures.’ (Sitaridou 2014a: 136, ex. 72b; glosses modified)

<sup>203</sup> The term “optative mood” is adopted here as a typological cover term for the most salient function of the particle *as*. However, it needs to be noted that this term is from a closer Greek linguistic point of view not ideal, since the particle *as* is used in a variety of functions, mostly exhortative. Furthermore, the optative has a specific meaning in the Greek grammatical tradition and the optative as a grammatical category was already lost in the Hellenistic period (M. Janse, p.c.).

Note that (281) features a form *ða* probably used as optative. *ða* is the future marker in SMG and most modern Greek dialects including PG (Drettas 1997: 298), which normally does not figure in Romeyka. But this form could be probably related to variants of *na* such as *ha* in some varieties (Section 4.3.2.4).

- (281) *ða diri ha n=evriški mi tši opsaræ*  
 ‘He shall see whether he will find the fish there.’ (04\_01072019F\_12; 27)

#### 4.3.6.4 Expressions of wish, desire or intention

While the optative (Section 4.3.6.3 above) covers next to voluntatives also other assertions close to irrealis mood, voluntative expressions of wish, desire and intention that are on the realis end of the continuum are realized by the modals *thelo* ‘want’ and *eyabo* ‘like/love’.

These modals are followed by the modal particle *na* and the finite verb (282). Both finite verbs agree generally in person, number (283), and tense (but cf. ex. 284), if the subject of main and complement clause is identical. Otherwise, the complement shows its own subject (285). Note, however, that according to Drettas (1997: 324) in PG the matrix verb can appear in a different tense than the dependent clause (286). Volitionals are in the present corpus found predominantly in present (indicative) tense, but see exs. (284) and (287; from Sitaridou 2014a: 123, ex. 14c) for imperfect (indicative) tense. From SMG perspective, volitionals use either the present (imperfective) subjunctive or the aorist (perfective) subjunctive, depending upon the aspectual meaning. The construction with both modals exists in positive as well as in negative contexts (288). It seems that both modals can be used interchangeably in positive (289a/b) and negative (289c/d) contexts; but cf. Sitaridou (2014a: 141), who notes that native speakers consider the construction with *thelo* in positive contexts as unacceptable and rather prefer *ayapo*. According to Sitaridou (2014a: 123–124), *thelo* can only govern a *na*-clause as complement when negated (287a vs. 287b); according to her findings, this does not hold true for the volitional *ayapo* ‘I love/like’, though, which always requires a *na*-clause.<sup>204</sup> This view seems not to be supported by the present (questionnaire) data, though (see 283b/c, 289b). Further naturalistic data are required to check the distribution of *thelo* + *na* in positive contexts. Negated past tense aorist volitionals with *thelo* require the infinitive, while the verb *eyabo* and all other tenses of *thelo* govern *na*-clauses (Sitaridou 2014a: 134, Table 4; see Sections 4.3.7 and 5.3.2.2.1).

- (282) *eyabo na troyo tiba ama utš exo*  
 ‘I want to eat something, but I don’t have (anything).’ (C1)

<sup>204</sup> From a SMG perspective, this is rare since *thelo* requires always *na*, while the semantic meaning of *ayapo* ‘love’ does not include a reading as ‘like’ and it governs a deverbal NP (M. Janse, p.c.). Since there also does not seem to be a functional split between *thelo* and *eyabo*, it is unclear whether Turkish could have an impact here with the two verbs *istemek* ‘want’ and *sevmek* ‘love/like’, although in the case of Turkish, several other verbal expressions like *beğenmek* and *hoşlanmak* complicate the picture.

- (283) a. *eyabo na=rxume*  
 ‘I want to come.’ (C1)  
 b. *esi θelis n=arse*  
 ‘You want to come.’ (H2)  
 c. *etšinos θel na=rde*  
 ‘He wants to come.’ (H2)  
 d. *u=θelnume na bame*  
 ‘We don’t want to go.’ (01\_04022016F\_1; 141)  
 e. *esist u=θelite na trote*  
 ‘You don’t want to eat.’ (A1)  
 f. *atin utš eyabune na troyune*  
 ‘They don’t want to eat.’ (A1)
- (284) a. *utš eθelena n=andriza*  
 NEG want.IPFV.1SG PRT=marry.AOR.1SG  
 ‘I did not want to marry.’ (02\_02022015F\_1; 014)  
 b. *etšin eyapene n=erxotune*  
 he love.IPFV.3SG PRT=come.IPFV.3SG  
 ‘He wanted to come.’ (H1)
- (285) *Ayapo na tšimaste.*  
 love.1SG PRT sleep.2PL  
 ‘I want you to sleep.’ (Sitaridou 2014a: 124, ex. 15d; glosses modified)
- (286) *sa=lutra eθelesen na epejnen met etšinon*  
 to.the=? want.AOR.3SG PRT go.IPFV.3SG with OPN.3SG  
 ‘He wanted to go to [place name] with him.’ (Drettas 1997: 324, ex. 57, presentation adapted)
- (287) a. *utš eθelna n= emaireva*  
 not want.IPFV.1SG PRT= cook.IPFV.1SG  
 ‘I didn’t want to cook.’ (Sitaridou 2014a: 123, exs. 14c; glosses modified)  
 b. *\*θelo na porpato*  
 want.1SG PRT walk.1SG  
 ‘I want to walk.’ (Sitaridou 2014a: 124, exs. 14e; glosses modified)
- (288) *u=θelo n=eftayo faji*  
 ‘I don’t want to make food.’ (A1)
- (289) a. *ayabai n=arde*  
 ‘He wants to come.’ (C1)  
 b. *θel n=arde*  
 ‘He wants to come.’ (B1)  
 c. *ate u=θel na tšimate*  
 ‘She doesn’t want to sleep.’ (A1)  
 d. *ate utš eyaba na tšimate*  
 ‘She doesn’t want to sleep.’ (A1)

Note that wishes can be also expressed solely by the subjunctive with *na*, i.e., with omission of the modals *θelo/eyabo* (see also 4.3.6.2 above). The examples (290–292) have been elicited by means of the questionnaire.

- (290) *eyo kati na tro ama utš exo*  
 ‘I want to eat something but I don’t have (anything).’ (C1)

- (291) *etšinos n=arde*  
 ‘He wants to come.’ (H3)

- (292) *emis=pal utše na troyume*  
 ‘We don’t want to eat.’ (A1)

As a full verb, *thelo* governs a direct object (293), which can be expressed by object clitics (294). The construction with *thelo* as full verb is applied next to the complementation strategy (295a vs. 295b/c).

- (293) a. *thel faedo*  
 ‘He wants to eat.’, lit. ‘He wants food.’ (B1)  
 b. *thelis mi faedun*  
 ‘Do you want to eat?’, lit. ‘Do you want food?’ (B1)

- (294) *beker ethelesan=ana edera thelesan=a*  
 ‘Unmarried [men] wanted her, others wanted her.’ (01\_04022016F\_1; 106)

- (295) a. *o mexmetis fai u=thel*  
 ‘Mehmet doesn’t want food.’  
 b. *u=thel na tro*  
 ‘He will not eat.’  
 c. *o mexmetis utš eyapai na tro*  
 ‘Mehmet does not want to eat.’ (A1)

Modal *thelo/eyabo* govern both intransitive (e.g., 283 above) and transitive (296) complements. ‘Want’-complement clauses can have their own subject which is not necessarily the same as that of the main clause (297–299), in which case the complement clause has its own nominal agreement; word order in the complement clause seems to be variable (300a/b).

- (296) *eyo esena thelo/eyabo ne lepo*  
 ‘I want to see you.’ (A1)

- (297) *batsi eyapume n=efda*  
 ‘We want her to make a girl.’ (01\_04022016F\_1; 091)

- (298) *o bedino gagani thel din kosaran (n=)erde*  
 ‘The rooster cackles, it wants its chicken to come.’ [folksong] (04\_01072019F\_1; 088)

- (299) a. *eyo n=arxume thelis*  
 ‘Do you want me to come?’ (H2)  
 b. *eyo eyabo n=arse*  
 ‘I want you to come.’ (H2)

- (300) a. *eyo utš eyabo esi n=eftas mairija*  
 b. *utš eyabo esi mairija n=eftas*  
 ‘I don’t want you to cook.’ (A1)

*θelo* and *eyabo* are used in questions both as modals and full verbs. The Turkish interrogative particle, that occurs in nearly all elicited interrogatives (except for one by a German heritage speaker, suggesting a potential interference from German) always follows the modal (301). In the switch-referent construction in (302), which strongly resembles the Turkish model, the interrogative particle follows the resumptive pronoun.

- (301) a. *eyabas mi na kahese*  
           ‘Do you want to sit?’ (C1)  
       b. *θelis mi n=arse*  
           ‘Do you want to come?’ (B1)  
       c. *trano šehirin ne lepome θelis mi* (H1)  
       d. *θelis mi trano šehirin ne lepome* (H1)  
           ‘Do you want to see a big city?’
- (302) *eyo n=erxume θelis=a mi*  
       I PRT=come.1SG want.2SG=OPN.3SG Q  
       ‘Do you want me to come?’, lit. ‘That I come, do you want it?’ (H1)

Apart from the periphrastic construction with *na*, both verbs can also be followed by a nominalized verb form, usually including the definite article (303), but cf. (304a/b, 305). This construction seems to be more frequent with the verb *θelo* (but see 306b). This might be explained by the hypothesis that *θelo* covers predominantly the modal meaning of ‘want’, whereas *eyabo* has more a semantic coverage of ‘like/love’.<sup>205</sup> Otherwise, the construction occurs in the same environments as the periphrastic construction with *na*, i.e., negative (307), and interrogative (308). It seems, however, that the nominalization strategy is in the present questionnaire data preferred to express different subjects of main and complement clause; although generally both nominalizations and *na*-clauses are possible (Sitaridou 2014a). The fact that the nominalized forms are especially frequent in the questionnaire data can be likely ascribed to calques of the Turkish structure (309a, see 309b–d for realizations of the translation task). While SMG expresses complement subjects overtly, Turkish uses an infinitival construction and leaves the subject of the ‘want’-complement implicit when it is coreferential with the wanter (Haspelmath 2013b). The nominalization strategy implies that next to the inherited method with finite complements, the Turkish-strategy is adopted. Interestingly, in the heritage data in (310), both the nominalization and the periphrastic strategy with *na* are combined.

- (303) *aðelfo=s to kliði esena to ðosimo θel*  
       ‘Your brother wants to give you the pencil.’ (B1)
- (304) a. *θel kaθinimo*  
           ‘He wants to sit.’ (B1)  
       b. *kaθinimo θelis mi*  
           ‘Do you want to sit?’ (B1)
- (305) *sade jad emena faji utš eθelisa pseθinimo*  
       ‘I did not want to cook just for myself.’ (C1)

<sup>205</sup> Sitaridou (2014b: 39) mentions that *eyabo* ‘love’ is associated with “non-obligatory control”, i.e., the occurrence of distinct subjects in complement and main clause; this would point to a similar direction.



- (306) a. *teso t=erθinimo θelo* (C1)  
 b. *esi.. teson t=erθinimon eyabo* (H3)  
 ‘I want you to come.’
- (307) *habaḏadžega to panimo eyo utš eyabo*  
 ‘I don’t want to go from here.’ (08\_04072019M\_1; 284)
- (308) *to kalem adena to ḏosimo θelis mi*  
 ‘Do you want to give the pencil to him?’ (B1)
- (309) a. *Ali ’nin gelmesini istemiyorum.*  
 b. *t=ali t=erθinimu u=θelo* (C1)  
 c. *alis-in de n=arde u=θelo* (B1)  
 d. *alis n=erde u=θel adonan* (H1)  
 ‘I don’t want Ali to come.’
- (310) *t=erθinimon eyabas n=arse*  
 ‘You want to come.’ (H3)

#### 4.3.6.5 Potential constructions

Following Sitaridou (2014a: 140), the potential construction with the auxiliary *boro* ‘can’ only exists for negative contexts; the positive potential is covered by present tense, e.g., *mairevo* ‘I can cook’ (Sitaridou 2014a: 142, Table 6).<sup>206</sup> Thus, the auxiliary *boro* is mostly accompanied by a negation particle, although Sitaridou (2014a: 139) argues that *boro* has itself become a negative polarity item and thus the negation particle can be dropped under certain circumstances (311, 312; for a discussion see Section 4.3.6.1). *Boro* usually requires a *na*-clause as its complement (in negated clauses) (Sitaridou 2014a: 123; see also Drettas 1997: 324); but see below for *eboro* used in positive environments and complemented by *tše* ‘and’. Negated past tense (aorist) modals with *boro* (but not questions!), i.e., antiveridical contexts, trigger, according to Sitaridou (2014a), the infinitive, although in the present data, the infinitive only occurs in once instance (313) and otherwise *na*-clauses are selected with negated aorist *boro* (314; see also exs. 30b/d in Neocleous 2020: 349). In potentials with *na*, the modal and the finite verb are congruent in person, number, and usually in tense. Note, however, that in PG, the volitional can be in aorist tense and the dependent verb in the imperfective (Drettas 1997: 325). The potential is in the present corpus attested in present (315–317), aorist (313), and imperfective tense (318).

- (311) *bori=nata jani*  
 ‘She cannot (do it).’ (07\_04072019F\_5; 28)
- (312) *boro na steko ḏa to poḏar=im bone*  
 ‘I cannot stay there, my leg hurts.’ (03\_30062019F\_6; 03)

<sup>206</sup> Note, however, that Tursun (2019: 419) lists positive examples with invariant (!) *pori* + *na* + finite verb (v, vi).

(v) *pori na erxome*  
 ‘I can come.’ (Tursun 2019: 419)

(vi) *pori na pas*  
 ‘You can go.’ (Tursun 2019: 419)

- (313) *utš eboresane dosin adonusine*  
 NEG can.AOR.3PL hit.INF OPN.3PL  
 ‘They could not hit them.’ (04\_01072019F\_12; 15)
- (314) *utš eborise n=eyvali=na*  
 NEG can.AOR.3SG PRT=take.off.3SG=OPN.3SG  
 ‘It could not take it off.’ (02\_29062019F\_2; 28)
- (315) *u=boro na krüy=a u=boro n=exlev=a*  
 ‘I cannot hit her, I cannot be angry with her.’ (01\_28062019F\_2; 30–31)
- (316) *u=boris n=eperis*  
 ‘You can’t buy it.’ (02\_05072019F\_1; 34)
- (317) *u=bori na landžef*  
 ‘It cannot jump.’ (02\_29062019F\_2; 17)
- (318) *tevekela tib=utš eborenam n=epiname*  
 ? something=NEG can.IPFV.1PL PRT=do.IPFV.1PL  
 ‘Without asking, we could not do anything.’ (02\_02022015F\_1; 089)

As for word order, adverbs modifying the whole verb phrase appear in pre-verbal position (319, 320), while adverbs modifying the finite verb occur post-verbally (321, also 320), but also before the *na*-clause (322); see also Sitaridou (2014a: 139), according to whom the adverbs *manaxos* ‘alone’ and *panda* ‘always’ can only appear in clause-initial position.

- (319) *emanaxos u=boro na tšališevu*  
 ‘I cannot work alone.’ (08\_04072019M\_3; 052)
- (320) *teke u=boro na yonuševum emorfa*  
 ‘Alone, I cannot speak nicely.’ (03\_07072019F\_1; 35)
- (321) *u=boro na steko bola boro na steko ōa*  
 ‘I cannot stay long, I cannot stay there.’ (03\_30062019F\_6; 02–03)
- (322) *u=borume alo n=eskumes*  
 ‘We cannot get up anymore.’ (02\_02022015F\_1; 162)

*Boro* can also govern a double verb construction (coordinated by the conjunction *tše* ‘and’, Section 5.3.1.2) in the *na*-clause (323). According to Sitaridou (2014a: 130) even the infinitive can be stacked (324).

- (323) *u=boro na payo tš=erxume*  
 ‘I cannot go (myself).’ (04\_01072019F\_2; 201)
- (324) *Utš eporesa ayapisini almeksini.*  
 not can.AOR.1SG love.INF milk.INF  
 ‘I couldn’t bring myself to liking milking the cows.’ (Sitaridou 2014a: 130, ex. 45a; glosses modified)

Importantly, there exists a second strategy of negative (for positive, see below) potentials with *boro* in the subjunctive, i.e., the particle *na* expresses future tense and the potential requires a complement with *tše* ‘and’ (325). *Tše*-complements with *boro* are in Sitaridou (2014a) only noted in (positive) questions<sup>207</sup> but seem not to appear beyond interrogatives (326). In the present corpus, *tše*-complements of *eporo* occur (a) with negative antiveridical future tense (in 325 with the infinitive!)<sup>208</sup>, (b) in (negative) antiveridical conditionals (327) where *na* + *boro* in imperfective tense express irrealis mood, (c) with negative correlative conjunctions, reminiscent of the corresponding construction in Turkish (328) where Turkish *ne* ‘neither/nor’ is used as clausal negator, and (d) in positive potentials and questions (see next paragraph).

- (325) *utše ne borum tš=erxumesini*  
 ‘We will not be able to come.’ (07\_04072019F\_5; 16)

- (326) *Eporis tše mairevis?*  
 can.2SG and cook. 2SG  
 ‘Can you cook?’ (Sitaridou 2014: 141, ex. 93c)

- (327) [...] *utše n eborinam dž=efdename don pešgo*  
 ‘[...] I could have not lit the stove.’ (04\_01072019F\_13; 50)

- (328) *ne boro tš=erxume ne esi*  
 ‘Neither I can come, nor you.’ (01\_15022015F\_1; 23–24)

Positive potentials with *boro* were found in questionnaire data of one speaker. *Boro* and the finite verb are coordinated by the conjunction *tše* (329–331). The second verb is finite and agrees with the modal. Positive potentials with *boro* are only attested in present tense, although Sitaridou (2014a: 140–141) also notes them for questions both present and past. Note that in the question in (332), optional the optative particle *as* is added to the modal construction.

- (329) *ti manam eporo tš=elepo*  
 ‘I can see my mother.’ (A1)

- (330) *eyo esas eboró tše elebome*  
 ‘I can see you.’ (A1)

- (331) *ton Axme i Aiše epori tše elep*  
 ‘Ayşe can see Ahmet.’ (A1)

- (332) *avudo to saxan (as) eboro tš=epero*  
 ‘Can I take this plate?’ (A1)

Apart from the attested positive potentials with *eporo* in one speaker, in the majority of speakers, positive ability is expressed either by present tense (333–335; see also Sitaridou 2014a), also with progressive aspect (336), or by the optative (337).

<sup>207</sup> Tursun (2019: 420) lists *na*-complements of *eporo* also in positive questions (vii).

(vii) *eporó me n=eftayo ula ta ðulias*  
 ‘Can I do all the work?’ (Tursun 2019: 420)

<sup>208</sup> The infinitive has not been described yet in contexts like in (425) and it is not clear what triggers its occurrence here; probably it occurs here in an antiveridical negated context, albeit in future tense rather than the aorist (but still with perfective aspect).

- (333) *ta mila elep*  
‘He can see the apples.’ (C1)
- (334) *esist emasuna elepide mi*  
‘Can you see us?’ (A1)
- (335) *do xali bola tšabuxe pletš*  
‘She can weave carpets quickly.’ (C1)
- (336) *elep tše kahete*  
‘She can see it.’ (A1)
- (337) *omorfo dobos en an θelide as bam etši*  
‘It is a nice place, if you like, we can go there.’ (08\_04072019M\_3; 189–190)

#### 4.3.6.6 Expressions of external obligation

No data have been elicited yet for modals like ‘must’ and ‘have to’. According to Sitaridou (2014a: 123), ‘must’ clauses can be expressed by the Turkish loan *ile* (< *ille* ‘no matter what’, related to Arabic *illa* ‘except’), which functions as an invariant modal, + *na* + present tense (ex. 338 from Sitaridou 2014a: 123, ex. 12b). To establish this etymology, further evidence would be required, though, especially, since the Turkish model construction seems not to be very frequent in colloquial spoken Turkish. This construction did not figure in the present corpus but see *ila* + *s* in (444) below in a negative context. No data is available yet for past contexts.

- (338) *Ile na porpato/porpatis/porpati.*  
must PRT walk.1SG/walk.2SG /walk.3SG  
‘I/you/s/he/it must walk.’ (Sitaridou 2014a: 123, ex. 12b; glosses modified)

Furthermore, Sitaridou (2014a: 142, Table 6, exs. 101) reports the use of *exo* ‘have’ + deverbal noun to express ‘have to’ (339).

- (339) *Exo maireman.*  
have.1SG cooking  
‘I have cooking to do.’ (Sitaridou 2014a: 142, Table 6, exs. 101)

Tursun (2019: 371) notes that necessity can be expressed by *na* + present/imperfective tense (340, 341). This is in line with the fact that the subjunctive can also include a deontic meaning like ‘must’ and ‘have to’ (342, 343; cf. also Sitaridou 2014a: 142, Table 6).

- (340) *as pai ipa me na pai*  
OPT go.3SG say.AOR.1SG Q PRT go.3SG  
‘I have said that he shall go, he has to go.’  
(Tr. *Gitsin dedim mi, gitmeli!*, Tursun 2019: 371)

- (341) *n=epejna*  
‘I had to go.’ (Tursun 2019: 371)

- (342) *ta patsiðes utš ejollaevane n=eðuilevame*  
‘They did not send the girls, we should/had to work.’ (02\_02022015F\_1; 044)

- (343) *ti mana=muna jardim n=epiname*  
 ‘We should/had to help our mother.’ (02\_02022015F\_1; 045)

Finally, Turkish strategies of expressing obligation and necessity are borrowed, i.e., *mecbur olmak* ‘have to’ and *lazim* ‘need’. However, this seems to be primarily a strategy of Turkish-dominant speakers as the borrowing/insertion of the Turkish forms seems rather spontaneous without involving any integration (344). In (345), the Turkish construction of *lazim* + dative (Göksel & Kerslake 2005: 305) is copied into Romeyka as *lazim* + object pronoun.

- (344) *esi medžburen na pas eyo medžbur deul-um m=ila so banemo*  
 ‘You will necessarily have to go, I am not obliged, I mustn’t go.’ (03\_30062019F\_6; 06)

- (345) *lazim [...] emasine mila*  
 ‘We need apples.’ (C1)

#### 4.3.6.7 Non- and antiveridical irrealis mood

See Section 5.3.3.2 on conditional clauses.

### 4.3.7 (Non-)Finiteness

This section is dedicated to morphologically non-finite, namely infinitival and nominalized verb forms. Like is already the case for the previous Section on Negation and modality, this section is placed at the fringe between morphology and syntax; since many aspects of verbal morphology in Romeyka are analytical rather than morphological, all aspects related to verb morphology are discussed in Chapter 4 to keep them apart from the syntactical processes at clause level presented in Chapter 5. (For a discussion of the syntax of infinitives, see Section 5.3.2.2.1.) Non-finite verb forms are defined here (in line with Payne 1997: 306) as dependent verb forms that are less finite than full verbs, be it only with respect to minimally one agreement feature such as tense (see inflected infinitives below). While complementation in Romeyka is predominantly finite, non-finite complementation strategies are increasing under contact with Turkish; for (non-)finiteness at clause level, see Section 5.3.2.

#### 4.3.7.1 Infinitives

The retention of the AG infinitive in Romeyka is a notorious topic in scholarly debate (see also Sitaridou 2014b: 50, Table 4 for an overview of earlier publications with regard to the infinitive). While Mackridge (1987: 127, Fn. 17) still stated that Romeyka lacks non-finite forms of the verb, based on the assumption that finiteness equates to “indexing person”, he discovered evidence of the genuine infinitive soon after (P. Mackridge, p.c.). The existence of the uninflected infinitive in at least some varieties of ROF as spoken in Çaykara has been demonstrated later by Sitaridou (2014a/b), who reports a genuine infinitive in *-in(i)* (346). However, it seems inflected infinitival forms, that show subject-verb agreement (347) are more wide-spread among Romeyka dialects. In fact, the uninflected infinitive seems to be only used in some mountain villages to the south of Çaykara administrative district, such as the villages of Karaçam/Köknar (Óšena) and Uzungöl (Saráchos) (P. Mackridge, p.c.) and in Romeyka of Tonya (Sitaridou 2014b: 49, Table 3). In the latter variety, the infinitive exists in a reduced variant in *-in* (Sitaridou 2014b: 35, Fn. 10). The genuine infinitive is not attested in Dernekpazarı and Of administrative districts, which are closer to the sea, although ROF as spoken in (the city of) Dernekpazarı and Erenköy/Çoruk (Asan 1996) has the inflected

derivative of the infinitive. Inflected infinitives are furthermore attested in RSür (Sitaridou 2014b: 49, Table 3), also RSür as spoken in Beşk y ( zkan 2013: 148), and in the present corpus even from a speaker from RO  as spoken in  aykara/Kara am.

- (346) *Ut  eporesa t imi ini.*  
 not can.AOR.1SG sleep.INF  
 ‘I could not sleep.’ (Sitaridou 2014a: 122; RO , glosses modified)

- (347) *ut  eboresa t ime ina*  
 NEG can.AOR.1SG sleep.INF.1SG  
 ‘I could not sleep.’ (B1, T1)

According to Sitaridou (2014a: 122), the plain infinitive in Romeyka is defined as not bearing any agreement features. The plain infinitive consists of the aorist (active/passive) stem followed by the infinitival ending *-ini* (< AG present infinitive suffix *ειν*, Sitaridou 2014b: 35). Sitaridou (2014b: 48–50) differentiates between plain infinitives and personal infinitives which both occur in RO  and Romeyka of Tonya, whereby personal infinitives are morphologically identical with plain infinitives but have a subject distinct from the subject of the matrix verb. Furthermore, the inflected infinitive which occurs in certain villages of RO  and especially in RS r, is formed by the perfective (aorist) verb stem + *(i)n* + active past personal endings ( zkan 2013: 148; for a paradigm of personal endings see Sitaridou (2014b: 51, Table 5). According to Sitaridou (2014b: 48, Fn. 24), inflected infinitives have (like personal infinitives) a subject distinct from the subject of the matrix verb and occur as complement, subject or adjunct (348). Other examples from RS r (and RO ) seem to counter this and show that inflected infinitives are also used for clauses with identical subjects (349).

- (348) *na ixa mairepsina etroyamen*  
 PRT have.IPFV.1SG cook.AOR.INF.1SG eat.IPFV.1PL  
 ‘If I had cooked, we would eat.’ (RS r; Sitaridou 2014b: 49, ex. 50, glosses modified)

- (349) *do iladz n=i e alibsines har galo jendusune*  
 ‘If you had applied the ointment, now you would have been well.’ (RS r;  zkan 2013: 148)

Sitaridou (2014a) attests plain infinitives in RO  in the following contexts: (i) following negated past tense modals and volitionals, (ii) in ‘before’ clauses with *prin* ‘before’, (iii) in counterfactuals with *ixa* ‘I had’, and (iv) as hapax legomena after verbs of perception or negated past motion verbs. The syntactic constructions in which the infinitive occurs is known to vary between dialects (Mackridge 1999: 102–103). For example, in RS r as spoken in Beşk y, the infinitive is not used in *prin* ‘before’-clauses ( zkan 2013: 149) and negated past tense modals/volitionals take the inflected infinitive. For a description of the syntax of infinitives, see Section 5.3.2.2.1. In the present Romeyka corpus, infinitival forms occur rarely but in similar environments (see discussion of individual contexts (i)–(iv) below). However, the form of non-finite verbs varies; a repeatedly occurring infinitival form ends in *-ine* (see also Sitaridou 2014a: 122 who notes this form of the infinitive for “very few speakers”).

(i) Modal and volitional complements: According to Sitaridou (2014a), the plain infinitive occurs in negated past tense modal (350) and volitional complements (351). In the present corpus, a single (reduced) infinitival form is attested as negative past tense complement of *boro* ‘can’ (352). Otherwise, *na*-clauses prevail (353, 354). Furthermore, an infinitival form of the type NEG + particle + modal + conjunction + infinitive is attested in a negated future tense

potential construction (355). This construction is striking since it differs from the past forms described by Sitaridou (2014a: 35) which pattern as NEG + modal + infinitive. Negated future tense potentials are, according to Sitaridou (2014b: 132, ex. 56a) formed by *na*-clauses, e.g., *sapale tši poro na porpato* ‘tomorrow I cannot walk’.

- (350) *Utš eporesa tšimiθini.*  
not can.AOR.1SG sleep.INF  
‘I could not sleep.’ (Sitaridou 2014a: 126, ex. 25, glosses modified)
- (351) *Utš eθelesa mairepsini.*  
not want.AOR.1SG cook.INF  
‘I didn’t want to cook.’ (Sitaridou 2014b: 39, ex. 17; glosses modified)
- (352) *ama utš eboresane dosin adonusine*  
‘But they could not shoot them.’, lit. ‘give them’ (04\_01072019F\_12; 15)
- (353) *utš eborise n=eyvali=na*  
‘It could not take it off.’ (02\_29062019F\_2; 28)
- (354) *tevekela tib=utš eborenām n=epiname*  
‘We could not do anything alone.’ (02\_02022015F\_1; 089)
- (355) *utše ne borum tš=erxumesine*  
‘We won’t be able to come.’ (07\_04072019F\_5; 16)

However, in complements to negated past volitionals only a rare nominalization of an infinitival form occurs in a speaker from Karaçam (356, see also below). Note that Sitaridou (2014a: 45, ex. 40, i.e., ex. 468 below) analyses this form as infinitive with possessive suffix =*mu*, which makes sense in cases where the subject of both matrix and dependent clause is 1SG, but for the ending *-mo* in forms other than 1SG see below.<sup>209</sup> The phonological form in (356) suggests *-mu/-mo* to be a potential phonological variant of the nominalization suffix *-mon* rather than a possessive clitic. Otherwise, negated past tense volitionals are in the present corpus (i.e., in a speaker from ROF as spoken in Çaykara in the village of Soğanlı) realized as *na*-clauses (357, 358).

- (356) *sade jad emena faji utš eθelisa pseθinimo*  
‘I did not want to cook just for myself.’ (C1)
- (357) *utš eθelene buθen (n=) ebejene*  
‘She didn’t want to go somewhere.’ (01\_04022016F\_1; 105)
- (358) *utš eθelena n=andriza*  
‘I did not want to marry.’ (02\_02022015F\_1; 014)

(ii) Adjuncts of ‘before’-clauses with *prin* ‘before’: Although according to Sitaridou (2014a: 36) adjuncts of ‘before’-clauses trigger the infinitive (359), only finite verb forms are

<sup>209</sup> Note also rare forms like *temon t=erθan=im* ‘my coming’ in ex. 440 (from Sitaridou 2014b: 41, ex. 27) in Section 5.3.3.1.4, which pattern in the present data as *teso(n) t=erθ-ini-mo(n) θelo/eyabo* ‘I want you to come’ (C1, H3, see ex. 406 in Section 4.3.6.4) and which are very likely calqued on Turkish inflected infinitives of the type stem + (short infinitive) *-mA* + possessive suffix, e.g., *gel-me-m* ‘my coming’.

attested in these environments in the present corpus (360a/b). This is in line with what Özkan (2013: 149) states for RSür as spoken in Beşköy.

- (359) *Prin spudzisini so mandarin, tsi pao.*  
 before clean.INF at.the barn not go.1SG  
 ‘I am not leaving before I clean the barn.’ (Sitaridou 2014a: 129, ex. 9a)

- (360) a. *so pazari brin bane na pao si fatimes* (C1)  
 b. *so bazar prin bao di fadime zijared n=eftao* (B1)  
 ‘Before going to the market, I will visit Fatma.’

(iii) Counterfactuals such as wishes, exclamatives, and conditionals as a complement of *ixa*: According to Sitaridou (2014b: 136), in antiveridicals the infinitive occurs as a complement of finite *ixa* ‘I had’ (361). However, in a counterfactual construction of the type *na* + neg + *ixa* in the present corpus, a different infinitival form is attested (362). Notably, inflected infinitives in RSür seem to take invariant *ixe* (Sitaridou 2014b: 54–55), while in RO of the imperfect form of ‘have’ *ixa* shows subject-agreement (Neocleous 2020: 56; also Tursun 2019: 219, e.g., *ixa iníne* ‘if I were’ vs. *iše iníne* ‘if he was’).

- (361) *As išen portpatesini sa rašia.*  
 OPT have.IPFV.3SG walk.INF to.the mountains  
 ‘S/He should have walked in the mountains.’ (Sitaridou 2014a: 136, ex. 72a; glosses modified)

- (362) *[...] eyo na mi ixa škisen-da da ksila dže eftena don pešgo laya ni isayumunese laya n=exlimunesine*  
 ‘If I had not chopped the wood and lit the stove, how could we dry, how could we get warm?’ (04\_01072019F\_13; 53–55)

(iv) After negated motion verbs: Sitaridou (2014a: 42) also notes an infinitive after the negated motion verb *pa(y)o* ‘go’ (363; but *na*-clauses after *erθa* ‘come’, Sitaridou 2014a: 42). A similar construction occurs with an inflected infinitive in the Romeyka corpus (364a). Although according to Sitaridou (2014a), neither full nor inflected infinitives are due to contact with Turkish, the example in (364a) mirrors the Turkish participial form in (364b). An infinitive as complement to the perception verb *iða* (which has been mentioned as a hapax legomenon in Sitaridou (2014a: 40)) is not attested in the present corpus. Furthermore, Sitaridou (2014a: 130) mentions nominalized forms of the infinitive as complements to aspectuals such as *biturevo* ‘finish’ and mental perception verbs such as *anespalo* ‘forget’ (365, see also Section 5.3.2.2.1). The nominalized infinitive is used with a, possibly obligatory, complex possessive *=(e)muneθe* whose internal composition is not transparent, but which may be a calque from Turkish nominalizations of the type *oku-ma-sı-nı* ‘read-INF-3SG.POSS-ACC’ (Sitaridou 2014a: 130), in which case *-(e)mon* could correspond to the Greek nominalizing suffix *-mon* and the 3SG weak possessive *=eθe* to the Turkish 3SG possessive *-sı*. In another analysis, it could be a combination of two weak possessive pronouns *=(e)mun* ‘our’ and *=eθe* ‘its’ (Papadopoulos 1955: 59; see also Section 3.2.2.3 on possessives). Finally, although less likely, the phonological similarity between *=(e)muneθe* and the 1PL imperfective form of the copula *ime* ‘be’, *emunestine*.IPFV.1PL, is striking. This complex nominalized infinitive is not attested in the present corpus.

- (363) *tš=epies almeksin=ata*  
 ‘You didn’t go to milk them.’ (Sitaridou 2014a: 42, ex. 32)



- (364) a. *utš ebiya deresina*  
 b. *Gidip bakmadım.*  
 ‘I did not go and check.’ (04\_01072019F\_17; 50)
- (365) *To tšimiθin=emuneθe epiturepsa.*  
 the sleep.INF=our.its finish.AOR.1SG  
 ‘I finished sleeping (=I woke up).’ (Sitaridou 2014a: 130, ex. 48a; glosses modified)

#### 4.3.7.2 Nominalization

Apart from the infinitival forms described above, the present corpus features nominalizations in *-mo(n)* (< HelGr/MedGr *-imon*, Sitaridou 2014a: 46), for example in copula clauses (366; cf. Sitaridou 2014a: 42, ex. 29). Apart from nominalizations in *-mo* as complements to predicative adjectives, nominalizations with *-mo* or *-ma* together with the locative preposition *s-* are used after an aspectual such as *bašlaevo* ‘start’ (367) where Sitaridou (2014a: 45) attested either an infinitive or a nominalization strategy (368); see also Sections 4.3.3.1 and 5.3.2.2.2.

- (366) *ena aletero ylossa maθama.. maθenimo bola zori e*  
 ‘Learning another language is very difficult.’ (C1)
- (367) *ebašlaepse so borbatima*  
 ‘She started walking.’ (04\_01072019F\_13; 30)
- (368) *epašlaepsa pola so ðipsasini=mu / ðipsasimo*  
 started.AOR.1SG a.lot at.the get.thirsty.INF=POSS.1SG / thirst  
 ‘I started to get thirsty.’ (Sitaridou 2014: 45, ex. 40; glosses modified)

Alongside Romeyka nominalizations, integrated Turkish nominalized infinitives occur in copula clauses (369).

- (369) *embro dera megbul utš edune d=egudema*  
 ‘In these times, the teaching was not good.’ (02\_02022015F\_1; 052)

#### 4.3.7.3 Participles

Morphologically, there are arguably two kinds of participle in Romeyka: (i) passive participles (basically verbal adjectives) ending in *-menos* (see also Sections 4.3.3.2 and 4.1.1.2), (ii) gerundive participles formed by two different strategies. According to Sitaridou (2014b: 121), though, passive verb forms in *-menos* are the only participles in Romeyka and no gerunds exist like *-ondas* which is productive in SMG (Neocleous 2020: 59; see also Manolessou 2005 for Greek diachrony of participles and gerunds).

(i) These participles are deverbal adjectives formed by the passive perfective stem + *-menos* (Neocleous 2020: 55). For the inflectional properties of participles in *-menos*, see Section 3.1.2.2 on predicative adjectives. Participles frequently occur predicatively in combination with the copula *ime* ‘be’ (in both present and imperfective tense). According to Neocleous (2020: 55), “the auxiliary verb always proceeds the participle”. This is confirmed by the available data in the present Romeyka corpus where the copula also precedes the participle. The 3<sup>rd</sup> person singular copula can be omitted which is clearly a calque of the Turkish structure (370; see also Section 5.2.1.5). In (371), however, it is not clear whether the copula has cliticized with the participle. In general, participles in *-menos* do not only occur as predicative adjectives but also

as complements to other verbs/adverbially (see exs. 163, 164 in Section 4.3.3.2 above). Note that participles as manner adverbs are in Turkish-dominant speakers apparently in competition with the equivalent Turkish participle (372).

(370) *etšinos xaremenos*

‘He (is) happy.’ (02\_29062019F\_2; 24)

(371) *jane xaremenje plesame bola opsaræ*

‘We were happy, (that?) we caught many fish.’ (08\_04072019M\_1; 162)

(372) *sevinmiş epiya*

‘I went happily.’ (08\_04072019M\_1; 172)

(ii) The term gerundive participle is used here for uninflected verbforms used as manner adverbs. Romeyka has gerundive adverbials formed by the perfective stem + the suffix *-ta* (Neocleous 2020: 59). The etymology of the suffix *-ta* is unclear; probably there is a relation to the MedGr participle in *-onta*. They exist in both present and past context, although it is unclear, whether they show any tense agreement, cf. (373) and (374).

(373) *porpatefta na pas so džidženin*

walking.GER PRT go.2SG to.the.ACC grocery.ACC

‘You should get to the grocery on foot.’ (Neocleous 2020: 59, ex. 56; presentation/glosses modified)

(374) *porpatixda epejnam son barxari*

‘We went to the pasture by foot.’ (04\_01072019F\_2; 306)

Note, however, that these participles seem not readily available to all speakers; several periphrastic constructions occur even where Turkish has a converb in *-(y)ArAk* as equivalent (Göksel & Kerslake 2005: 89), e.g., Tr. *severek* ‘lovingly’ in (375), which is realized as two coordinated clauses (see also 376).

(375) *etšina ta ðulias eyapenam tš=ebinam=ada*

‘We did these works lovingly.’ (08\_04072019M\_2; 120–121)

(376) a. *xizlija ekonusefte istejan epije s=ospi* (B1)

b. *daxtira daxtira epije tše sade jelane eseve so spidi* (C1)

‘He ran quickly and headed home laughing.’

(Tr. *Hızlı koştu ve gülümseyerek eve döndü.*)

Furthermore, a non-finite gerundive verbform in *-ene* appears in subordination which is probably modelled on Turkish *-(y)IncA* converbs used for subordination, e.g., *nej-ene to zo* ‘when an animal gives birth’ (04\_01072019F\_1; 172), (377); see also Section 5.3.3.1.1 on temporal adjuncts.

(377) *ade ba derene šaka en ejelane*

‘She, seeing that it is a joke, laughs?.’ (02\_02022015F\_1; 146)

Tr. *O da bakınca şaka olduğunu anlayınca güler.*

## 4.3.7.4 Other non-finite verb forms

Furthermore, the corpus contains some examples of reduced verbal endings that seem to function as infinitival forms of otherwise finite verbs (378, 379a). Note that the form in (379a) is probably modelled on the Turkish converb (379b), whereas the verb form in (378) could be simply a reduced finite verb form.

(378) *spundžezo tše **kai** ka o kadar*  
 ‘I sweep and sit down, that’s all.’ (04\_01072019F\_2; 245)

(379) a. *bođar utš exo me ta sturatša **katsa ga** katsa ga*  
 b. *ayaklarım yok cubuklarla **otura kalka***  
 ‘I don’t have legs, I get up and sit down with the stick.’ (03\_30062019F\_7; 16)

## 5 Syntax

This chapter deals primarily with clausal syntax. For some aspect of syntax that are related to the verb, for example, negation and non-finiteness, see Section 4.3. on verbal morphology.

### 5.1 Structure of the noun phrase

Noun phrases in Romeyka are head-final; all determining modifiers, i.e., determiners, (attributive) adjectives and other modifiers such as genitives (Section 4.2.2), but except for weak possessive pronouns, occur to the left of the head noun, which occurs in the right-most position of the phrase (1).

- (1) *o emorfos o peðas*  
 the.M handsome.M the.M boy.M  
 ‘the handsome boy’ (constructed example)

In general, nominal modification features different kinds of attributive expressions that show different grammatical encoding and word order properties (Gil 2013c). Three different semantic types of attribution are (i) genitives such as adnominal possessives, (ii) adjectival attributes, (iii) relative clauses. In Romeyka, the three types of attribution pattern as follows: (i) Adnominal possessives precede the head and are marked by the genitive case on nouns or possessive pronouns (i.e., strictly GenN, Neocleous 2020: 287); (ii) Adjectival (and other) modifiers also precede the head noun. They are congruent with the head noun in a number of morphological properties, and they are sensitive to definiteness: definite NPs trigger determiner spreading; (iii) Relative clauses are (unlike in SMG and like in Turkish, see Dryer 2013g, Map 90A) predominantly pre-nominal and are marked by certain relativizers (but cf. Neocleous 2020: 238 for both orders). The typological tendency that languages with pre-nominal relative clauses are OV languages (Dryer 2013h) might shed an interesting light on the dominant word order of Romeyka as a language set on the intersection of a linguistic area with predominantly VO order in Europe and the circum-Mediterranean region, and predominantly OV further northwards into the Caucasus and Central Asia, and eastward across Iran into the Indian Sub-Continent (see Haig 2017). Relative clauses have been shown to be highly susceptible to contact influence. Considering that SMG has VO order and post-nominal relative clauses (NRel) and Turkish has OV order and pre-nominal relative clauses (RelN; Dryer 2013h, Map 96A); it seems that Romeyka is on its way to undergo metatypy modelled on Turkish, at least with regard to word order and NP structures, i.e., OV word order and RelN relative clauses. However, considering the typological correlation between word order and adposition, the Greek heritage of Romeyka is still visible, i.e., VO order and prepositions like SMG and unlike Turkish, which has OV and postpositions (Dryer 2013i, Map 95A). As Dryer (2013i) points out, the adpositional type of a language is more resistant than word order patterns (see also Haig et al., to appear, who confirm the relative inertia of adpositional order in settings of contact-induced change), so it is likely that cross-linguistically uncommon combinations of OV + prepositions derive from a recent change in word order patterns.

Furthermore, there has been a debate since Greenberg (1963) whether the order of adjectival modifier and head noun is likely to correspond to a dominant word order pattern, although, according to Dryer (2013j), an implicational relationship between the two cannot be established, since all (four) combinations are cross-linguistically wide-spread. Turkish and SMG are both categorized as AdvN while Turkish is OV and SMG is VO (Dryer 2013j, Map 97A). Still, within both languages, strictly head-final syntax of the NP is a property of Turkish, while in SMG, also right-branching NPs are possible. Karatsareas & Lekakou (2016: 198) argue for Cappadocian that these head-final properties of the NP are due to contact with Turkish.

Ultimately, it is necessary to explain predominantly left-branching word orders at NP level, i.e., prenominal genitives and relative clauses, together with frequently SOV word order (in subordinate clauses, predicatives and also often in declarative clauses) which are in line with Turkish word order patterning in contrast to SVO as unmarked word order (Neocleous 2020, 2022) and prepositionality, which are both inherited from Greek.

In Romeyka, the NP consists at least of a pronoun (not modifying pronouns such as possessives, though; 2) or a noun/proper noun (3, 4). A nominal phrase may contain either a full possessive pronoun or an enclitic pronoun but note the elicited example in (5) below, where both a free and a clitic pronoun occur within the NP – possibly influenced by Turkish. Enclitic possessive and object pronouns are generally the only elements that can follow the head.

- (2) *atin erθan*  
they came.3PL  
'They came.' (05\_03072019M\_4; 01)
- (3) *eyo ospi utš exo*  
'I don't have a house.' (B1)
- (4) *emine erθe*  
'Emine came.' (04\_01072019F\_5; 18)
- (5) *tema ta dört tane ta kosaras mu efeyane*  
'My four chickens escaped.' (C1)

The default order of modifiers in the NP is outlined in (6).<sup>210</sup>

- (6) [Rel], [Quant], [Pron], [(Art), Num], [Adv], [(Art), Adj], Def./indef. article, N/NN, [=Pron]

Definite nouns are in general preceded by the definite article, this principally also applies to proper nouns. Adjacency between the head noun and the definite article is required, thus no other element can appear in between. Indefinite referential singular nouns carry the indefinite article; other indefinite nouns appear bare (see Section 4.2.3). In definite NPs, determiner spreading (Sections 3.1.2.1, 3.2.1.1) occurs to attributive adjectives and partially numerals (Section 3.1.3): each attributive adjective is obligatorily preceded by the definite article in general congruent with the head noun in number, gender, and case (but for recent developments in the gender system leading to "mixed declensions", see Sections 3.1.2.1, 4.2.1.1). Possessive pronouns diachronically incorporate the definite genitive article, but demonstratives and quantifiers do not receive an article. Adjectives can be modified by adverbs of manner or degree that precede the adjective (7).

- (7) *bola ketine ðulias ixame*  
'We had very difficult work.' (50, 02\_9062019F\_2)

Cardinal numerals occur to the left of adjectives; in definite NPs, they also receive a congruent definite article (8). Modifying pronouns like possessive (see 5 above) or demonstrative pronouns (9) and quantifier (10) occur to the left of the noun (and adjective and/or numeral, if existing), whereby it seems that quantifier precede pronouns (11). In interrogative phrases, the

<sup>210</sup> Note that the overview of the maximal projections of the NP presented in (6) is a tentative one since no examples containing all these elements of a NP are available. Furthermore, discourse particles that also occur within the NP are not included in this representation; see at the end of this section.

interrogative pronoun precedes any noun (12) and can be modified by a preceding adverb (13). The NP can be modified by a relative clause; relative clauses in Romeyka are pre-dominantly pre-nominal, the relativizer occurs within the relative clause in immediate preverbal position (14; see also Section 5.3.4 on relativization).

- (8) *ts=aišes ta tria t=aðelfae bane so mektebi*  
 ‘Ayşe’s three siblings go to school.’ (C1)
- (9) *etšino t=emorfon i patsi*  
 that.NOM.N the.NOM.N=beautiful.NOM.N the. NOM.F girl. NOM.F  
 ‘this nice girl’ (NN2017: 36, ex. 22; presentation/glossing modified)
- (10) *olon da tehlikeliya da dobe*  
 ‘the most dangerous places’ (091, 08\_04072019M\_2)
- (11) *alo tibu tš=esuni*  
 ‘There was nothing else.’ (04\_01072019F\_17; 15)
- (12) *me do bio karðia*  
 ‘with which heart’ (04\_01072019F\_1; 068–069)
- (13) *alo din avres*  
 ‘Who else did you find?’ (04\_01072019F\_17; 28)
- (14) *aso istanbolin d=örðen i jinega temon i jenge en*  
 ‘The woman who comes from Istanbul is my aunt.’ (H1)

Although the topicalizer *pa(l)* (Section 3.2.4.5.1) usually follows the NP it emphasises, it can also appear within the NP when coordinating modifiers (15, 16) or emphasizing a modifier (17, 18). *Pa(l)* always follows the element it emphasizes. Note that the coordinating conjunction *tše* seems to fulfil similar functions within the NP, for example as a sort of additive particle in (19); see also Section 3.2.6.1.1.

- (15) *ta tria pal mikri da aðelfa*  
 ‘the three small siblings’ (A1)
- (16) *emorfon bal guvetin yarðelin emune*  
 ‘I was a handsome and strong child.’ (02\_21042018M\_2; 15)
- (17) *pola bal tšešida exorepsane*  
 ‘They danced a variety (of dances).’ (01\_06042017F\_4; 008–009)
- (18) *trana ba obsarae eplezame bola ba*  
 ‘We caught as well big as many fish.’ (08\_04072019M\_1; 159)
- (19) *dört dane beš dane tše yarðele*  
 ‘four or five children’ (04\_01072019F\_2; 093)

## 5.2 Independent clauses

### 5.2.1 Simple declarative clauses

A clause exists at least of a predicative element (20). Usually, there is one finite verb per clause (for verb serialisation which is frequent in Romeyka and shows coordination of several finite verbs, see Section 5.3.1).

- (20) *vreš*  
rain.3SG  
'It rains.' (cf. Drettas 1997: 284)

The basic structure of simple declarative clauses is shown in (21). Note that this is a tentative overview for which no complete example is available (but see 22). Furthermore, note that the order of adjuncts is variable according to information structure (see Section 5.3.3.1).

- (21) [Adjunct], [PP], [left-dislocated topic], [Subj NP], [PP], V, [Obj NP], [Adjunct]  
or [Adjunct], [PP], [left-dislocated topic], [Subj NP], [Obj NP], [PP], V

- (22) *ombron ta patsiðæ so kulin tš=epoliyan=æ*  
in.the.past the.ACC girls.ACC to school.ACC NEG=send.IPFV.3PL=OPN.3PL  
'In the past, they did not send the girls to school.' (Neocleous 2020: 120, ex. 14; presentation/glosses modified)

Simple declarative clauses have a flat intonation contour with a slight fall in the end (Section 2.4.2). Information structure/discourse effects and pragmatic reasons cause alternations of the basic word order, like topic dislocation (for the difficulty of establishing 'basic word order' in PG, see Drettas 1997: 277). It remains controversial whether we should assume a basic OV or VO structure for Romeyka as both variants are frequently attested; the figures from the WOVA data base (Schreiber 2021) for nominal direct objects indicate an almost exact 50% split between OV and VO. Rather than pointing out the dominant word order in declarative clauses, Romeyka could be rather considered to have more flexible word order determined by pragmatic reasons with two dominant basic orders (for the definition of word order adopted here, see the following section): VO and OV. This is in contrast to SMG, which has two dominant orders SVO and VSO (Dryer 2013k), whereby the position of the subject varies relative to the verb; unlike in Romeyka, where the position of the object relative to the verb is variable. MedGr shows VO and focus in situ (Neocleous 2020: 273–274) but word order in Late MedGr was mainly determined by pragmatic principles (Holton et al. 2019, vol. 4 syntax).<sup>211</sup> Following Dryer (1991: 475–476), it is assumed that the most fundamental word-order dichotomy is between OV and VO languages, while the position of the subject (S) is of less impact in determining the overall typological profile of the language. Nevertheless, subject placement will be discussed at various points below. Following standard practice in word order typology where the ordering with nominal constituents is generally taken as diagnostic, predominantly nominal (lexical) objects will be considered, as opposed to pronominal objects, which may exhibit different ordering patterns.

<sup>211</sup> For example, in MedGr "left-dislocated topics are typically used to establish or re-establish an entity as the current discourse theme and to initiate discussion of it" (Holton et al. 2019: 2022). It seems this also applies to Romeyka topics, resulting in frequent DO-V-IO orders (Section 5.2.1.4).

### 5.2.1.1 Word order in simple declarative clauses

Before presenting the basic word order properties of Romeyka, a brief definition of ‘word order’ is required. In the present study, word order is investigated predominantly by comparing frequencies in word order patterns of unmarked declarative clauses. Discourse-pragmatic aspects are considered only as a second step; since the present grammar sketch is based on a naturalistic spoken language corpus, it is often difficult to tease discourse and information structure apart from basic word order. While Neocleous (2020) provides a clear picture of word order in Romeyka based on grammaticality judgements and syntactic (generative) analysis, the present picture with naturalistic data is much messier and it is at times difficult to decide how the finding of Neocleous (2020) relate to the present data. For example, Neocleous (2020: 282) states that if Romeyka has been considerably influenced by Turkish word order, we would expect OV orders as default order also in declarative clauses. In terms of mere frequency, in the present Romeyka corpus, OV is the most frequent order in both declarative and subordinate clauses, but it is certainly difficult to estimate the influence of movement for pragmatic/discourse purposes. In the following section, it is therefore attempted at providing an overview both over findings from existing literature and the patterns in the present corpus (as well as the sub-corpus in Schreiber 2021).<sup>212</sup>

The unmarked basic word order of predicate and noun in declarative main clauses is arguably SVO (23; Neocleous 2020, 2022), although due to contact influence from Turkish, SOV default order (24) seems to be increasingly frequent in Turkish-dominant speakers. Furthermore, SOV order seems to apply if both subject and object are pronominal, e.g., (*eyo esena bola eyabo* ‘I love you a lot’). Note, however, that word order with pronouns is generally not considered the basic one in typology; often, the word order with nominal arguments is more conservative, while pronouns are more prone to shift and align with a contact language (Haig, Noorlander & Schiborr, under review; Haig, Molin & Noorlander 2022). It remains controversial whether the basic unmarked word order in Romeyka is predominantly OV though permitting VO order (25, 26) or whether Romeyka’s canonical order is “underlying” VO, with OV resulting mostly from focus fronting and from Turkish contact (27 vs. 28; Neocleous 2020, 2022, for a diachronic account, see especially Tables 1 and 6 in Neocleous 2022: 149, 157, respectively, going back to Sitaridou 2016; also Michelioudakis & Sitaridou 2016, 2012: 216); even the assumption of two dominant word orders VO and OV would be possible (see ex. 29). According to Michelioudakis & Sitaridou (2016: 9, Fn. 3), “there is some indication that the VP can be superficially head-final as well [...] without, however, excluding unmarked VO orders. Our working hypothesis here is that ROF is underlyingly VO, with OV arising in all/most contexts with objects presented as (new and contrastive) information, possibly also influenced by its contact with Turkish [...]”. Neocleous (2020: 234) concludes (based on his findings of VO in Romeyka declarative clauses and OV in subordinate clauses) in the words of Dryer (2013k) that Romeyka is a “third subtype of language lacking a dominant order [which] consists of languages in which different word orders occur but the choice is syntactically determined”. Apart from the importance of discourse for Romeyka word order, it may be hypothesized that word order might be sensitive to the nature of the bilingual repertoire of speakers, in a way that the more exposure to Turkish as speaker had and the more (s)he uses Turkish in daily life, the

<sup>212</sup> This section (as does the whole chapter) attempts to give a general overview of word order and information structure in Romeyka. Since it is not possible within the scope of this study to apply proper syntactical tests to determine possible word orders, the subsequent sections will give some estimation on the frequency of word orders found in the present corpus (see Dryer 2013o) for a “rule of thumb” in determining dominant word order: “if text counts reveal one order of a pair of elements to be more than twice as common as the other order, then that order is considered dominant, while if the frequency of the two orders is such that the more frequent order is less than twice as common as the other, the language is treated as lacking a dominant order for that pair of elements”). For the grammaticality of occurring word orders, reference is made to Neocleous (2020).



higher the likelihood for OV word order to dominate also in declarative clauses.<sup>213</sup> This hypothesis needs further research, though.

- (23) *ebiyane obis*  
‘They went back.’ (04\_01072019F\_12; 17)
- (24) *mis ađatšes erθafame*  
‘We grew up here.’ (08\_04072019M\_1; 207–208)
- (25) *etšine bal đotš emena milo*  
‘She gave me an apple.’ (C1)
- (26) *esi do kitabi don ađelfo=s eđotšes*  
‘You gave the book to your brother.’ (C1)
- (27) *ekatsame me ti mana=s*  
‘We sat together with your mother.’ (03\_07072019F\_1; 10)
- (28) *me di mana=m erθafame*  
‘We grew up with my mother.’ (02\_02022015F\_1; 004)
- (29) *ebiya gubisa=da džeha s ena pošet eban džeha da kadas evoksa voksa da kada*  
‘I went to pour them there over a plastic bag and called the cats, I called the cats.’  
(03\_30062019F\_11; 064–066)

According to Neocleous (2020, 2022), the pragmatically unmarked word order in Romeyka matrix clauses is SVO,<sup>214</sup> with possible orders of SOV and OSV arising for pragmatic reasons through left-dislocation of foci and topics (the same applies to interrogative clauses with direct questions; subordinate clauses and indirect questions have SOV as unmarked order, but also apply OSV). Romeyka allows all types of phrases (i.e., subject/object NPs, predicative complements, PPs, adverbials, quantifier, predicate, gerund) in focus and topic position; focus and topic position is immediately preceding the verb (Neocleous 2020: 186; Drettas 1997: 276; cf. also Holton et al. 2019: 2024 for “simple topics” in MedGr). If both focus and topic appear in the left periphery, the aboutness topic is the left-most element (Neocleous 2020: 124). Non-contrastive given information may appear post-verbally (Neocleous 2020: 121). It seems, however, that a subject may be post-posed for emphasis (30), with a rare OVS order arising. Thus, other than stated by Neocleous (2020), Romeyka seems to allow VS order, albeit very rarely (see below).

- (30) *butün jazun etšina ula ixa=ta eyo*  
‘The whole summer, I had them all.’ (04\_01072019F\_2; 139–140)

Nominal object and verb occur predominantly in unmarked VO order (31; cf. SMG which also has VO order, Dryer 2013l); although OV order is very frequent as well, arguably affected by information structural processes such as focus (32). However, it has to be noted that with the present naturalistic data, it is difficult to determine whether OV arises for pragmatic reasons

<sup>213</sup> This hypothesis is supported by the finding from the WOWA corpus (Schreiber 2021) that the variable “speaker” is decisive for different patterning of word order. Idiolectal differences in word order could be easily explained by different bilingual profiles of the speakers which arise due to individual upbringing, education, work life, migration biographies, marriage patterns, etc.

<sup>214</sup> In matrix and subordinate clauses with auxiliary, the order is Aux VO (Neocleous 2020: 273).

under left-dislocation or as the unmarked order; according to Neocleous (2020), the unmarked order of object and verb is VO. In the WOWA-corpus (Schreiber 2021), 66% of nominal direct objects are post-predicate. Potentially, phonological weight may be a decisive factor here, whereby nominal direct objects of smaller phonological weight (i.e., till 10 segments) are preferred in post-predicate position; and direct objects NPs with higher phonological weight tend to be pre-verbal. However, this is typologically very rare and could be a secondary effect of information structural processes. Furthermore, animacy might be an influential candidate factor, in the sense that [-HUM] nominal direct object NPs tend to be pre-verbal. Definiteness seems not to play a role; neither does the presence or absence of a definite article.<sup>215</sup>

In general, full pronominal object NPs seem not to differ from full nominal object NPs in terms of their position: The data of the WOWA corpus reveal that compared to nominal direct object NPs which are 66% post-predicate, full pronominal direct object NPs are 58% post-predicate; further research is required here to explore the patterning of word order in nominal and full pronominal objects. Weak pronominal object NPs are strictly post-verbal.

- (31) *eftjem da xordaræ*  
 ‘We make hay.’ (08\_04072019M\_1; 195)

- (32) *ta tšehere koftume*  
 ‘We cut the meadows.’ (08\_04072019M\_1; 197)

As for the order of auxiliary and verb, there is no difference in the word order of verb phrases involving auxiliaries (33), full (finite) verbs and infinite verbs (34) in declarative clauses. The order of auxiliary and main verb is Aux VO (Neocleous 2020: 234; see also Neocleous & Sitaridou 2022). Within the VP, no elements are placed between the verbs apart from the conjunction *tše*, the modal particle *na*, and, according to Neocleous (2020: 102), some adverbs like *kal* ‘again’ which can interpolate between auxiliary and main verb.

- (33) *i batsi=m eš ebine nife*  
 ‘My daughter was going to be a bride.’ (01\_06042017F\_4; 055)

- (34) *eyo na mi ixa škisenda da ksila [...]*  
 ‘If I had not chopped the wood, [...].’ (04\_01072019F\_13; 53)

As for the order of subject and predicate in intransitive clauses (Section 5.2.1.2), SV is the dominant order (35, 36; Neocleous 2020: 237). However, VS order is possible as well, although not occurring frequently. It seems VS order occurs more with nominal subjects (37), while its occurrence with pronominal subjects is very rare and marked (38).<sup>216</sup> Neocleous (2020: 105) argues that all subjects in pragmatically unmarked orders are actually left-dislocated.

- (35) *argos erθen*  
 ‘The bear came.’ (05\_03072019M\_3; 40)

- (36) *eyo udže na bayo*  
 ‘I will not go.’ (01\_04022016F\_1; 112)

<sup>215</sup> In the interpretation of these quantitative data, it needs to be considered that the size of the WOWA corpus is very small (i.e., 500 tokens); for more reliable data, a larger dataset would be required. This is in the sense of Dreyer (2013o): “Of course, unless one examines a large number and a broad variety of texts, one cannot be sure that differences in frequency may not occasionally reflect the idiosyncratic properties of a particular set of texts. It is likely that in some cases, further text counts would lead to classifying a language differently.”

<sup>216</sup> Cf. also exs. (62) in Chapter 4 cited from Michelioudakis & Sitaridou (2012: 219, ex. 10a).

- (37) *bonun da bođare=djes*  
 ‘Her legs have hurt.’ (03\_07072019F\_1; 05)
- (38) *aidika đulias eftame mis ha*  
 ‘Yes, we have these (kind of) works.’ (03\_30062019F\_11; 072)

The order of verb, object and obliques (including adverbial modifiers) is flexible in Romeyka and may vary according to the type of oblique. Oblique phrases are defined here as either noun phrases or adpositional, i.e., prepositional, phrases that are not arguments to a predicate but adverbial modifier functioning as adjuncts, i.e., they have scope over the verb of the clause (Dryer & Gensler 2013). Obliques that are predominantly expressed by PPs include the semantic roles of benefactive, instrument, comitative, purpose, location, direction. Time and place are often expressed by (simple) adverbial modifiers. Dryer & Gensler (2013) suggest that temporal obliques should be regarded separately, as they often pattern differently from a typological point of view and occur frequently at the beginning of a clause. An overview of the order of oblique and verb in Romeyka is given in Table 36 for different types of obliques.

**Table 36:** Order of oblique and verb in Romeyka

Type of oblique	Dominant word order <sup>217</sup>
benefactive ( <i>ja</i> ‘for’)	predominantly pre-verbal
instrument ( <i>me</i> ‘with’)	predominantly pre-verbal; but also frequently post-verbal
comitative ( <i>me</i> ‘with’)	predominantly pre-verbal; but also frequently post-verbal
other uses of <i>me</i> ‘with’	pre- and post-verbal
location	predominantly pre-verbal (Schreiber 2021)
direction (goal)	predominantly post-verbal (Schreiber 2021)
Simple temporal adverbials	predominantly pre-verbal; also between AUX and lexical verb (Neocleous 2020: 102)
Spatial adverbials	predominantly pre-verbal
Manner adverbials	pre- and post-verbal; also in between AUX and lexical verb (Neocleous 2020: 102)

While it is important to note that pragmatic aspects such as topicalization and focus cannot be accounted for in counting pattern frequencies (see Fn. 215), Table 36 reveals that pre-verbal obliques seem to form the majority in Romeyka across all types of obliques (cf. Turkish XOV order, Dryer & Gensler 2013). Especially, instruments (39–41) and comitatives (42, 43) with the preposition *me* ‘with’ appear pre-dominantly pre-verbal but also in considerable number post-verbally (in the Schreiber 2021 WOVA-corpus, 47% of post-predicate “other obliques”, i.e., comitative, instrument, benefactive, other). However, it can be assumed that the high numbers of pre-verbal PPs with *me* involve some kind of pragmatic focus, e.g., contrastive (43, also ex. 28 above); assumably, these operations are frequent with obliques for pragmatical reasons. It can therefore be assumed in line with Neocleous’ (2017, 2022) assumption of unmarked SVO word order in Romeyka, that post-verbal obliques with *me* are the unmarked option (see ex. 44, 45 for instruments and ex. 43 for comitative; according to Neocleous (2020: 149), focused adverbials appear in immediate pre-verbal position) which should reveal linear order without any discourse-effects as all information are new in discourse, roughly answering a question like ‘What happened?’, for the diagnostics see Neocleous 2020, 2022). For pragmatic

<sup>217</sup> Dominant word order is assessed based on quantitative occurrences in the present Romeyka corpus; this method does, however, not account for movement for purposes of topicalization or focus and does, thus, not help to identify the unmarked word order.

purposes, instruments appear frequently in pre-verbal position resulting in XO<sub>V</sub> (39), XVO (40), and OX<sub>V</sub> (41) orders (cf. also rare VXO in 46).

- (39) *ason barxari asi mazira me t=aleyo ta yomarae=m eferena*  
 ‘From the pasture, from the village square, I carried my loads with the horse.’  
 (08\_04072019M\_2; 136–137)
- (40) *me ti raša=muna fortunestin ta xortare*  
 ‘We carried the hay in our backing.’ (04\_01072019F\_2; 223)
- (41) *ta kabanae me to tšagudža siθaðo*  
 ‘I smoothed<sup>?</sup> the walls with the ?.’ (08\_04072019M\_3; 047)
- (42) *erθen me ta ðio patsið=atjes*  
 ‘She came with her two daughters.’ (07\_04072019F\_5; 04)
- (43) *mi tsi aulus udž epikan=a*  
 ‘We had not done it with the men.’ (01\_06042017F\_4; 068)
- (44) *do kindi esevenam so potami me d=opserangišo*  
 ‘In the mid-afternoon, we went down to the valley with the fishing rod.’  
 (08\_04072019M\_1; 147)
- (45) *eyrafeme me do yalemi*  
 ‘We wrote with the pencil.’ (04\_01072019F\_2; 042)
- (46) *tšen ebina med edži dolma*  
 ‘Usually<sup>?</sup>, I used to make dolma from them.’ (03\_30062019F\_11; 053–054)

The order of obliques expressing location and verbs is in Romeyka apparently subject to variation (cf. WOVA corpus, Schreiber 2021): While the majority of goals (both prepositional and adverbial) are post-predicate (47, 48; 78% post-predicate goals), locations (prepositional and adverbial) tend to be more often pre-verbal (49; 48% post-predicate locations). As for adjuncts, a tentative estimation of spatial adverbials in the present Romeyka corpus are also frequently pre-verbal (i.e., often clause-initial), although they also occur in post-verbal position (50–52). While usually adjuncts are inserted between the subject and predicate (Neocleous 2020: 109 following Alexiadou 1997), they can be also dislocated to the left of the subject (53). The same seems to apply to temporal adverbials; here, again, the post-verbal position might be the unmarked one (54). According to Michelioudakis & Sitaridou (2016), temporal adjuncts precede local adjuncts, which are in turn preceded by comitatives (55). Manner adverbs appear both pre- and post-verbally; although topicalization may play a role here as well in pre-verbal orders (56 vs. 57). Neocleous (2020: 102) states furthermore that some temporal and manner adverbs can interpolate between auxiliary and the verb.

- (47) *erxusuni so spidin*  
 ‘You came to the house.’ (03\_07072019F\_1; 22)
- (48) *feren ana so jan=is*  
 ‘Bring her to your side!’ (01\_04022016F\_1; 038)

- (49) *ena s ena sakul espringsam=a mono dže evalam=a s ena betra bugo*  
 ‘In a bag, we squeezed it a bit and put it under a stone.’ (03\_30062019F\_1; 21)
- (50) *ebiyane obis*  
 ‘They went back.’ (04\_01072019F\_12; 17)
- (51) *erθa ađaha s=ospe*  
 ‘I came here to the house.’ (03\_30062019F\_11; 044)
- (52) *bola estaθan edži*  
 ‘He stayed there a while.’ (03\_30062019F\_11; 051)
- (53) *eskiden beri ađatšes emis jašajevume*  
 ‘We have lived here since long.’ (08\_04072019M\_1; 009)
- (54) *pudžeka istin he na stin so inegölü joksa so xorio mi na stjetšidi panda*  
 ‘Will you come to Inegöl or will you be the whole time at the village?’  
 (03\_07072019F\_1; 45)
- (55) *pote mo tinan pu-merea epies*  
 when with whom where-side went.2SG  
 ‘When, where and with whom did you go?’ (Michelioudakis & Sitaridou 2016: 6, Table 1)
- (56) *emanaxos u=boro na tšališevo*  
 ‘I cannot work alone.’ (08\_04072019M\_3; 052)
- (57) *andan jertedje so inegölen imistine endaman eyonušeftam endaman ekatsame me ti mana=s*  
 ‘When you came to Inegöl together with us, we talked, we sat together with your mother.’ (03\_07072019F\_1; 09-10)

Finally, the word order of copula complements is pre-verbal (just 8% post-verbal copula complements, Schreiber 2021), while ‘become’-complements are predominantly post-verbal (69% post-verbal become complements, Schreiber 2021); see Sections 5.2.1.5. Neocleous (2020: 117) suggests that OV order in copula complements (especially of the 3SG copula form *en*) must be the result of a phonological change, since the copula is otherwise considered enclitic.

#### 5.2.1.2 Intransitive clauses and pro-drop

In Romeyka, there is no difference between the position of the subject of a transitive clause and an intransitive clause: The default word order of intransitive clauses – certainly for pronominal subjects – is SV (58–60; cf. previous Section 5.2.1.1), while VS order is possible with nominal subjects as well (61, 62). The unmarked word order SV is affected by discourse-pragmatic strategies such as contrastive focus (see exs. 59, 60); the focus position is in Romeyka in immediately preverbal position (Neocleous 2020: 141).

- (58) *eyo biya dahabuko ados epije eyo biya đabuko*  
 ‘I went down, he went, I went down.’ (03\_30062019F\_6; 40)

- (59) *i alin tšimundane*  
 ‘The others sleep.’ (04\_01072019F\_12; 22)
- (60) *ta peðia airija portanenane ta patsiðes airija porpatenane*  
 ‘The boys walked on their own, the girls walked separately.’ (02\_02022015F\_1; 076–077)
- (61) *tšimundane tu spidi i arθob*  
 ‘The people of the house sleep.’ (05\_03072019M\_3; 35–36)
- (62) *ondan sonra ište erxundan da yarðelæ*  
 ‘Then the children came.’ (04\_01072019F\_1; 046–047)

Finally, Romeyka is a pro-drop language, i.e., the person value is specified by subject affixes on the verb and free subject pronouns are not obligatory, if the contextual reference is clear; in fact, subjects are frequently omitted in Romeyka (63, 64). Free subject personal pronouns occur in topic shifts (65, 66), for reasons of contrast (67, 68), and emphasis (69). Also, in subordinate clauses, co-referential subjects do not have an overt subject pronoun, e.g., (*eyo*) *θelo na (eyo) pao* ‘I want to go’ (Neocleous 2020: 98).

In Romeyka, the influence of pragmatics and discourse on word order is relatively high: note that in discourse, frequently subject or object NPs can be omitted, if the reference is clear from the context (70).

- (63) *tu spid i arθob eskoθane eksimniyanerune nunizune*  
 ‘The people of the house got up, are awakened, they look, think.’ (05\_03072019M\_3; 43–44)
- (64) *eši ðiyo tin aiše*  
 ‘I am giving [you] to Ayşe.’ [on the telephone] (01\_04022016F\_1; 147)
- (65) *ifedi eyo ebiya ište muskaræ*  
 ‘I bought calves this year.’ (08\_04072019M\_1; 225–226)
- (66) *kal eyo kruy=atona*  
 ‘I hit him again.’ (02\_2906019F\_1; 16)
- (67) *ade bal ekovalisen dž=eyo estivaksa*  
 ‘She moved them and I piled them up.’ (01\_28062019F\_2; 15)
- (68) *eyo jelo hade duru dže esta tše deri=me*  
 ‘I laugh, she keeps looking at me.’ (01\_28062019F\_2; 27–28)
- (69) *ula tsi mamikandus uluns eyo kovalina*  
 ‘All, everything of the elderly women, I carried.’ (04\_01072019F\_2; 146)
- (70) *eyo har utš exo*  
 ‘I don’t have [any] at the moment.’ [omitted: *za* ‘animals’] (06\_03072019M\_2; 45)

## 5.2.1.3 Transitive clauses

Transitive clauses are understood here in their basic form as consisting of a predicate with two argument NPs, whereby one is the subject and the other the (direct) object in accusative case (for passive voice see Section 5.2.2.2). As outlined in Section 5.2.1.1 above, the unmarked word order of transitive clauses has been suggested to vary according to the speaker's multilingual profile between OV (71) and VO (72, 73). It seems, though, that for basic declarative clauses without obliques, VO order is preferred (also Neocleous 2020), while clauses containing additional constituents seem to trigger OV order (74). According to Neocleous (2020: 140), Romeyka applies OV to monotransitive verbs under focus, whereby the verb remains in-situ and the object occurs in immediate pre-verbal position.

- (71) *ksila ekovalename*  
'We carried wood.' (09\_04072019\_7; 35)
- (72) *valis eliyo nero*  
'You put a bit water.' (03\_30062019F\_11; 101)
- (73) *n=ekovalina xoma n=ekovalina njezma*  
'I carried soil, I carried cow pats.' (04\_01072019F\_2; 071–072)
- (74) *to xargo θenume eban*  
'We put the cooking pot' above [the fire].' (03\_30062019F\_11; 082)

A feature which is striking in PG (and other Greek varieties such as AMG, SMG, MedGr), is clitic doubling of nominal object topics in left-dislocated position by means of a referential resumptive pronoun (75a; see Holton et al. 2019: 2023 for MedGr). In PG, clitic doubling exists for all topics to distinguish a topic from pre-verbal focus, since focalization never triggers clitic doubling. Thus, leaving out clitic doubling in PG automatically puts the dislocated NP into focus. In Romeyka, however, clitic doubling of object topics is not obligatory (75b). Ex. (76) exemplifies some form of resumptive long-distance anaphora by means of a weak object pronoun, which is not exactly clitic doubling, though. In MedGr, clitic doubling might be sensitive to indefiniteness or unspecificity, but this seems not to apply to the Romeyka examples. Neocleous (2020: 120) assumes clitic doubling via clitic left-dislocation of aboutness topics (77), although what he analyses as 3<sup>rd</sup> person singular object clitic =*æ* is phonologically very similar to the 3PL verbal ending -(y)*ane* (the same applies to ex. 78 by Michelioudakis & Sitaridou 2016: 11, ex. 23a). Note, however, that in (79), also the 3SG object clitic is used as anaphoric reference to a 3PL topic. In (80), however, the 3PL weak object pronoun is used as a resumptive pronoun of a pronominal topic, which is also rare. Finally, note the occurrence of the enclitic 3SG resumptive pronoun in the main clause of headless relative clauses (81a/b).

- (75) a. *avudo to saxan epero=to* (constructed example)  
b. *avudo do saxan eboro tše epero* (A1)  
'Can I take this plate?'
- (76) *fasulijas bote en erxume bero=na*  
'When there are beans, I come and take them.' (03\_30062019F\_11; 041)
- (77) *ombron [ta patsiðæ] s okulin tš=epoliyan=æ*  
in.the.past the.ACC girls.ACC to school.ACC NEG=send.IPFV.3PL=OPN.3PL  
'In the past, they did not send the girls to school.' (Neocleous 2020: 120, ex. 14; presentation/glosses modified)

- (78) *[Pion fai]<sub>i</sub> pios epitšen=æ<sub>i</sub>?*  
 which.ACC food.ACC who.NOM make.AOR.3SG=OPN.3SG  
 ‘Who made what food?’ (Michelioudakis & Sitaridou 2016: 11, ex. 23a; glosses modified)
- (79) *eyo lata atšin [da tšitšege xasun=a]*  
 ‘I told her to remove the flowers.’ (01\_28062019F\_3; 03)
- (80) *butiün jazun etšina ula ixa=ta eyo*  
 ‘The whole summer I had them all.’ [i.e., the animals] (04\_01072019F\_2; 139–140)
- (81) a. *[aso Katoxor **op** erθe] iða=na* (A1)  
 b. *[aso katoxor erθe] o mexmetis iða=na* (A1)  
 ‘I saw the one/Mehmet who came from Çaykara (Katohor).’

Like subjects, (direct) objects can be left out, if their reference is clear from discourse (82; for verb serialisation see Section 5.3.1). This does not apply solely to verb serialization, though, nor is it limited to pronominal NPs (see ex. 70 in Section 5.2.1.2). However, it is frequent with pronominal direct object NPs, for example, in imperatives (83, 84). There is likely a parallel to spoken Turkish where those direct objects that are recoverable from the context can also be left out (Göksel & Kerslake 2005: 127). Note that sometimes also an indirect object of a ditransitive clause can be omitted (85).

Apart from omission of objects, note that objects can for pragmatic reasons also be resumed at the end of a clause (86).

- (82) *ebedžede da idas plinun dže spingune dže valun edži*  
 ‘Then they wash the things, squeeze them and put them there.’ (03\_30062019F\_11; 087)
- (83) *đos emen*  
 ‘Give [her] to me.’ (01\_04022016F\_1; 148)
- (84) *feride=me*  
 ‘Bring [it] to me.’ (04\_01072019F\_2; 201)
- (85) *eši điyo tin aiše*  
 ‘I am giving [you] to Ayşe.’ (01\_04022016F\_1; 147)
- (86) *emena an erotai ulu t.. ulan anlateva dona da hikajeres*  
 ‘If you ask me, I will tell you the whole story.’ (C1)

With pronominal objects, weak object pronouns appear behind the verb (87, 88), while full object pronouns appear often pre-verbally (89, 90). However, not few full object pronouns occur post-verbally, which is a dubious topic that requires further research. Finally, OSV order seems possible as well (91).

- (87) *so džob n=evalena=da*  
 ‘I will put them into the waste.’ (03\_30062019F\_11; 068)
- (88) *o kadar eyabo=se*  
 ‘I love you so much.’ (01\_14012015F\_1; 4)



- (89) *etšinena eynorizes me*  
 ‘Do you know her?’ (01\_04022016F\_1; 020)
- (90) *ta yarðela=m emena hidž hesap udž ebidže*  
 ‘My children did not give me a credit at all.’ (01\_06042017F\_4; 134–135)
- (91) *esena bal ade bola eyaba*  
 ‘She loves you (also) a lot.’ (01\_14012015F\_1; 2)

As for the word order in light verb constructions (LVCs): since LVCs are largely copied from Turkish/Ottoman, many of them appear with verb-final order (see ex. 90), although the construction for ‘to give a phone call’ is VO (92).

- (92) *i xismedžeji n=aftai=me telefon*  
 ‘The servant was about to call me.’ (03\_30062019F\_7; 11)

#### 5.2.1.4 Ditransitive clauses

Ditransitives are clauses where the verb governs two arguments next to the subject: one usually an addressee or recipient and the other in the semantic role of theme (Haspelmath 2013c). In Romeyka, both the direct and the indirect object are coded by accusative case (“double-object construction” in terms of Haspelmath 2013c), as are monotransitive objects. However, since Greek (and Turkish) traditionally have indirect object constructions where direct and indirect object are coded differently by case, Romeyka seems to have arrived at the double-object construction due to a reduction of case inflection in objects, thus departing originally from an indirect object construction in Med Gr: All datives are realised as morphological accusative NPs, whereby this case marking does often not surface (Michelioudakis & Sitaridou 2012: 231). There seems to be no differential object marking in Romeyka, unlike in Turkish, i.e., monotransitive objects and ditransitive themes are not marked differently when definite (but see Section 4.2.3 for differential subject marking in PG), although there is further research needed on the use of definite articles with definite objects. The predicate does only show subject agreement and is not congruent to any object. To differentiate between direct and indirect object, word order can be used as a cue, or the grammatical role is inferred from context.

In the Romeyka corpus, the most frequent order of nominal objects (involving a recipient) is (S-)DO-V-IO (93–95), although the reverse order (S-)IO-V-DO does exist as well (96, 97). It is not in all cases easy to determine the role of topicalization (but note for example contrastive topic in ex. 98), though, but it seems that in principle, the order (S-)V-IO-DO is the most unmarked order (ex. 99; cf. SVO as default word order in Neocleous 2020; see also Drettas 1997: 274, 279 on PG). However, this order does hardly appear in the present corpus, probably since (a) in free narratives, often discourse-relevant processes such as topicalization and focus override the unmarked word order; and (b) in questionnaire data elicited by a translation task from Turkish, Turkish word order (which is DO-IO-V, Göksel & Kerslake 2005: 342) influences the Romeyka data.<sup>218</sup> In the WOVA sub-corpus (Schreiber 2021), 61% of all (i.e., nominal and pronominal) direct objects appear in post-verbal position, and even 66% of nominal direct objects (both definite and indefinite, so definiteness seems not to play a role), while 58% of pronominal direct objects are post-verbal (enclitic pronouns are not considered); see also Section 5.2.1.1. It is not clear yet whether there might be a difference in word order with regard to the semantic role of the object, e.g., between recipients, benefactives and addressees; the present data seem not to support this; according to

<sup>218</sup> For a critical reflexion on this methodology, see Section 1.2.3.

Michelioudakis & Sitaridou (2010: 139, 143), both recipients and benefactives appear in both pre- and postverbal position.<sup>219</sup> According to Michelioudakis & Sitaridou (2012: 232), both object orders (IO-DO/DO-IO) occur (98), although in their data – like with the present data – DO-V-IO order is most common. In any case, ditransitive clauses are likely to be VO, rather than OV, although DO-IO-V order exists (99), but they are likely calqued on the Turkish structure (as in ex. 99 in response to a translation task from Turkish). However, verb-final orders also occur for pragmatic reasons: According to Neocleous (2020: 193–195), in the linearization of objects in ditransitive verb constructions involving topic and focus a topicalized NP precedes a focused NP and both precede the verb; this order is irrespective of the grammatical function of both NPs (100, 101).

- (93) *ta mila đokan ti mana=tuna*  
‘They gave the apples to their mother.’ (C1)
- (94) *o hasanis.. to ketši=nat.. eđotše to.. to mustafa*  
‘Hasan sold his goat to Mustafa.’ (C1)
- (95) *inetši to psomi eđokan ena auro*  
‘The women gave the bread to the man.’ (C1)
- (96) *din batsi eđotše ena ido.. sturatši*  
‘He gave a stick to the girl.’ (C1)
- (97) *eleyane ištera ta peđia kopela ta patsiđes elayane yospiđes*  
‘Later, they said to the boys girl and to the girls they said prostitutes.’  
(02\_02022015F\_1; 073–074)
- (98) *Aiše epitše to Mehmet pide / pide to Mehmet.*  
Aise.made.3SG the.ACC Mehmet.ACC pie.ACC / pie.ACC the.ACC Mehmet.ACC  
‘Aise baked Mehmet a pie.’ (RSür; Michelioudakis & Sitaridou 2012b: 235; glosses modified)
- (99) *esi do kitabi don ađelfo=s eđotšes*  
‘You gave the book to your brother.’ (C1)
- (100) *[Ti sandalian]<sub>C-TOP</sub> [ton TŠIri=m]<sub>C-FOC</sub> eđotšen.*  
the chair.ACC the father.ACC=POSS.1SG give.AOR.3SG  
‘[He] gave the chair to my dad.’ (Neocleous 2020: 194, ex. 152b, presentation/glosses modified)
- (101) *[Ti mana=m]<sub>C-TOP</sub> [to saAtin]<sub>C-FOC</sub> efitšen.*  
the mother.ACC=POSS.1SG watch.ACC bequeath.AOR.3SG  
‘[He] gave the watch to my mother.’ (Neocleous 2020: 195, ex. 153b, presentation/glosses modified)

In pronominal NPs, however, the most frequent word order differs from that of nominal object NPs: With full object pronouns, (S-)V-IO-DO order seems to be more frequent (102, 103, see also S-IO-DO-V order in ex. 104), which corresponds to the assumed unmarked word order; while (S-)DO-V-IO order, which is most prominent with nominal objects, exists as well (105),

<sup>219</sup> Note that recipients (and goals) of ditransitive constructions cannot alternate with prepositional phrases, while benefactives may alternate with prepositional phrases headed by *đæ* ‘for’ (ROf) or *ja* ‘for’ (RSür; Michelioudakis & Sitaridou 2012: 231–236; also Michelioudakis & Sitaridou 2010: 139).

although possibly resulting from a marked information structure. While ditransitive constructions with pronominal object NPs are often post-verbal (103, 104), verb-final orders also exist, probably arising under focus, such as with the interrogative in (106; see also Section 5.2.3.3 on word order in interrogative clauses).

- (102) *etšine bal dotš emena milo*  
 ‘He gave me an apple.’ (C1)
- (103) *ta yarđeli=nat osad ebejin eđokan adona ena suri hedijeđes*  
 the.PL children=POSS.3SG when went.3SG gave.3SG OPN.3SG a number presents  
 ‘His children gave him some presents when he left.’ (C1)
- (104) *O Mehmejis adona etšino fanerose.*  
 the Mehmet.NOM him.ACC this.ACC show.AOR.3SG  
 ‘Mehmet showed this to him.’ (RSür; Michelioudakis & Sitaridou 2010: 141, ex. 10f; glosses modified)
- (105) *esi do kitabe eđotšes atšinena*  
 ‘You gave her the book.’ (C1)
- (106) *emenan dona na ðiyune*  
 ‘What do they give me?’ (08\_04072019M\_1; 272)

In pronominal NPs involving clitic pronouns, the default order of objects is – like with full pronominal object NPs, and acknowledging the fact that the position of the enclitic object pronoun is fixed – (S-)V-IO-DO (107, 108), although DO-V-IO order exists (again like with full pronominal object NPs) under focus condition (109, 110). Like with full pronominal object NPs, the most frequent word order shows post-verbal objects, although pre-verbal objects (both direct and indirect) are possible under focus, e.g., IO-V-DO order in (111).

- (107) *esi na ðis=me kabane*  
 ‘You will give me coat.’ (08\_04072019M\_3; 059)
- (108) *sa xuljera apopis na ðivune=me psomi*  
 ‘The spoons from behind, they will give me bread.’ [saying meaning ‘I work in order to get food.’] (07\_04072019F\_6; 37)
- (109) *to biberoni enđiyum=a*  
 ‘We gave it the feeding bottle.’ (02\_2906019F\_1; 09)
- (110) *i aiše to psomi eđotše=me*  
 ‘Ayşe gave me the bread.’ (C1)
- (111) *ipen havu to peđin habađahandžeka eparen ja do hūkūmetin na ðiyum=a*  
 ‘She said, “Take this child from here or we will give him into a children's home!”’  
 (02\_21042018M\_2; 19–20)

Clitic clusters of enclitic pronominal object NPs are possible as well, whereby the only possible order of weak object pronouns is V-IO-DO (exs. 112–114). This order corresponds to the assumed unmarked word order in ditransitive clauses and to that of full pronominal object NPs (see above). Clitic clusters are in the present corpus only attested with first- and second-person

indirect objects (cf. Drettas 1997: 279). Note that Michelioudakis & Sitaridou (2012: 237) question the clitic nature of first- and second-person object pronouns and thus conclude that there are no clitic clusters in Romeyka. Furthermore, according to Michelioudakis & Sitaridou (2012: 218), the third person singular object clitic *-æ* cannot be combined with another clitic (115), which, however, seems to be possible in ex. (113). According to the exs. (112, 113), it seems to be possible in ROF to combine first- and second-person indirect objects with third-person direct objects besides full pronominal object NPs (116, 117; cf. Michelioudakis & Sitaridou 2012: 237–238).

- (112) *eđotše=m=ada*  
 ‘She gave them to me.’ (04\_01072019F\_2; 086)
- (113) *eđiksa=s=a*  
 ‘I showed it to you.’ (08\_04072019M\_3; 159)
- (114) *Eđiksane=m=ese.* / *\*eđiksane=s=eme*  
 showed.3PL=OPN.1SG=OPN.2SG / showed.3PL=OPN.2SG=OPN.1SG  
 ‘They showed me to you.’ (RSür; Michelioudakis & Sitaridou 2012: 238; glosses modified)
- (115) *\*O Mehmetis eđotše=m=æ*  
 the Mehmet gave.3SG=OPN.1SG=OPN.3SG  
 ‘Mehmet gave it to me.’ (RSür; Michelioudakis & Sitaridou 2012: 218; glosses modified)
- (116) *Eđiksan(e) =æ/aton(a) emenan.*  
 showed.3PL =OPN.3SG / OPN.3SG OPN.1SG  
 ‘They showed him to me/ \*They showed me to him.’ (RSür, ROF; Michelioudakis & Sitaridou 2012: 238; glosses modified)
- (117) *Eđiksane=me aton(a).*  
 showed.3PL=OPN.1SG OPN.3SG  
 ‘They showed him to me.’ (RSür; Michelioudakis & Sitaridou 2010: 141, ex. 10c)

### 5.2.1.5 Predicatives

Expressions of state, identity, and location can be formed with non-verbal predicates coupled with the verb *ime* ‘be’. While in predicate nominals and adjectives the copula *ime* is in clause-final position sometimes omitted in third person singular present tense (especially in elicited questionnaire data facilitating copying from Turkish), the verb *ime* is obligatory in third person existentials. Although a contact explanation with structural copying from Turkish, which marks non-verbal predicates with a copula, except for third person singular present tense predicatives, seems natural, an internal explanation seems vital as well since the third person singular form of the copula could be also dropped in AG under certain circumstances (M. Janse, p.c.). Furthermore, the status of the verb *ime* in some modern Greek varieties may offer another explanation. In other AMG varieties (Cappadocian, Phrasiot, but not SMG), two forms of the verb *ime* exist; (i) the full verb *ime* ‘be’ used in existentials, which is never dropped; (ii) the copula *ime* ‘be’ which behaves as a clitic (=me) since MedGr times, but which is never left out in third person singular present tense. Finally, as opposed to variable word order of the copula *ime* in predicatives, complements of ‘become’-clauses in Romeyka are predominantly post-verbal (also see Schreiber 2021).

## 5.2.1.5.1 Predicate nominals

Predicate nominal sentences are clauses where a subject occurs with a nominal predicate (Stassen 2013a). In Romeyka, predicate nominals are linked to the subject by a copula in both present and non-present tense. Stative predicate nominal sentences, including equatives, with the copula *ime* ‘be’ are in ROF nearly always verb-final (118–122, but cf. ex. 123, 124; see also Neocleous 2020: 53; but cf. Drettas 1997: 271, who states SVX as the default order). In questionnaire data translated from Turkish the clause-final third person singular present tense copula is frequently omitted, which is likely due to direct copying from the Turkish prompt (125a/b, but cf. 125c; see also Neocleous 2020: 54). Predicate nominals in Turkish are marked by a nonverbal copula, that is a pronoun in form of a suffix that serves as the linker between subject and predicate nominal (126), and which is zero-marked in third person (see ‘pro-copulas’ in Stassen’s (2013a) terminology). From a typological point of view, third person copulas behave often differently compared to other person forms (Stassen 2013a).

- (118) *eyo ixtiari inega ime*  
 ‘I am an old woman.’ (01\_28062019F\_3; 13)
- (119) *eyo mualimis ime*  
 ‘I am a teacher.’ (C1)
- (120) *emine mualimis edune*  
 ‘Emine was a teacher.’ (C1)
- (121) *bola emorfo batsi en*  
 ‘She is a very beautiful girl.’ (01\_28062019F\_2; 34)
- (122) *ade ba derene šaka en ejelane*  
 ‘Realizing that it is a joke, she laughs.’ (02\_02022015F\_1; 146)
- (123) *i dri en tu spid i arθob*  
 ‘The three are the people of the house.’ (05\_03072019M\_4; 02)
- (124) *šeher oldi ne har en šeheri*  
 ‘It became a city, yes, now it is a city.’ (04072019F\_10; 19)
- (125) a. *bu benim*  
       b. *ado demon* (B1)  
       c. *hado demo e* (C1)  
       ‘This is mine.’
- (126) a. *(Ben) öğretmen-im.*  
       ‘I am a teacher.’  
       b. *(O) öğretmen.*  
       ‘He is a teacher.’

Dynamic predicate nominal sentences involving the change-of-state verb ‘become’ are in turn not verb-final (127, 128). This also applies to ex. (129), where ‘become’ is not used as a change-of-state verb but rather in the meaning of ‘come into existence’.

- (127) *do ylidži ejendune diri*  
 ‘The milk became cheese.’ (03\_30062019F\_1; 20)

- (128) *ejendune omorfo diri*  
‘It became a good cheese.’ (03\_30062019F\_2; 07)

- (129) *ondan sora ejendune ena yarðel*  
‘Thereafter a child came.’ (01\_06042017F\_4; 123)

#### 5.2.1.5.2 Predicate adjectives

Like predicate nominals, stative predicate adjectives are formed with the copula *ime* in clause-final position (130–135). Note, however, that for pragmatic reasons, the copula may also occur in clause-initial position (136). Again, in data from the translation task, the third person present tense copula is frequently omitted (134, 135a, but cf. 135b).

- (130) *eyo bola kaðakesa ime*  
‘I am very ?.’ (translation unclear; 01\_28062019F\_2; 16)

- (131) *bola magri ise*  
‘You are very tall.’ (C1)

- (132) *ta mandria=muna boš en*  
‘Our stables are empty.’ (08\_04072019M\_1; 204)

- (133) *da kadas me do limo ini*  
‘The cats are hungry.’ (03\_30062019F\_11; 068–069)

- (134) *havudo spidi bola trano*  
‘This house is big.’ (C1)

- (135) a. *tsi fatimes to zo aspro*  
‘Fatma’s cow is white.’ (C1)  
b. *teso o šgilo mavro en*  
‘Your dog is black.’ (B1)

- (136) *en da malia=ndes enanero*  
‘Her hair is soaking wet.’ (04\_01072019F\_13; 30)

In dynamic predicate adjectives with *ejendune* ‘become’, most complements of ‘become’ are post-predicate (i.e., 70% in the corpus of Schreiber 2021), which is the unmarked order (137, 138), while OV arises from topicalization (139, 140), e.g., as indicated by the topic marker *ba(l)* in (139) or focus (141).

- (137) *ejendune xastas*  
‘He became sick.’ (01\_06042017F\_4; 124)

- (138) *omon d=eroise s=ormi ejendune na-tsurula*  
‘When she fell into the water, she became soaking wet.’ (04\_01072019F\_13; 45)

- (139) *heralda kalabaluk ba ejendune*  
‘Apparently it became crowded.’ (03\_30062019F\_6; 15)

- (140) *argoðes ena edun ðio dria tesera bende ejendane*  
‘The bears were one, they became two, three, four, five.’ (05\_03072019M\_4; 38)

- (141) *i patsi adjes xasta.. xastas ijendun*  
 ‘Her daughter became sick.’ (07\_04072019F\_5; 14)

#### 5.2.1.5.3 Predicative locatives and other predicatives

In predicative locatives (142–144) and other predicative adpositional phrases involving a preposition (145) like in benefactives or accompaniment, complements of the stative copula *ime* ‘be’ are pre-verbal like in other predicatives above. (Romeyka is a share-language in the terminology of Stassen 1997, which means that both nominal and locative predicatives are realized by the same encoding strategy).

- (142) *s=oros abes-mereja ime*  
 ‘I am inside the forest.’ (01\_28062019F\_3; 42–43)
- (143) *öretmenis so kul en*  
 ‘The teacher is at school.’ (A1)
- (144) *ta yarðela so džebi ine*  
 ‘The children are in the garden.’ (B1)
- (145) *jetmiš iki jašinda ime*  
 ‘I am seventy-two years old.’ (01\_28062019F\_3; 13–14)

Complements of *ejendune* ‘become’, are post-verbal (146).

- (146) *ejendune ab emas*  
 ‘She became [one] of us.’ (01\_04022016F\_1; 076)

Note that like with predicate nominals and adjectives, the third person present tense copula is occasionally left out (147).

- (147) *t=opsaræ so spid embro*  
 ‘The fish [are] in front of the house.’ (05\_03072019M\_4; 18)

#### 5.2.1.5.4 Existentials

Existentials are formed with the copula *ime* ‘be’ showing person agreement with the predicative complement, often accompanied by a spatial adverb (148) or locational expression (149), which is not obligatory, though (150–152).

- (148) *omorfo ospid en etši*  
 ‘There is a nice house.’ (08\_04072019M\_3; 173)
- (149) *sa rašia bola artši ine*  
 ‘There are many bears in the mountains.’ (C1)
- (150) *xameleda en*  
 ‘There is a mill.’ (07\_04072019F\_6; 02)
- (151) *opsara ine*  
 ‘There are fish.’ (08\_04072019M\_1; 144)

- (152) *hae bidonja ine*  
 ‘There are such cans.’ (03\_30062019F\_2; 05)

Negative contexts are marked by clausal negation with *utš* (153).

- (153) *do džüsdan=im utš en edži*  
 ‘My wallet is not there.’ (08\_04072019M\_1; 166)

According to Neocleous (2020: 117), the predicative complement generally precedes the copula, although counter examples seem to be possible as well (154). As for adjuncts, both orders SVX (155) and SXV (156) are attested, whereby the latter is likely pragmatically marked (see Neocleous 2020: 148 for focus of PPs). Furthermore, the prepositional phrase can be fronted for pragmatic reasons resulting in XSV (157–159). The spatial adverbial *etši* ‘there’ seems to be always post-verbal (160, 161).<sup>220</sup>

- (154) *esane i mamikandi*  
 ‘There were grandmothers.’ (04\_01072019F\_2; 184)
- (155) *uš tane orendžiðes ine aða so mehele*  
 ‘There are three pupils here in our neighbourhood.’ (08\_04072019M\_2; 021)
- (156) *dört dane za so kato ine*  
 ‘There are four cows at the stable.’ (04\_01072019F\_1; 148)
- (157) *so xorio=muna ena is edone*  
 ‘There was a person in our village.’ (C1)
- (158) *si yoryora ido zias en*  
 ‘There is a Ziya at Gorgoras.’ (01\_04022016F\_1; 018)
- (159) *temetero tso xorio bola mejveðes ine*  
 ‘There are many fruit trees at our village.’ (02\_02022015F\_1; 181)
- (160) *pola ospidæ ine tši*  
 ‘There are many houses there.’ (A1)
- (161) *baja misafe edun etši*  
 ‘There was a relatively long distance.’ (04\_01072019F\_2; 066)

Like with other types of predicates outlined above, the 3SG copula can be omitted (162).

- (162) *so spidi embro enan ido ormi*  
 ‘In front of the house [there is] a river.’ (04\_01072019F\_13; 04)

In interrogatives involving existentials (see also Section 5.2.3.3 for word order in interrogative clauses), word order is variable, although apparently never verb-final. SVX seems to be the most frequent order (163, 164), which includes topicalization of the constituent that is asked for. A PP can be fronted (165), although this does not lead to verb-final order. The copula can be fronted as well (166); this is in line with interrogative existentials in AG where copula-initial

<sup>220</sup> For the hypothetical focus particle =*tši*, see Section 3.2.4.5.1.



position is possible (M. Janse, p.c.). Interestingly, the copula appears also in clause-initial position in indirect polar interrogatives (167), which are argued to have only verb-final orders by Neocleous (2020: 90). The use of the Turkish question particle *mi* is optional in direct and indirect polar interrogatives.

- (163) *muxtera ine sin almanjan*  
‘Are there pigs in Germany?’ (H2)
- (164) *jaralmasia ine mi so xorafi*  
‘Are there potatoes at the field?’ (04\_01072019F\_17; 50)
- (165) *si holandan ine xtina*  
‘Are there cows in Holland?’ (H3)
- (166) *ine mi sa ðorme bola ksila*  
‘Is there a lot of wood on the road?’ (04\_01072019F\_17; 40)
- (167) *lege as dero ine mi dibo na troyume*  
‘He said, “Let me see whether there is something to eat”.’ (04\_01072019F\_12; 10–11)

#### 5.2.1.5.5 Predicative possession

Predicative possession is expressed by transitive have-clauses with the verbal predicate *exo* ‘have’ (168); for the paradigm of *exo* see Section 4.3.1, Table 24. For negative mode, clausal negation with *utš* is applied (169). There is no difference between alienable and inalienable possession.

- (168) *exo tesera peðia*  
‘I have four sons.’ (03\_30062019F\_7; 23)
- (169) *eyo patsi utš exo*  
‘I don’t have a daughter.’ (03\_30062019F\_7; 21)

The possessee is according to Michelioudakis & Sitaridou (2012) marked by accusative case (170), which is the expected case marking of direct objects. Accusative case of the direct object could not be confirmed with the data from the present corpus, though (171), possibly due to the lack of word-final /n/ (see Section 4.2.1.3, Fn. 161).

- (170) *škilon          exo*  
dog.ACC    have.1SG  
‘I have a dog.’ (ROf; Michelioudakis & Sitaridou 2012: 216, ex. 4a)
- (171) *eš aðelfo*  
‘Does he have a brother?’ (A1)

Word order in have-clauses varies between OV and VO. While OV order seems to be the default order (172–177), VO order seems sometimes to be triggered by temporal adverbials (178, 179, but cf. 180). Rare VSO order seems to be also possible (181). In contrastive focus, OSV order can occur (182).

- (172) *andras=im sarija malia eš*  
‘My husband has got blond hair.’ (A1)

- (173) *hat=ena peði iše*  
 ‘She had a daughter.’ (C1)
- (174) *eyo ena batsi exo*  
 ‘I have one daughter.’ (C1)
- (175) *eyo ospi utš exo*  
 ‘I don’t have a house.’ (B1)
- (176) *emistine ōmo utš ixame*  
 ‘We did not have a road.’ (08\_04072019M\_1; 062)
- (177) *bal kad exo apes=am*  
 ‘I have something within me.’ (03\_30062019F\_7; 35)
- (178) *embro utš išen ōromo*  
 ‘Before we did not have a road.’ (08\_04072019M\_1; 212)
- (179) *embro ixame za*  
 ‘Before we used to have animals.’ (08\_04072019M\_1; 202)
- (180) *har za utš exume*  
 ‘Now we don’t have animals.’ (08\_04072019M\_1; 203)
- (181) *aōadžega ba exo eyo botamin*  
 ‘Here I have a valley.’ (08\_04072019M\_1; 275)
- (182) *sekiz dane eyo=xo*  
 ‘I have eight [children].’ (04\_01072019F\_1; 052)

Note that there appears to be instability in the use of subject pronouns vs. possessive pronouns in pronominal predicative possession elicited by means of a translation task from Turkish (183, 184). Apparently, Turkish predicative possession by means of the possessive pronoun (185) interferes with the inherited strategy. Note furthermore that some of the Romeyka data elicited by the translation task yield intransitive possessives of the existential type with a genitive possessive (Stassen 2013b), namely possessive pronoun + copula *ime*, i.e., of the adnominal modifier type “of me there is X” (186, also 187a alternating with 187b; note that this construction is not attested in negative contexts). This seems to be linked to an alternative strategy for definite possessed NPs consisting of a genitive NP for the possessor and the copula *ime* ‘be’ (188a/b, also 186; cf. Stassen 2013b on the typological existence of this construction based on the parameter of definiteness) which was already used with verb-final word order in AG (M. Janse, p.c.). However, since in (188a) genitive possession is used with an indefinite possessor NP, the genitive construction could be also analysed as a direct copy of the Turkish intransitive genitive possessive strategy using the existential particle *var* in clause-final position.

- (183) *eyo eh.. demo ena ospit utš exo*  
 ‘I don’t have a house.’ (C1)
- (184) *teso ena.. esi ena peði ešis*  
 ‘You have a child.’ (C1)

(185) *Benim bir keç=im var.*  
 ‘I have a goat.’

(186) *tema tria ketšia ine*  
 ‘I have three goats.’ (A1)

(187) a. *temo ena peði en*  
 b. *eyo=na peði exo*  
 ‘I have a child.’ (C1)

(188) a. *ato to kitap temo en*  
 ‘This book is mine.’ (A1)  
 b. *teso en*  
 ‘It is yours.’ (A1)

Finally in nominal predicative possession, the example in (189) is found, where it is not quite clear whether the possessor noun has been assigned the neuter definite article (not very likely) or whether rather the genitive case is used here (together with the inherited verb *exo* ‘have’, though), probably induced by the Turkish genitive marking of nominal possessors.

(189) *t=ali ðio yarðelæ iše*  
 ‘Ali had two children.’ (C1)

Note also that occasionally ‘have’-predicates seem to be used to express existentials (190, 191; see Section 5.2.1.5.4 above). It is not clear whether this could be also attributed to influence from Turkish *var* existentials (but see 192).

(190) *eš kliði*  
 ‘There is a key.’ (lit. ‘It has a key.’) (08\_04072019M\_3; 157)

(191) *ula ði eš*  
 ‘Everything is there.’, lit. ‘It has everything.’ (08\_04072019M\_3; 168)

(192) *nenekalar var eh.. esane i mamikandi*  
 ‘There were grandmothers.’ (04\_01072019F\_2; 184)

#### 5.2.1.6 Comparative constructions

Comparative constructions usually concern comparison of inequality (Stassen 2013c). This section deals next to comparatives also with superlatives and equatives. Comparative constructions in Romeyka have not been previously described, although they do not differ significantly from those described for PG in Drettas (1997: 156, e.g., ex. 8).

Romeyka has a grammaticalized means of forming comparative constructions with a subject, a standard NP against which the subject is compared, a preposition as comparison marker, and a gradable adjective as predicate. The standard of comparison is introduced by the preposition *as* (193) or *ap* (194). Possibly, *as* is selected with animate or human standard NPs (see 193, and 195) and *ap* with inanimate. However, according to Drettas (1997), anaphors in the standard NP are more likely to trigger *ap*.<sup>221</sup> In the terminology of Stassen (2013c),

<sup>221</sup> See, e.g., the following example from a Turkish-dominant speaker, who applies *as* in the first clause (ia, note also the incorrect definite article) and corrects in the second to *ap* (ib).

Romeyka has fixed-case locational comparatives of the ‘from’-type, since both *as* and *ap* have an ablative function and mark the standard NP as source of a movement. This is the same type of comparatives like in Turkish (see also Haig 2017 on “ablative”-type comparative constructions as an areal feature of Anatolia), while SMG has particle comparatives with derived case of the standard NP (Stassen 2013c: Map 121A).

Following ex. (193), there seems to be no case marking neither on the standard NP nor on the comparee NP. This is remarkable, as the preposition *as* usually governs accusative case in SMG (also in PG, Drettas 1997: 156, e.g., ex. 3) and the standard NP is typologically usually an object NP (Stassen 2013c). However, considering the “tooth prayer” from RSür (195), the standard NP is in genitive case. Case assignment on masculine and feminine standard NPs seems to be thus subject to idiolectal variation. Unfortunately, no data are available for case marking on pronominal standard NPs.

- (193) *O Mehmetis aso Osmanis alo tranos en.*  
 the Mehmet from.the Osman more big be.3SG  
 ‘Mehmet is bigger than Osman.’ (RSür; Özkan, n.d.; glosses added)

- (194) *temetero to xorio ap=esetero [ ] emorfi-tero en*  
 our the village from=your beautiful-CMP be.3SG  
 ‘Our village is more beautiful than yours.’ (A1)

- (195) *urano na ðond ðome ðond asi gada-s asbri-dero*  
 sky one tooth give.IMP.2SG tooth from.the.F.GEN cat.F.GEN white-CMP  
*asu şgil gajmo-dero*  
 from.the.M.GEN dog.M strong-CMP  
 ‘Oh sky! There’s a tooth, give me a tooth whiter than the cat’s, stronger than the dog’s.’  
 (RSür; Özkan 2013: 145; glosses added)

The constituent order in comparative clauses is, like often in predicative clauses, verb-final (cf. Section 5.2.1.5). The comparee NP precedes the standard NP. The standard NP can be elliptical with an anaphoric pronoun, if the referent of both comparee and standard NP is identical, as in (194) above, or clear from the context (196). The comparative marking is shown on the predicate either by means of a comparative suffix or an adverb<sup>222</sup> and followed by the copula in clause-final position. However, like it is the case for predicate adjectives in general, the word-final third-person singular copula is frequently left out (197, also see 198 below).

- (196) *temetero.. emist exome ospi ama tatinuna ap=emeteroji emorfi-tero en jengi en*  
 ‘We have a house but theirs is nicer than ours, it is new.’ (A1)

- (197) *haşin to spidin habalahan aso spidin trani deron*  
 ‘This house is bigger than that over there.’ (H2)

The choice of comparative marker is subject to micro-dialectal variation: While (193) from RSür forms the comparison periphrastically with *alo* ‘more’, ROF still uses predominantly the inherited comparative suffix *-tero(n)* (194) which is derived from AG

- 
- (i) a. *temetero o xorio as=esetero emorfi-tero en* (A1)  
 b. *temetero to xorio ap=esetero emorfi-tero en* (A1)  
 ‘Our village is more beautiful than yours.’

<sup>222</sup> But note ex. (ii) from a heritage speaker who applies double marking by use of both strategies.

- (ii) *hatşino t=osbidin ap=etşesao t=osbidin alo trani-deron edune*  
 ‘This house was bigger than that over there.’ (H1)

verbal stem+*o+teþ*+inflectional ending. It is, however, not fully clear which factors determine the selection of both comparative strategies. Interestingly, both PG and Cappadocian use for simple comparatives (i.e., without any focus) just the positive form of the adjective (Drettas 1997: 156-157; Dawkins 1916: 116, respectively). Note also that Neocleous (2020: 134, ex. 39b) still attested the ancient comparative in *-ion*, at least in the (possibly lexicalized) form *kal-ion* ‘better’, which, however, does not figure in the present corpus (but cf. ex. 208 from P. Mackridge below). According to Papadopoulos (1955: 56), the periphrastic type of comparison with *alo* is the normal type of comparison in PG, whereas the inherited type with the ancient suffix *-tero* is used rarely. In the present Romeyka corpus, comparatives with *-tero(n)* dominate in elicited data, while periphrastic *alo* comparatives occur in free speech (198, 199). Turkish also applies comparison marking on the predicate by means of a particle, *daha* ‘more’, which might support the use of the periphrastic comparison strategy. Note that occasionally also the Turkish adverb *fazla* ‘more’ is borrowed as marker of comparison (199b, 200).

(198) *t=adaha d=ospid alo mikro*  
 ‘The house here is smaller.’ (B1)

(199) a. *to šehiri alo masraflı en* (08\_04072019M\_1; 312–313)  
 b. *i šeher kalo fazla masraflı eftes* (08\_04072019M\_1; 314)  
 ‘The city is more expensive.’

(200) *talo fazla tar en*  
 the.other more narrow be.3SG  
 ‘The other is narrower.’ (A1)

Unfortunately, not enough examples are attested to determine whether the comparative suffix *-tero(n)* inflects like other adjectives, namely for gender and plural. However, since both comparative forms (periphrastic and affixal) are also used as temporal adverbs (201a/b, note, however, that in the adverbial form – like in superlatives (see below) –, *olon* is used instead of *alon*), there is evidence for the fact that *-tero(n)* is sensitive to number (but note the apparent lack of number agreement in 201b vs. 201c). Furthermore, ex. (212) from superlatives below features gender inflection of *-tero(n)* for feminine NPs as *-teresa*. This suggests the following inflectional paradigm for *-tero*: *-tero(s)*.M.SG, *-tero(n)*.N.SG, *-teresa*.F.SG, *-tera*.PL (cf. Holton et al. 2019: 841).

(201) a. *olon embro(n)*.SUP  
 ‘before, in old times’ (03\_30062019F\_11; 084)  
 b. *embro-dera*.CMP  
 ‘before, in old times’ (01\_04022016F\_1; 113)  
 c. *embra-dera*.CMP  
 ‘before, in that times’ (02\_02022015F\_1; 043)

Superlatives are built periphrastically by *olon* (N.SG < AG ‘all-GEN.PL) as superlative marker on the predicate (202, 203; cf. Drettas 1997: 158-159). As a pronominal element, *olon* is normally involved in determiner spreading (204, 205) and it is not clear what leads to the omission of definite articles in exs. (202) and (206).

(202) *olon emorfo ospi temo en*  
 all beautiful house POSS.1SG be.3SG  
 ‘The nicest house is mine.’ (A1)

- (203) *olon o tranon ađelfo=m*  
 ‘my eldest brother’ (Saráchos, example P. Mackridge, p.c.)
- (204) *du dūnja olon da tehlikelija da dobe ta rđome ta viradže hadudžega ine*  
 ‘The world’s most dangerous places.. roads.. road bends are over there.’  
 (08\_04072019M\_2; 090–092)
- (205) *audo d=ospī eh.. so xorio en tranon do en d=ospidin.. olon dranon do en d=ospī*  
 ‘This house is the largest of the village.’ (B1)
- (206) *hado to spidin sin klisuran olon tranon ospidin*  
 ‘This house is the largest of the village.’ (H2)

Note that Neocleous (2020: 196, ex. 154b) also presents a pronominal superlative based on the genitive plural of the quantifier *ulo* ‘all’ (207); but cf. other superlatives in ROF as spoken in Saráchos (Uzungöl) (exs. 208–210, data P. Mackridge, p.c.). In general, *ulo(n)* is arguably not identical with *olon* (see Section 3.2.3).

- (207) *ulunon o mikron*  
 all.GEN the.NOM young.NOM  
 ‘the smallest’ (Neocleous 2020: 196, ex. 154b)
- (208) *t=olon embrotikon*  
 ‘the very first’ (Saráchos, example P. Mackridge, p.c.)
- (209) *olon embron*  
 ‘at the very beginning, first of all’ (Saráchos, example P. Mackridge, p.c.)
- (210) *olon kalon*  
 ‘[the] best of all’ (Saráchos, example P. Mackridge, p.c.)

Sometimes, *olon* is accompanied by the comparative marker *-tero(n)* (211, 212, cf. also Fn. 222 above). It is not clear whether this implies a semantic difference compared to superlatives formed just with *olon*.

- (211) *havu t=ospitin so mehelen=muna tosdon t=ospidin olon trani-deron edune*  
 ‘This house was the biggest of the houses of our neighbourhood.’ (H1)
- (212) *eyo olon emorfi-teresa eyo ime*  
 ‘I am the most beautiful of all.’ (A1)

As for the marking of the standard NP, there seems to be variation. Unlike with comparatives,<sup>223</sup> in superlatives the standard NP is often omitted and locational adjuncts with *s* ‘at, to’ are added (205, 206, 211). In (204), an (elliptical) genitive is used for the standard NP. Only ex. (211) shows a full standard NP including a locational adjunct. If reference to the standard NP is omitted, *olon* seems to take its place as anaphoric pronoun (202, 212, also cf. 207).

<sup>223</sup> Note the rare nominalized<sup>2</sup> superlative with *as* ‘from’ in (iii).

(iii) *pao ebero asa kalase*  
 ‘I go and buy from the best.’ (03\_30062019F\_6; 61)

The word order parameter in superlative constructions is predominantly Adj–Cop whereby the adjective precedes the copula (213), but Cop–Adj order is also possible (214). Again, in clause-final order, the third-person singular copula can be left out (215).

- (213) *olon dranisa eyo mune*  
 ‘I was the oldest.’ (04\_01072019F\_2; 070)
- (214) *olon emorfo temo en to fistani=m*  
 all beautiful my be.3SG the dress=POSS.1SG  
 ‘My dress is the most beautiful of all.’ (A1)
- (215) *havu do spidi so.. aso xorio tu spidi olon do trano [...]*  
 ‘This house is the largest of the village’s houses.’ (C1)

For the formation of equatives insufficient data are available. However, a single example (216) shows that equatives are formed with the preposition *omon/emon* (for the use of *omon* as conjunction see Section 5.3.3.1.1, i.a.) which links the two comparee NPs. Both comparee NPs can be omitted in the equative clause, if clear from the context. Word order is (like often the case with predicatives, see Section 5.2.1.5.2) verb-final.

- (216) *temo tospi mikro en ama teso pal emon temon=a mikro en*  
 ‘My house is small but yours is as small as mine.’ (A1)

## 5.2.2 Valency-changing processes

### 5.2.2.1 Causatives

No morphological causatives are attested in Romeyka. Note, however, that also in SMG, there is no productive morphological means of expressing causatives; they are either realized (a) based on the AG/MedGr verbal suffix *-izo* (Özkan 2013: 147), (b) periphrastically with ‘do’, or (c) by means of non-causative labile verbs that may have an intransitive (or transitive) meaning and may take an object (either overt or understood from context), e.g., *se jelasa* ‘I made you laugh’, lit. ‘I laughed you’. Further research on causativity in Romeyka is required in order to test for any of these options. However, in Romeyka a periphrastic paraphrase equalling a causative meaning may be formed by a matrix verb ‘let, allow’, and a complement clause introduced by *na* (217) (Sitaridou 2014a: 125).

- (217) *Efikane=sas na skaftete ta xorafæ=suna.*  
 let.AOR.3PL=OPN.2PL PRT dig.2PL the fields=POSS.3SG  
 ‘They let you dig his fields.’ (ROf; Sitaridou 2014a: 125; glosses modified)

Furthermore, in the Romeyka corpus use is made of Turkish loan verbs including the Turkish causative suffixes *-DIr* or *-t-* (Nemeth 1962/2020: 87) to express causativity (218, 219, 220).

- (218) *ta patsiðes e..gez-dur-epsane*  
 ‘The girls were not walked around.’ (02\_02022015F\_1; 071)
- (219) *tut-tur-evane=masine nestia*  
 ‘They made us fast.’ (04\_01072019F\_2; 127)

- (220) *ta patsiðes utš eku-t-evane*  
 ‘The girls where not taught.’ (02\_02022015F\_1; 039)

Finally, in a translation task with the Turkish causative prompt *yönlendir-mek* ‘instruct’ active clauses with a nominalized complement are provided (221a/b).

- (221) a. *i aiše tin batsi=nates ebolise diri so piðenimo* (C1)  
 b. *i aiše tin batsi=nades diri so biðimo erlaeps* (B1)  
 ‘Ayşe teaches her daughter how to make cheese.’

#### 5.2.2.2 Passives

The syntax of passives in Romeyka is still awaiting investigation; but see Section 4.3.5 on voice above. The syntax of datives under passivization in RSür has been investigated by Michelioudakis & Sitaridou (2012). They note that a theme argument can be a regular subject of a passive verb (222, 223a), while a benefactive or recipient cannot advance to subject under passivization (223b). Interestingly, in (222) a passive aorist paradigm would be expected, however, the form provided shows active aorist endings (M. Janse, p.c.). The example in (224a) seems to feature an active present ending to be used in passive meaning, possibly mirrored on the Turkish equivalent (224b).

- (222) *To harti eyrafte tin Aiše.*  
 the.NOM letter.NOM written.PASS.3SG the.ACC Ayşe.ACC  
 ‘The letter was written for Ayşe.’ (RSür; Michelioudakis & Sitaridou 2012: 236; glosses modified)

- (223) a. *I para tin Aiše eðoste.*  
 the.NOM money.NOM the.ACC Ayşe.ACC given.PASS.3SG  
 ‘The money was given to Aise.’ (RSür; Michelioudakis & Sitaridou 2012: 236, ex. 44a; glosses modified)  
 b. *\*I Aiše eðoste tin paran*  
 the.NOM Ayşe.NOM given.PASS.3SG the.ACC money.ACC  
 ‘Ayşe was given the money.’ (RSür; Michelioudakis & Sitaridou 2012: 236, ex. 44c; glosses modified)

- (224) a. *ap-að-tšeka apolisimo utš eftene*  
 b. *Buradan gönderme yapılmıyor.*  
 ‘There is no mail delivery here.’ lit. ‘Mail delivery is not done here.’ (Tursun 2019: 142)

In the present ROF corpus, no passives could be attested in natural speech. Elicitation by means of a translation task resulted in a copy of the Turkish prompt (225a): either a full copy including lexical borrowing of the Turkish analytical marker *tarafından* ‘by’ but ignoring the Turkish passive affix *-l-* (225b) or a copy of the Turkish word order with simultaneous adherence to Romeyka argument structure (including object marking) but without morphological passive marking which allows for misinterpretation (225c, 226b/c). The realization in (225c) resembles the competence of aphasic patients in Turkish who misinterpret passives on the basis of default word order under disregard of morphological cues (Yarbay Duman et al. 2011).



- (225) a. *Fatma keçi tarafından vuruldu*  
 b. *i fadime kedži **tarafından** endokanane* (B1)  
 c. *i fadime endotše to ketši* (C1)  
 ‘Fatma was hit by the goat.’
- (226) a. *Hasan’a babası tarafından izin verilmedi*  
 b. *o xasanis ton džiri=nad ežin utš eđodže* (B1)  
 c. *o hasanis to peđin=at utš eđotše izini* (C1)  
 ‘Hasan did not get the approval of his father.’ lit. ‘Hasan was not given..’

### 5.2.3 Non-declarative clauses

#### 5.2.3.1 Imperative clauses

An imperative clause consists at least of a single predicate (227). If negated, the negative particle precedes the verb (228, 229). The way patient arguments are coded in imperative clauses does not differ from that of declarative clauses. Word order in imperatives seems to be predominantly verb-initial (230); local adjuncts and goals are mostly post-predicate (231, 232, also 237), although verb-final orders arise after topicalization of direct objects (233), focus (234) and with light verb constructions calqued on Turkish (235). Full object pronouns appear predominantly in post-verbal position (236a, also 229), while weak object pronouns are enclitic anyway (236b-238).

- (227) *ela*  
 ‘Come!’ (08\_04072019M\_4; 28)
- (228) *mi bas*  
 ‘Don’t you go!’ (04\_01072019F\_13; 11)
- (229) *hiš ma afinis*  
 ‘Never leave it!’ (01\_15022015F\_1; 25)
- (230) *arajepson adina*  
 ‘Call her!’ (04\_01072019F\_17; 67)
- (231) *ayome katuhaž*  
 ‘Go down there!’ (07\_04072019F\_6; 40)
- (232) *apopse elate s=emena*  
 ‘Come to me tonight!’ (04\_01072019F\_2; 186)
- (233) *havu to peđin habađahandžeka eparen [...]*  
 ‘Take this child from here! [...].’ (02\_21042018M\_2; 19–20)
- (234) *hem d=apiđe ba fa hem do dajaye fa*  
 ‘Eat the pears as well as the beating!’ (04\_01072019F\_2; 296–297)
- (235) *emena ifeti jardimi bisun*  
 ‘Help me this year!’ (08\_04072019M\_3; 100)

- (236) a. *đos emen*  
 ‘Give her to me!’ (01\_04022016F\_1; 148)  
 b. *đo=me nero na pino*  
 ‘Give me water so that I shall drink.’ (A1)
- (237) *feron=a hađa-mer*  
 ‘Bring it to this side!’ (08\_04072019M\_3; 001)
- (238) *feride=me u=boro na payo tš=erxume*  
 ‘Bring it to me, I cannot go!’ (04\_01072019F\_2; 201)

### 5.2.3.2 Exclamatory clauses

The Romeyka corpus contains a single hit of an (elliptical) exclamatory clause following the English type of grammaticalized ‘what’/‘how’-clauses (239). In Romeyka, the exclamative clause is preceded by the Turkish interjection *ya* expressing astonishment and the imperative *der* ‘look’. The exclamatory clause consists solely of a NP with ellipsis of the copula.<sup>224</sup> Due to the reduced form of the determiner *d=* it is unclear whether this is simply the definite neuter article or the interrogative particle *ndo* ‘what’. Exclamatory clauses in Romeyka await further investigation.

- (239) *ja de d=omorfo ospidi*  
 ‘Look, what a nice house!’ (04\_01072019F\_12; 05)

### 5.2.3.3 Interrogative clauses

Interrogative clauses share for a large the word order of declarative clauses, although movement of the *wh*-element in content questions is the inherited Greek feature: In content questions, both clause-initial question elements and topicalized arguments leading to question elements in immediate pre-verbal position exist. This means that in direct interrogatives, SOV orders are frequently attested since *wh*-elements are obligatorily left-dislocated (Neocleous 2020: 180). According to Neocleous (2020), interrogative clauses with direct questions, i.e., both polar and content questions, have the same unmarked word order like matrix declarative clauses: SVO. The following other orders (resulting from movement) are also attested: SOV, OSV (Neocleous 2020: 87). Polar questions are frequently marked by the Turkish interrogative particle *mi*, which frequently occurs in clause-final position after VO order but also allows for verb-fronting and for putting focus on various other constituents, like with the Turkish interrogative particle *mI*. Inherited polar questions without the Turkish interrogative particle, are solely marked by an interrogative intonation contour. The intonation contour of polar questions does not differ from that of constituent questions but is usually a bit more pronounced in the latter (Section 2.4.2). Predicate nominals with the copula *ime* ‘be’ seem to yield a slightly different word order, as they very frequently display fronted arguments with question elements in immediate pre-verbal position resulting in OV. Embedded interrogative clauses are pre-dominantly SOV (such as subordinate declarative clauses) with OSV order being also attested (Neocleous 2020). However, since SOV orders are at the present corpus very frequent even in direct content questions (and declarative clauses), word order of embedded interrogatives seems in fact not to differ considerably from that predominantly found in direct interrogatives, i.e., initial or pre-verbal *wh*-elements in content questions, and SVO order + *mi* in polar questions.

<sup>224</sup> Note that in the narrative texts elicited by means of a picture prompt, the copula is frequently left out in existentials (Section 5.2.1.5.4), e.g., *havuda da rašia* ‘here [are] the mountains’ (04\_01072019F\_12; 01 and following segments).

## 5.2.3.3.1 Polar questions

Unmarked word order of polar questions is SVO, with SOV and OSV orders arising due to movement for pragmatic purposes (Neocleous 2020: 87). Polar questions are solely marked by intonation (ex. 240a/b; for the typical intonation contour of polar questions with a rise followed by a decline, see Section 2.4.2).

- (240) a. *eš kliði*  
 ‘It has a key.’ (08\_04072019M\_3; 156)  
 b. *eš zo*  
 ‘Does he have animals?’ (04\_01072019F\_5; 29)

However, there is extensive borrowing of the Turkish question particle *mi*, especially in Turkish-dominant bilingual speakers. Note, however, that *mi* seems to occur less frequently in existentials and (third person) possessive clauses with the verb *exo* ‘have’ (see ex. 240; but cf. *xoraf eis mi* ‘do you have a field’ (03\_30062019F\_6; 51)). The Turkish interrogative particle is borrowed in the invariant form *mi* (occasionally in free variation with *me*; 241–243).<sup>225</sup> *Mi* is used in addition to the interrogative intonation contour.

- (241) *ekuses=a mi*  
 ‘Did you hear it?’ (01\_15022015F\_1; 27)
- (242) *etšinena eynorizez me*  
 ‘Do you know her?’ (01\_04022016F\_1; 020)
- (243) *hen na stin so inegölü joksa so xorio mi na stjetšidi panda*  
 ‘Will you come to Inegöl or will you be the whole time at the village?’  
 (03\_07072019F\_1; 45)

In terms of default word order, *mi* appears generally in post-verbal position (244a–246). This is also the case in Turkish, although in Turkish this leads to the particle occurring by default in clause-final position (244b). Interestingly, since SMG has a polar question particle in clause-initial position (Dryer 2013d, Map 92A) and Turkish in clause-final position, the position in Romeyka appears to be in middle position. In clauses with more than one verb, e.g., in modal clauses, *mi* follows the main verb (247, 248).

- (244) a. *ana ebijes mi so xorio*  
 b. *Anne, köye gittin mi?*  
 ‘Mother, did you go to the village?’ (04\_01072019F\_17; 04)
- (245) *ler mi i mana=s*  
 ‘Does your mother say this?’ (07\_04072019F\_5; 51)
- (246) *n=evriški mi tši opsaræ*  
 ‘Will he find fish there?’ (04\_01072019F\_12; 27)
- (247) *tšimate mi tše kahete*  
 ‘Is he sleeping?’ (A1)

<sup>225</sup> The invariant form of the Turkish question particle is partly explained by particularities of regional Trabzon Turkish phonology, which features reduced vowel harmony due to contact with Greek (see Brendemoen 2002).

- (248) *kalos u=poris me na inese*  
 good NEG=can.2SG Q PRT be.2SG  
 ‘Will you not be good?’ (Tursun 2019: 147; glosses added)

Unlike in Turkish, person marking can never be expressed on the question particle (249a vs. 249b) but is rather always expressed on the verb. Enclitic (weak) object pronouns cannot be separated from the verb, so *mi* follows the clitic pronoun (see 241 above). Strong object pronouns seem to occur predominantly in pre-verbal position, resulting in the question particle occurring in clause-final position (see 242 above); topicalization is likely to account for this order. It would be a topic for further investigation whether the occurrence of the question particle correlates to (some kind of) topicalization or focus. According to Michelioudakis & Sitaridou (2016: 26), the question particle occurs in ROF only in indirect questions (optionally) and obligatorily in direct questions “of total ignorance” (thus, roughly corresponding to the notion of polar questions).

- (249) a. *eyridžes mi* (03\_30062019F\_11; 071)  
 b. *Anlıyor mu-sun?*  
 ‘Do you understand?’

Like in Turkish, *mi* shows a great flexibility in placement, i.e., it can occur after all constituents for constituent questions (250). For example, fronting of NPs for the sake of topicalization is possible; the interrogative particle occurs at the end of the fronted NP (251). Even fronting of the verb is possible with *mi* in post-verbal position (252).

- (250) *teso i patsi layo en emorfo mi en*  
 POSS.2SG the.NOM daughter.NOM how be.3SG beautiful Q be.3SG  
 ‘How is your daughter, is she beautiful?’ (A1)

- (251) *tesetera ta tahadžes mi inane haje bola šimu*  
 ‘Was there so much rain on your side (as well)?’ (03\_30062019F\_6; 49)

- (252) *ine mi sa đorme bola ksila*  
 ‘Is there a lot of wood on the road?’ (04\_01072019F\_17; 40)

Like in Turkish, *mi* is also used in alternative (Section 5.2.3.3.3) and tag (Section 5.2.3.3.2) questions as well as in embedded questions (Section 5.2.3.3.5).

#### 5.2.3.3.2 Leading questions

Romeyka does not code any tag questions like in English. However, the positive responsive particle *ne* (253, 254) and the Turkish conjunction *yoksa* (see also Section 5.2.3.3.3) are used in leading questions to ask for confirmation at the end of an interrogative clause. In (255), the inverse question carries additionally to the particle *joksa* the interrogative particle *mi*.

- (253) *ta laxana layo koševume dž=eftame laxana ne*  
 ‘How we prepare? cabbage, right?’ (translation unclear; 03\_30062019F\_11; 081)

- (254) *inete mi ne*  
 ‘Did it work, didn’t it?’ (03\_30062019F\_2; 33)

- (255) *adžaba maksuz efte=a mi joksa mi*  
 ‘Does she actually do it intentionally, or not?’ (01\_28062019F\_3; 30–31)

A request for confirmation of a declarative clause can be marked without particle solely by interrogative intonation, i.e., a flat intonation contour with a final drop, e.g., *kalos e* ‘he is well, isn’t he’ (04\_01072019F\_5; 28). As a particle asking for confirmation, regional Turkish *hemi* is used (256).

- (256) *evra bola insani - hemi dona*  
 ‘I found many people.’ - ‘Really, whom?’ (04\_01072019F\_5; 13–14)

#### 5.2.3.3.3 Alternative questions

Alternative questions in Romeyka make use of the Turkish contrastive particle *yoksa* ‘or’ and/or may adopt the Turkish construction of alternative questions: In Turkish, both alternative clauses are marked with the question particle *mi* placed after each alternative phrase; optionally, the contrastive options are coordinated by the conjunction *yoksa* ‘or’ (257; Göksel & Kerslake 2005: 254).

- (257) *Eve gidecek misin (yoksa) kalacak mısın?*  
 ‘Will you go home or will you stay?’

In Romeyka, alternative questions are often marked by the coordinating conjunction *joksa* ‘or’ (258; see also 04\_01072019F\_17; 33–34), albeit not obligatorily (see 260). The borrowed interrogative particle *mi* is also not obligatory, but can occur for example, to mark contrastive focus (259), although in Turkish both alternative phrases would have been marked by *mi*. Finally, the coordinating conjunction *joksa* can be left out in which case both contrasting options are marked by the question particle (260), which is identical with the Turkish strategy of alternative questions.

- (258) *ina dina joksa mutš eyriga*  
 ‘Does she do this intentionally or do I not understand?’ (01\_28062019F\_3; 32–33)
- (259) *hen na stin so inegölü joksa so xorio mi na stjetšidi panda*  
 ‘Will you come to Inegöl or will you be the whole time at the village?’  
 (03\_07072019F\_1; 45)
- (260) *do tšai ađatška me na pinume s=ospidi me*  
 ‘Will we drink the tea here or in the house?’ (08\_04072019M\_2; 161–162)

According to Neocleous (2020: 158), contrastive foci are in Romeyka realized in pre-verbal position; any phrase can appear in contrastive focus (Neocleous 2020: 160).

#### 5.2.3.3.4 Content questions

Unlike polar questions, content questions (or question-word questions, Payne 1997: 299) contain an interrogative phrase which is in English often formed by a *wh*-element (Dryer 2013m). While SMG has, as in most of the languages of Western Europe, obligatory fronting of *wh*-words to clause-initial position (Dryer 2013m, Map 93A), Turkish is normally analysed as having in situ *wh*-words (interrogative pronouns) commonly in pre-verbal position which can appear in other (pre-verbal) positions as well, (Göksel & Kerslake 2005: 265–266). Romeyka seems to have moved closer to Turkish by losing obligatory fronting of *wh*-words. In Romeyka, two different orders of interrogative phrases occur in neutral questions which do not depend on the type of interrogative element (pronominal or other). But all interrogative elements occur obligatorily in pre-verbal position; *wh*-elements cannot remain in situ (Michelioudakis & Sitaridou 2016). Firstly, although interrogative phrases are not obligatory

in phrase-initial position in Romeyka like in SMG, clause-initial interrogatives are very frequent (261). According to Neocleous (2020: 179), the position of *wh*-elements in Romeyka is strictly order preserving (i.e., SVO). In the present corpus, this entails frequently immediate pre-verbal position, although topicalized NPs can interpolate between the *wh*-element and the predicate (261d). Appositions are fronted together with the question word ('pied-piping'; 262, see also Michelioudakis & Sitaridou 2016: 26–27).

- (261) a. *tina na pulis=a*  
 'Whom will you sell this?' (03\_07072019F\_1; 07)  
 b. *s posa i da yarðela*  
 'How many children do you have?' (01\_04022016F\_1; 080)  
 c. *pu je i mana=s*  
 'Where is your mother?' (01\_04022016F\_1; 038)  
 d. *laya yomare fordumune tše erxumunesten*  
 'How did we move the loadings?' (04\_01072019F\_1; 195)  
 e. *oðande de udž ebidžes*  
 'Why didn't you make [another]?' (01\_04022016F\_1; 083)  
 f. *donas en so dolabi*  
 'What is in the wardrobe?' (04\_01072019F\_17; 10)  
 g. *dohna na leyo alom*  
 'What else shall I say?' (03\_07072019F\_1; 21)
- (262) *os pudžega pijede*  
 'Till where did you go?' (04\_01072019F\_17; 33)

Secondly, interrogative elements occur in immediate pre-verbal position, i.e., in focus position, with a subject (263, 264) or object (265, also 253 above) in initial position. According to Neocleous (2020), SVO is the unmarked order in content questions, although SOV and OSV orders are possible as well; all *wh*-elements are obligatorily left-dislocated (Neocleous 2020: 180). SOV order is especially frequent among interrogative predicate nominals (266–269), but cf. (261b/c/f above). Apart from arguments, locative adjuncts are often fronted (270). Note that interrogatives following the argument(s) differ strongly from the order of SMG (271), (irrespective of whether this is topicalization or the default order), so influence from Turkish interrogative clauses with interrogatives in immediate pre-verbal position is possible here (see also Neocleous 2020: 242).

- (263) *i mana=s d=eftjej*  
 'What is your mother doing?' (03\_07072019F\_1; 02)
- (264) *i nife=nades pote n=arde*  
 'When will her daughter-in-law come?' (04\_01072019F\_5; 20)
- (265) *emenan dona na ðiyune*  
 'What do they give to me?' (08\_04072019M\_1; 272)
- (266) *havudo pios en*  
 'Who is this?' (03\_07072019F\_1; 07)
- (267) *to xorio=s pio en*  
 'Who is your village?' (01\_04022016F\_1; 013)

- (268) *teso i patsi layo en*  
 ‘How is your daughter?’ (A1)
- (269) *to tsubađi ta fasüljas laya sane*  
 ‘How were the corn and the beans?’ (04\_01072019F\_17; 49)
- (270) *iben mana daha ndo kahese*  
 ‘He said, “Mother, why are you sitting here?”’ (03\_30062019F\_6; 30)
- (271) *Layo en teso i patsi?*  
 ‘How is your daughter?’ (SMG, example M. Janse, p.c.)

According to Michelioudakis & Sitaridou (2013: 360, 2016), multiple interrogatives are (unlike in SMG) acceptable in Romeyka; two or more *wh*-phrases can be fronted in the left periphery. Movement is obligatory as *wh*-elements are not permitted *in situ*. As for the order of interrogatives, apart from discourse-linked interrogatives, multiple *wh*-fronting is order-preserving (Michelioudakis & Sitaridou 2013: 363; see ex. 272).

- (272) *Pios tinan pote efilise?*  
 who.NOM who.ACC.HUM when kissed.3SG  
 ‘Who kissed whom and when?’ (ROf; Michelioudakis & Sitaridou 2013: 363)

#### 5.2.3.3.5 Embedded interrogatives

Embedded questions are often arguments of a verb like ‘know’ or ‘find out’. Only those interrogative phrases are considered here that are arguments, and not, for example, locational adjuncts. Interrogative arguments in embedded content questions include *wh*-interrogative elements like *doxna* ‘what’ (273, 274), *ndo* ‘what’ (275–279), *pio* ‘who’ (280), *poso(n)* ‘how much’ (281). According to Neocleous (2020), the unmarked word order in indirect questions (both polar and content questions) is like in subordinate declarative clauses SOV with OSV order also attested. In the present corpus, however, word order in the embedded clause seems not to differ from word order in regular content questions since they also show predominantly OV due to the obligatory fronting of the interrogative elements that occur in pre-verbal position. The order of matrix and subordinate clause is variable, although initial matrix clauses seem to be the unmarked order (see also ex. 28 in Michelioudakis & Sitaridou 2016: 12). Note also the exs. (274, 280) where the argument of the embedded predicative clause is extracted and moved to focus position of the matrix clause (possibly also 278 although the function of *hare* is ambiguous, see Section 3.2.4.5.2).

- (273) *as derume [d=alo doxna en]*  
 ‘Let us see what else will happen.’ (04\_01072019F\_12; 23)
- (274) *t=aleyo kseris dohna e*  
 ‘Do you know what ‘aleyo’ is?’ (08\_04072019M\_2; 066)
- (275) *to leyo utš ekserun*  
 ‘They don’t know what I say.’ (01\_28062019F\_2; 41)
- (276) *do thelis bas koftjis tše feris*  
 ‘What you want, you go, cut and bring [it].’ (08\_04072019M\_1; 219)

- (277) *nde na troyume u=ksero*  
 ‘I don’t know what we will eat.’ (02\_05072019F\_1; 25)
- (278) *[hare] iba [nde š=iftao]*  
 ‘I said, “What am I doing now?”’ (03\_30062019F\_11; 045)
- (279) *iben adona [d=ebines adadž]*  
 ‘She said to him, “What did you do here?”’ (04\_01072019F\_13; 47)
- (280) *akome kseris pia ine*  
 ‘Do you still? know who they are?’ (02\_21042018M\_2; 04)
- (281) *n=eksernes poso ayapo ta yulia*  
 ‘If know how much I love cabbage soup.’ (Tursun 2019: 224)

Note that the examples provided above comprise different kinds of headless relatives as well as examples of (in)direct speech which are not straightforward to analyse and ambiguous between different readings. Headless and embedded object relative clauses including interrogative elements are relatively frequent (282, 283); for their discussion see also Section 5.3.4.

- (282) *Eyrapses etšinon do ipamen?*  
 record.AOR.2SG that.ACC REL say.AOR.1PL  
 ‘Did you record what we have said?’ (Gandon 2016: 221; glosses modified)
- (283) *t=ado pe at jeft i nife=m eftjej*  
 ‘That what is made (lit. what she makes), is done by my bride.’ (07\_04072019F\_8; 14)

An example for an embedded interrogative subject argument with the interrogative element *to* ‘what’ is shown in (284). For embedded interrogative arguments with an indefinite pronoun like *opi/oti* ‘whoever’, see ex. (273) in Section 3.2.2.7 for a subject argument and ex. (274) in that Section for object arguments.

- (284) *har to leyo inšallah ormofa ine*  
 ‘What I say now is hopefully nice.’ (03\_07072019F\_1; 33)

Interrogative oblique arguments can be embedded in pre-verbal position (285–287; see also Section 5.4 on adverbial clauses).

- (285) *erpu en epejename*  
 ‘Wherever it was, we went [there].’ (09\_04072019\_7; 14)
- (286) *džaka pu en*  
 ‘Down there is where she is.’ (translation uncertain; 07\_04072019F\_5; 37)
- (287) *fasulijas bote en erxume bero=na*  
 ‘When there are beans, I come and take them.’ (03\_30062019F\_11; 041)

Embedded polar questions are formed with the interrogative particle *mi* and carry the meaning of ‘whether’ (288, 289); see also Section 5.3.2.1.1. The word order in the embedded clause is interestingly in both ex. (288) and (289) SVO/SVX with the interrogative particle in clause-final position and a sentence-initial matrix clause, although according to Neocleous (2020: 90),



indirect polar questions are SOV (290). But see also the fronted embedded nominalization in the purpose adverbial clause in (291).

- (288) *u=ksero θane fayane eb edže mi*  
 ‘I do not know whether they came and ate from it.’ (03\_30062019F\_11; 067)
- (289) *as derume na krui etunesine mi*  
 ‘Let us see whether he will hit him?’ (04\_01072019F\_12; 15–16)
- (290) *rotas=me                      alis              tin              aišen              efilisen*  
 ask.NOM.2SG=OPN.1SG   Alis.NOM   the.ACC   Ayşe.ACC   kiss.AOR.3SG  
 ‘You asked me, did Ali kiss Ayşe?’ (Neocleous 2020: 412, ex. 193; glosses modified)
- (291) *apsimo s=epario me erθes*  
 ‘Did you come to take fire?’ (Tursun 2019: 199)

### 5.3 Complex clauses and clause combining

#### 5.3.1 Coordination

##### 5.3.1.1 Clause coordination

Coordination at clause level shows the lowest degree of grammatical integration of two or more verbs. The strategies of conjoining clauses are largely identical to that of phrases, although not all conjunctions used for clause coordination can be used at phrase-level (see Section 3.2.6.1). The coding strategies usually involve coordinating conjunctions in intermediate position between the clauses, but also a zero strategy, i.e., mere juxtaposition of clauses, is possible (Payne 1997: 337; Stassen 2013d; for juxtaposition of VPs see this section below). While the coordinating conjunctions for disjunction and adversative coordination are borrowed from (Ottoman) Turkish, the conjunction for ‘and’ *tše* is inherited (see also Section 3.2.6.1).

Clauses (as well as words and phrases) are combined by the coordinating conjunction *tše* ‘and’ which appears between two clauses (for details on the various functions of *tše* see Section 3.2.6.1). Enumerations can be also juxtaposed without conjunction (292).

- (292) *epiname axlaðia ekoftame me di kerendi tše ta za ederename*  
 ‘We made hay, we cut [it] with the scythe and we looked after the cows.’  
 (09\_04072019\_7; 10–12)

Disjunction is marked by Turkish/Ottoman loans: *(ve)jaxod(e)* ‘or’ (293–295), *ja(da)* ‘or’ (296), *joksa* ‘if not, or, otherwise’ (297). Like *tše*, the conjunctions appear between the clauses, but see ex. (295) where the conjunction may occur at the beginning of a phrase, if a pause after the previous clause suggests two separate clauses. Turkish disjunctive *veja* ‘or’ occurs in the corpus only at phrase level (see also Section 3.2.6.1).

- (293) *xandilaza=na jahoda ekuntena=na*  
 ‘I tickle her or I hug her.’ (02\_02022015F\_1; 144–145)

- (294) *havu to peđin habađahandžeka eparen ja do hūkūmetin na điyum=a jaxota hao to peđin n=eperis*  
 ‘Take this child from here or we will give him into a children's home or you will take this child from here.’ (02\_21042018M\_2; 19–21)
- (295) *vejaxute nejenene to zo epukan=eθe n=epiname fila*  
 ‘Or when a cow gives birth, we put leaves under it.’ (04\_01072019F\_1; 172–173)
- (296) *d=akšemi n=arde ište jada ja.. belki na bai so xorio*  
 ‘He will come in the evening or maybe he will go to the village.’ (04\_01072019F\_5; 60)
- (297) *os pudžega pijede do-tšes eperan joks a eb-edž-eberan erde bireza sinora o dursuni*  
 ‘How far did you go, did you go over and back or from opposite until Dursun’s garden line?’ (translation partly unclear; 04\_01072019F\_17; 33–34)

Adversative coordination is marked by the Turkish loan *ama* ‘but’ in inter-clause position (298), but the conjunction can again also appear clause-initially (299). If the subject or direct object of both clauses are identical and mentioned in the first clause (or otherwise recoverable from discourse), it can be left out in the second clause (300; but see the use of an anaphoric weak object pronoun in 294 above). Note that in (301a/b) in response to a translation task to elicit the adversative conjunction, both clauses are juxtaposed without conjunction.

- (298) *makra ise ama o kadar eyabo=se*  
 ‘You are far, but I love you so much.’ (01\_14012015F\_1; 3–4)
- (299) *ama tem anθrobo ba sin almanjan dune dže inedžikse*  
 ‘But my husband was also in Germany and married [there].’ (01\_04022016F\_1; 116–117)
- (300) *ayabo na troyo tiba ama utš exo*  
 ‘I want to eat something, but I don’t have [anything].’ (C1)
- (301) a. *epiya so gurbeti bašlaepsa so tšelišema apetsenaen ensorane=m asi đulia* (C1)  
 b. *so šeher do đulema bašlaepsa mono išdera eksankane=me* (B1)  
 ‘I started to work at the city but soon after they fired me.’

Correlative conjunctions borrowed from Turkish follow the Turkish syntactical patterns: *ja ... ja* ‘either ... or’ (302), *ne ... ne* ‘neither ... nor’ (303), and *hem ... hem* ‘as well as’ (304) appear paired whereby one of them occurs in initial position of each clause (see Göksel & Kerslake 2005: 121). For *ne ... ne* and *hem ... hem*, no data are available whether also NPs alone can be conjoined like it is the case in Turkish. Note that in ex. (305a), the inherited topicalizer *ba(l)* is used to emphasize both adjectives within a single clause where the correlated NPs are otherwise just juxtaposed without conjunction. The speaker continues the clause using the Turkish conjunctions, though (305b).

- (302) *ja dajaxi n=etroyame ja..*  
 ‘Either we got the bashing or..’ (04\_01072019F\_2; 291)
- (303) *ne faje borum tše droyum ne rahatje borum tš=eftjeme*  
 ‘Neither can I eat bread nor can we make it easily.’ (07\_04072019F\_6; 58–59)

- (304) *hem d=apiðe ba fa hem do dajaye fa*  
 ‘Eat the pears as well as the bashing!’ (04\_01072019F\_2; 296–297)

- (305) a. *trana b=obsaræ eplezame bola ba*  
 ‘We caught as well big as many fish.’ (08\_04072019M\_1; 159)  
 b. *hem trana sani ham bola*  
 ‘They were as well big as many.’ (08\_04072019M\_1; 160)

### 5.3.1.2 Verb coordination (serial verbs)

A striking feature in Romeyka are different forms of VP coordination: (i) by mere juxtaposition of VPs; (ii) by coordination of different VPs by the conjunction *tše* (see Section 3.2.6.1.1). (i) Juxtaposition of VPs is a salient feature in Romeyka. It differs from serial verb constructions (see below) in lacking the coordinating conjunction *tše*. In many instances, this is just a strategy for expressing temporally tightly linked sequences of events carried out by the same actor, and most cases involve shared subjects. The events may be conceptually distinct, so that it is probably not appropriate to refer to this as verb serialization. Rather, it appears to be a strategy for clause combining. However, the borderline to some kind of verb serialization may not be easy to draw (see below on verb serialization). A minimum of two (up to four or five) VPs with the same subject are juxtaposed under OV word order (306, 307). The verbs share the same person/number inflections and the same tense values. Interestingly, this juxtaposition of VPs is not attested with negated VPs; most likely it is not possible to have positive and negative verbs together which indicates the close syntactic correlation of the verbs. Direct objects are often not repeated beyond the first verb (308). If so, the verb chains seem semantically to be related to an enumeration; subsequent verbs have limited grammatical features (like argument government) and seem to be reduced to their semantic meaning (see 307 and 308). However, this is not the case for all verb chains: in exs. (309, 310), VPs including locational adjuncts are juxtaposed. Note that the locational adjuncts are post-posed in both cases. In (311), the direct objects are repeated for all subsequent verbs (but note the rare insertion of ‘go’ between verb and object of the second VP). Although the semantic function of these verb chains is not fully clear, they seem to correspond (at least in the case of two juxtaposed verbforms; 312a) to the Turkish subordinating suffix *-(y)Ip* (312b), which is used to conjoin clauses of equal semantic status (Göksel & Kerslake 2005: 439).

- (306) *ados ba gedžejle na bai goft ksila*  
 ‘He will go chop wood in the night.’ (04\_01072019F\_13; 10–11)
- (307) *n=epina ðaðia epiya epejna fortunune nasi*  
 ‘I made ?, I went, I did, how I carried’. (04\_01072019F\_2; 142–143)
- (308) *ondan sonra ta za eyvalo [pause] payo epero erxome ðeno*  
 ‘Then I take out the cows, I go, take, come and bind.’ (04\_01072019F\_1; 007–008)
- (309) *ebijen dotše sin borda eseven apes*  
 ‘She went, knocked on the door, and entered inside.’ (04\_01072019F\_13; 33–34)
- (310) *ta is ebidže erθen eseven apes*  
 ‘He made footprints and entered inside.’ (04\_01072019F\_12; 24)

- (311) *so xor ha pies kadi pao eftayo kadi fidevo kadi tsakono*  
 ‘When you went to the field, I go [and] make something, I plant something, I break something.’ (03\_30062019F\_7; 07)

- (312) a. *epejna perena da xuljera se do tasi*  
 b. *gidip alıcaktım kaşıkları ve tası*  
 ‘I went and took the spoons and the cup.’ (07\_04072019F\_6; 42)

(ii) Serial verb constructions are generally a special form of VP coordination showing a high degree of grammatical integration of the verbs involved, as well as a closer semantic integration of the events. They consist of two or more verbs which are not compounded and form part of the same clause (Payne 1997: 307). Unlike in verb coordination, the individual verbs cannot have different TAM marking (Payne 1997: 308). Serial verb constructions (SVC’s) in Romeyka (see also Section 3.2.6.1) consist of two or more finite verbs which are all fully congruent in person, tense and mood marking (for an example with imperatives see 319 below). While it is usually the last verb in the chain that is coordinated by the conjunction *tše* (but see exs. 315 and 320 below), all other verbs are merely juxtaposed. Intonation of the sentence is like that of a single clause. Serial verb constructions often have a verb-final order as many of the coordinated verbs do not govern any arguments (but see 318 and 322 below for post-posed nominal direct objects; interestingly, no weak object pronouns are attested in serial verb constructions but the verb coordination with *tše* to express progressive aspect is attested with a weak object pronoun on the first verb, see Section 3.2.6.1.1). Serial verbs are negated by means of a single clausal negator that has scope over the whole clause (313). Interestingly, according to Drettas (1997: 397), serial verbs coordinated by *tše* only occur in positive assertions; in negative contexts, two verbs can be juxtaposed without conjunction.

- (313) *tipu tš=eperenam tš=edroyame uli hendan*  
 ‘We could not buy something to eat.’ (04\_01072019F\_2; 250)

It is not clear whether serial verb constructions have a particular semantics that differs from that if all VPs were coordinated. In typological literature, serial verb constructions describe often different aspects of one complex event (Payne 1997: 307). This could also pertain to Romeyka. However, certain verbs are more likely to occur in a serial verb construction, among them *fero* ‘bring’ (314), *eftayo* ‘do’ (315), and motion verbs like *erxume* ‘come’ (316) and *pao* ‘go’ (317), but also other verbs occur like *troyo* ‘eat’ (318).<sup>226</sup> Drettas (1997: 397) notes for PG, that mainly motion verbs can occur as first verb and the verb *pero* ‘take’, which evokes inchoative aktionsart.

- (314) *ta xordare=mun do thelis bas koftjis tše feris*  
 ‘Our fields: what you like you go, cut and bring it.’ (08\_04072019M\_1; 219)

- (315) *landževame landževame tš=epinam epesame*  
 ‘We jumped, jumped and made, we played.’ (04\_01072019F\_2; 019–020)

- (316) *manaxe ebenam tš=erxumis*  
 ‘We went and returned alone.’ (08\_04072019M\_1; 018)

- (317) *hajese efikame tš=epiyame*  
 ‘We left it like this (and we went).’ (04\_01072019F\_2; 323)

<sup>226</sup> The second *tše* in ex. (318) is probably the ‘additive particle’ *tše* (see Section 3.2.6.1.1).

- (318) *osa da šindže as es.. asirame tš=etroyame tše o koliva*  
 ‘If there was [any] now we could pull and eat boiled corn.’ (04\_01072019F\_2; 273)

While the combination of verbs like *troyo* ‘eat’ or *fero* ‘bring’ yields a complex meaning of the described event, the semantic meaning of *eftayo* ‘do’ and motion verbs seems reduced. Especially verb serialization with *erxume* ‘come’ seems to lack the original semantics of the verb (319, 320). This semantic bleaching may possibly indicate the begin of a grammaticalization process, although the direction of such a development is not clear (yet). According to Payne (1997: 310), however, serial verb constructions carry from a typological point of view often an aspectual meaning that is especially expressed based on verbs of motion (Drettas 1997: 396 confirms this for PG but it seems in various respects that the serial verb construction found in Romeyka at present has further grammaticalized than that presented by Drettas for PG).

- (319) *ayome dere tše erse*  
 ‘Go, look(, and come)!’ (08\_04072019M\_3; 043–044)
- (320) *da xordaræ kofteme tše eberum tš=erxumis*  
 ‘We cut the meadows, we take it and return.’ (08\_04072019M\_1; 200–201)

Serial verb constructions in Romeyka have a parallel with Turkish *-(y)Ip* converbs of the type verb stem + *-(y)Ip* + *finite verb* which usually target the same kinds of verbs (although in Turkish they can occur with all verbs), i.e., frequently a combination of ‘take’ or ‘bring’ with a motion verb (321, 322). Turkish *-(y)Ip* functions as subordinating conjunct which conjoins two VPs with the same semantic status in terms of TAM (Göksel & Kerslake 2005: 439–440). The two verbs are phonologically as well as syntactically conjoined to a close semantic meaning of factually a single event which might be expressed in other languages by a single lexeme (exs. 321, 322; for another example, cf. also Tursun 2019: 217). The converb itself does not show any inflectional marking; usually, only the second (finite) verb is negated, although the converb can be principally negated as well.

- (321) a. *am pao so raši n=epero liyo diri tše na fero* (C1)  
 b. *Dağlara gittiğim zaman, biraz peynir alıp götürüyorum.*  
 ‘When going to the mountains, I take a little cheese with me.’
- (322) a. *pao so raši ebero tše bayo ena demligi*  
 b. *bir tane demlik alıp dağlara gidiyorum*  
 ‘I go to the mountain and take a tea pot with me.’ (08\_04072019M\_1; 264–265)

Finally, the serial verb construction with motion verbs might have influenced in turn the regional Turkish variety (323), leading to mere juxtaposition of verbs.<sup>227</sup>

- (323) *hep rumdža konusurduk giderduk*  
 ‘We always spoke Rumca (and went).’ (04\_01072019F\_2; 003)

<sup>227</sup> Note that the author is only aware of this single example in the Romeyka corpus. So, further research on the serial verb construction in regional Turkish and the actual origin of it (i.e., which language exerts an influence on which) is required.

### 5.3.2 Complementation

A complement clause (CC) is a clause embedded into another clause which functions as an argument to the main verb of the matrix clause (Payne 1997: 313). In Romeyka, there are four main strategies for complementation depending on the type of predicate, two finite and two non-finite: (i) finite zero-complementation (i.e., clausal complements that lack an overt complementizer) for perception verbs, epistemic predicates, some emotive verbs, and verbs of saying (Section 5.3.2.1.1); (ii) finite CCs with the complementizer *na* for potentials, volitionals, mental perception verbs, some emotive verbs, certain causatives, and other predicates like ‘wait’ (Section 5.3.2.1.3). In addition, the areal complementizer *ki* is borrowed in verbs of saying (Section 5.3.2.1.2). In potentials, as a very rare trait, the conjunction *tše* ‘and’ is found occasionally as complementizer instead of *na* (Section 5.3.2.1.3); (iii) infinitives (among others) with negated past tense potentials and volitionals (Section 5.3.2.2.1);<sup>228</sup> (iv) non-finite deverbal nouns with volitionals and aspectual predicates like ‘start’ (Section 5.3.2.2.2). Importantly for some complement types, more than one complementation strategy is available, such as voluntatives and mental perception predicates such as *enespala* ‘I forget’ that can be complemented by a verb in imperfective, *na*-clauses, nominalized infinitives, or deverbal nouns. While Romeyka complement clauses are mostly finite, and the non-finite strategy of using infinitives is an archaic trait, non-finite deverbal nouns as a strategy have increased under contact with Turkish. Within the finite complementation strategies, complementation by means of the complementizer *na* is more restricted in Romeyka compared to SMG, while juxtaposition with paratactic syntax (which is also found in Cappadocian and other Greek dialects, Sitaridou 2011) without complementizer is widespread, for example with utterance verbs and in (in)direct speech (Section 5.3.2.1.2; cf. also Neocleous 2020: 85).

The structural integration of finite complement clauses is rather weak (with “balanced” verb forms in the CC, in terms of Christofaro 2013a). Finite CCs can have their own subject and individual TAM marking, while non-finite CCs have an infinite verb form neither displaying person or tense marking (“deranking” verb forms in terms of Christofaro 2013a) and their subject is often coreferential with the matrix clause. The position of CCs in the whole sentence is mostly post-posed, which is a common feature of VO languages (Payne 1997: 314), although indirect questions and (in)direct speech allow for much for variability.

SMG has balanced utterance complements while Turkish has both, balanced and deranked. Since deranked forms are typologically less frequent, according to Christofaro (2013b), an implicational hierarchy (the “Subordination Deranking Hierarchy”) suggests that languages that use deranked verb forms in utterance complements will likely also use deranked verb forms in all other complement types, as well as for most types of relative and adverbial clauses (see Christofaro 2013b). In Romeyka, finite complements are widespread in subordination and complementation, while the impact of Turkish deranked forms (e.g., deverbal nouns) is growing.

#### 5.3.2.1 Finite complementation

##### 5.3.2.1.1 Clausal complements lacking an overt complementiser<sup>229</sup>

Several types of complement clause occur without an overt complementiser; while null complementizers are scarce in SMG (Sitaridou 2014a: 127), null complementation is otherwise common in varieties of AMG. Sitaridou (2014a: 127) describes the zero-strategy as “extremely productive” in Romeyka. Clause types without complementizer include the complements of the

<sup>228</sup> But note the existence of inflected, personal or nominalized infinitives (cf. Section 4.3.7, also 5.3.2.2.1), which are strictly speaking not infinite anymore.

<sup>229</sup> For the sake of clarity, all complement clauses are put in square brackets throughout this Section 5.3.2.1.1 and Section 5.3.2.1.2.

following semantic verb classes: (i) perception verbs like ‘see’ (324) and ‘hear’ (325; see also Sitaridou 2014a: 127)); (ii) epistemic predicates like *θaro* ‘think/believe’ (326) and ‘know’, the latter especially in indirect questions with ‘whether’ or *wh*-words (Sitaridou 2014a: 128; see same section below); (iii) some emotive/psych verbs such as ‘fear’ (327, but cf. 328) (but not in all emotive verbs; others are expressed by *na*-clauses such as *xarenume* ‘be happy’, Sitaridou 2014a: 127); (iv) verbs of saying (Sitaridou 2014a: 128). Verbs of saying are predominantly zero-marked, but since other strategies exist with them as well and their differentiation from direct speech is difficult, they are treated in a separate section (5.3.2.1.2) below.

- (324) *eterezen [o argo erθen]*  
 ‘He saw that the bear came.’ (05\_03072019M\_4; 24)
- (325) *eguθame [o jiajia evren arkon]*  
 ‘We heard that Yahya has found a bear.’ (02\_2906019F\_1; 02)
- (326) *eθaresa [ab=emetera so xorio isen]*  
 ‘I thought that you are from our village.’ (01\_15022015F\_1; 10)
- (327) *etsakonam tš=etroyame ama laya foyumunesine [kanis me lepe=masine]*  
 ‘We cracked them [=walnuts] and ate them, but how we were afraid, that nobody should see us.’ (04\_01072019F\_2; 288)
- (328) *da insane bal foyundan [na=xundane s=emasu]*  
 ‘The people are afraid of coming to us.’ (02\_29062019F\_2; 37–38)

Complement clauses with zero-marking of object complements are post-posed by simple juxtaposition. From a typological view, complement clauses of VO languages are likely to be post-posed (Payne 1997: 314). Word order within the CC is verb-final with predicatives (329) and predicative possession (330), and also with some other complements (e.g., 324) but SVO order exists as well (325). Note that according to Neocleous (2020), the unmarked word order in subordinate clauses is SOV. The CC is less closely structurally integrated and can appear in its own tense (see, e.g., ex. 326 where the matrix clause is in the aorist and the CC in present tense). There appear to be no constraints on the occurrence of null complements in a particular tense, e.g., aorist past tense.

- (329) *Eyriko [aðakes pola kleftes ine].*  
 believe.1SG here very thieves are.3PL  
 ‘I believe that the people here are real thieves.’ (ROf; Sitaridou 2014a: 128)
- (330) *oh so spidin iða [televisjonin ixame]*  
 ‘Oh, at the house, I saw that we have a TV!’ (02\_21042018M\_2; 52)

Indirect questions are often complementing epistemic predicates or verbs of cognition like *nunizo* ‘think’, *ksero* ‘know’, *eyriko* ‘understand’ but also perception verbs like *tero* ‘see’ or verbs of saying. Indirect polar questions lead to ‘whether’-complements, while content questions as complements include *wh*-elements (see also Section 5.2.3.3.5 on embedded questions). Both types of indirect question complements are zero-marked.

‘Whether’-clause clauses evoking a polar answer are formed in principle like direct polar questions and using the Turkish interrogative particle *mi* (331, 332)<sup>230</sup>. The complement clause

<sup>230</sup> Note that in ex. (332), *na* does not occur as a complementizer but in its function as future particle.

follows the matrix clause. The word order of the interrogative complement clause seems to coincide with that of direct polar questions (Section 5.2.3.3.1), i.e., SVO/SVX with the interrogative particle in final position, although Neocleous (2020) attests SOV order in indirect questions and subordination in general.

- (331) *ebedženaen as dirume [argos erden edži mi]*  
 ‘Then, let us see whether a bear will come.’ (04\_01072019F\_12; 41)

- (332) *as derume [na krui etunesine mi]*  
 ‘Let us see whether he will hit him.’ (04\_01072019F\_12; 15–16)

Indirect content questions including a *wh*-element are also juxtaposed as complements without any complementizer, so it is sometimes not clear how to distinguish them from direct speech (333, 334), especially, as there is no grammatical integration of the CC which can have its own TAM features.<sup>231</sup> The structure of the complement clause seems to follow in general the word order of direct content questions (Section 5.2.3.3.4) with OV especially in predicatives but with SVO possible in the complement clause (1; unlike Neocleous 2020: 90). Stacking of CCs seems possible (333). Often, the CC follows the matrix clause (333, 334), although it may also precede it (335). Circumposition is possible in marked environments (336).

- (333) *nunizune [pu ine to psare=m do ejendane t=opsari=m do ebaθane]*  
 ‘They think, “Where are my fish, what happened to my fish, what happened [to them]?”’  
 (05\_03072019M\_3; 44–46)

- (334) *adin ba nunizune leyone [to.. harin nde n=eftame]*  
 ‘They think, they say, “What will we do now?”’ (05\_03072019M\_4; 05)

- (335) *[to leyo] utš eyrigo*  
 ‘She<sup>9</sup> does not understand what I say.’ (01\_28062019F\_3; 24)

- (336) *[t=aleyo] kseris [dohna e]*  
 ‘Do you know what “aleyo” is?’ (08\_04072019M\_2; 066)

#### 5.3.2.1.2 Clausal complements of verbs of saying

This section deals with different complementation strategies of verbs of saying as well as with the continuum between direct and indirect speech.

Complement clauses introduced by utterance predicates such as ‘say’ are embedded clauses with zero-complementation (ex. 337, 338; Sitaridou 2014a: 128; see also Sections 5.2.3.3.5 and 5.3.2.1.1 for indirect (embedded) questions). In the speech of Turkish-dominant bilinguals, the areal complementizer *ki*, also common in Turkish, can occur (339, 340), although these examples should be rather analysed as direct speech as becomes evident from the person marking of the pronouns (for a discussion of *ki* in the “speech act phrase” of Pharasiot Greek, see Bağriacık 2018: 289–440). Utterance complements have a finite verb form in the dependent clause. The dependent clause is either post-posed, which is the unmarked order (337), center-

<sup>231</sup> Note also the existence of indirect content questions which are apparently subordinated to the non-transparent non-verbal<sup>9</sup> form *eyunde* ‘what do I know/I don’t know’, although an analysis as parataxis would be possible as well (iv, v).

(iv) *eyu-nde ndoxnan n=anlatjevo=se*  
 ‘I don’t know, what to tell you.’ or ‘..., what shall I tell you?’ (07\_04072019F\_6; 33)

(v) *eyu-nde odene n=efteme*  
 ‘I don’t know what we do.’ or ‘..., we do everything.’ (04\_01072019F\_1; 036)



embedded (341, 342),<sup>232</sup> or circum-posed (343, 344). In Neocleous (2020: 73), also a pre-posed dependent clause is attested (345). CCs introduced by *ki*, are always post-posed like in Turkish.

- (337) *emine iben [sabale n=arde]*  
 ‘Emine said that she will come tomorrow.’ (04\_01072019F\_5; 22)
- (338) *la=na [so xorafi=m]*  
 ‘I told him [that I was] at the field.’ (03\_30062019F\_7; 13)
- (339) *eyo ton džiri=m ipa ki [bola eyabo=se]*  
 I the.ACC father.ACC=POSS.1SG told.1SG COMP much love.1SG=you  
 ‘I told my father that I love him much.’ (A1)
- (340) *ama ade lej ki ena emena son [...]*  
 ‘But she says, one is enough for me [...].’ (01\_04022016F\_1; 096)
- (341) *as eřtayo=se faji ama [utše na troyo] ipes=me*  
 ‘I had prepared lunch for you, but you said to me that you won’t eat it.’ (A1)
- (342) *O Alis esenan yrapson ipen*  
 the Alis.NOM you write.IMP.2SG tell.AOR.3SG  
 ‘Alis told/ordered you to write.’ (ROf; Sitaridou 2014a: 129, ex. 41; glosses modified)
- (343) *[permeno n=arxundane] ibe [to gülfe]*  
 ‘She said: “I wait for the plague to come.”’ (01\_04022016F\_1; 125–126)
- (344) *ište [n=epinam] ipen [ta za]*  
 ‘“In fact, we should do”, she said, “the animals.”’ (04\_01072019F\_2; 137)
- (345) *[Papos=mu ađatřaka stetř] lei.*  
 grandfather.NOM=POSS.1SG here live.3SG say.3SG  
 ‘S/he says that my grandfather lives here.’ (Neocleous 2020: 73, ex. 92; glosses modified)

Circum-posed CCs around the utterance verb ‘say’ are a salient feature in Romeyka, but it is not clear whether they can be fully explained by pragmatic causes such as topicalization (346) and the flexible position of locational adjuncts (347).

- (346) *[jad ado] iba [erřa s ađaha katsa ga]*  
 ‘“Therefore”, I said, “I come here and sit down.”’ (03\_30062019F\_6; 34)
- (347) *[sabale] ibe [n=arxume]*  
 ‘“Tomorrow”, she said, “I will come”.’ (07\_04072019F\_5; 30)

Due to their finiteness and lack of syntactic dependency on the matrix clause, all utterance complements presented so far could be interpreted as direct speech. So, by making use of zero-complemented CCs, Romeyka has no means to distinguish between direct and indirect speech and it remains unclear whether there are syntactical means to differentiate the two (cf. Christofaro 2013b, and references therein). According to Christofaro (2013b, citing Munro

<sup>232</sup> It seems Turkish influence on this construction is strong resulting in OV order in the matrix clause; while in SMG, the predicate would precede the CC.

1982), “[a] direct report differs from an indirect report in a number of structural and functional ways, and is not an argument of the utterance verb introducing it”. In this sense, the complementation strategy reported by Özkan (2013) in (348), with the interrogative element *ndo* ‘what’ used as complementizer (‘that’) is remarkable in that it can only be analysed as reported speech. In a translation task from Turkish, speakers from the present corpus provided some examples of indirect speech introduced by complementizers, namely *ndo* (349a/b), *ot* (349c) or with a *na*-clause in the dependent clause (349d), a strategy probably influenced by SMG, along zero-complementation (349e). Note that clitic doubling occurs in some examples (349a/d/e).

- (348) [Eyo nt=eyapesa=se] son kosmo ipa.  
I that=loved.1SG=you to.the.ACC world.ACC told.1SG  
‘I have told the world that I loved you.’ (RSür; Özkan, n.d.; glosses added)

- (349) a. ton džiri=m [**ndo** eyabo ane] iba=do (B1)  
b. o džiri=m [**d**=eyape sen] ipe=me. (H3)  
c. don džiri=m [eh.. **ot** eyabo etšinona] iba (C1)  
d. ton džiri=m [eyo dži **n=erapesa**] ipa to (H1)  
e. don džiri=m iba [eyo eyabo=ya] (H2)  
‘I told my father that I love him.’

Furthermore, while many utterance complements suggest an analysis as direct speech rather than reported speech (see below), embedded imperatives as complements to verbs of saying (see Sitaridou 2014a: 129–130) seem to suggest more a reading as indirect speech, probably due to the reduced set of morphological features available for imperatives (350–352). The order of the subordinate imperative and the main verb is variable (350–352 vs. 342 above). Furthermore, both OV (350) and VO (353) orders within the CC are possible.

- (350) eyo lata atšin [da tšitšege xasun=a]  
‘I told her to remove the flowers.’ (01\_28062019F\_3; 03)
- (351) i inega=nat ipen adona [mi bas sin avlea]  
‘His wife told him, “Don’t go outside!”’ or ‘... not to go outside’ (04\_01072019F\_13; 13)
- (352) iben adona mi.. m=eftjes=ada hajes  
‘She told him, “Don’t do these like this!”’ or ‘... not to do these like this’ (04\_01072019F\_13; 21–22)
- (353) Tin patsi=m ipa pison xapsia na troyume.  
the daughter=POSS.1SG tell.AOR.1SG make.IMP.2SG whitebait.ACC PRT eat.2PL  
‘I told my daughter to fry whitebait to eat.’ (Sitaridou 2014a: 129, ex. 39b; glosses modified)

In order to provide an attempt at an explanation based on the examples of the present corpus, the following preliminary observations can be drawn: If we assume a continuum between reported and direct speech, the reading of utterance complements as direct speech is (in a range from pragmatic aspects to structural) especially likely if (i) the CC includes an address noun (354, 355); (ii) the CC is repeated (356); (iii) the CC contains an answer particle (357) or elliptical (answer) clause (358, 359); (iv) the subject or object of the embedded clause remains unaltered in relation to the speaker (co-referentiality), e.g., 2<sup>nd</sup> person verb forms or pronouns

instead of 1<sup>st</sup> or 3<sup>rd</sup> person reported speech (360–362); (v) the CC has its own TAM marking such as optatives (363, 364), but not the subjunctive, though, which facilitates a ‘subordinate reading’ (365).

- (354) *ti mana=muna eleyame [ema epinasame]*  
 ‘We said to our mother: “Mom, we are hungry!”’ (02\_02022015F\_1; 114)
- (355) *iben [mana daha ndo kahese]*  
 ‘He said, “Mother, why are you sitting here?”’ (03\_30062019F\_6; 30)
- (356) *ama ade lej ki [ena emena son ena emena son]*  
 ‘But she says, “One is sufficient for me, one is enough.”’ (01\_04022016F\_1; 096–097)
- (357) *iba joox esi hado katuresis*  
 ‘I said, “No, you have peed this!”’ (02\_29062019F\_2; 29)
- (358) *ipa [na m=eftayo šamada apes]*  
 ‘I said, “In order to no make noise inside”.’ (03\_30062019F\_6; 31)
- (359) *[hare i almanja] ibe epijen*  
 ““Now Germany!”, she said and went.’ (01\_04022016F\_1; 116)
- (360) *eraepsane=m ipane [omorfo duvari eftes]*  
 ‘They called me and told me, “You make good walls.”’ (08\_04072019M\_3; 072)
- (361) *leya=na [esena u=düşünefkume]*  
 ‘I told her, “I am not thinking of you.”’ (01\_04022016F\_1; 097–098)
- (362) *ibane [na rezume na krume ton argo]*  
 ‘They said, “We will wait and shoot the bear.”’ (05\_03072019M\_4; 06–07)
- (363) *lege [as dero ine mi dibo na troyume]*  
 ‘He said, let me see whether there is something to eat.’ (04\_01072019F\_12; 11)
- (364) *iba [as pao ber aso xoraf]*  
 ‘I said, let me go and bring (some) from the field.’ (03\_30062019F\_11; 036)
- (365) *öretmenis elej=machine [türki na lede]*  
 ‘The teacher told us to sing folksongs.’ (04\_01072019F\_2; 040–041)

### 5.3.2.1.3 *na*-Clauses

In comparison to SMG, the use of *na*-clauses is in Romeyka more restricted (Sitaridou 2014a). *Na*-clauses are used for complementation (for other modal functions of *na* see Section 3.2.4.4) of (i) (negated present tense) modals like the potential with *boro* ‘can’ (366; Section 4.3.6.5); (ii) volitionals with *thelo* ‘want’ and *eyabo* ‘like/love’ (367; Section 4.3.6.4); (iii) complements of *permeno* ‘wait’ (368), also *erazo* ‘keep watch’ (369); (iv) some emotive verbs like *exara* ‘be happy’ (370) but apparently not *fovume* ‘fear’ (Sitaridou 2014a: 126–127), but cf. (371) and Drettas (1997: 324, ex. 55) for *na*-clauses with *fovume*; (v) verbs of mental perception like *anespalo* ‘forget’ (but a strategy with deverbal nouns exists as well, Sitaridou 2014a: 125; see 5.3.2.2.2); and (vi) ‘causatives’ with *afino* ‘leave’ (372; Sitaridou 2014a: 125). Complement clauses introduced by the particle *na* are finite and can have their own person, number (e.g., ex.

368) and tense marking (e.g., ex. 373) apart from co-referentiality with the subject of the matrix clause, which is expressed by the person-number affix on the finite verb (cf. Haspelmath 2013b); in the majority of cases, person, number and tense of both verbs are congruent, though. *Na*-clause complements appear always post-posed.

- (366) *u=borume alo n=eskumes*  
 ‘We cannot get up anymore.’ (02\_02022015F\_1; 162)
- (367) *utš eθelena n=andriza*  
 ‘I didn’t want to marry.’ (02\_02022015F\_1; 014)
- (368) *permeno na=rxundane ibe to gülfe*  
 ‘She said, “I wait for the plague to come.”’ (01\_04022016F\_1; 125–126)
- (369) *erazune na skodonone t=argures*  
 ‘They keep watch to kill the bears.’ (05\_03072019M\_3; 23–24)
- (370) *Exara na mairevo.*  
 be-happy.AOR.1SG PRT cook.1SG  
 ‘I was happy I had cooked.’ (ROf; Sitaridou 2014a: 126, ex. 23; glosses modified)
- (371) *ta insane bal foyundan n=erxundane s=emasuna*  
 ‘The people are afraid of coming to us.’ (02\_902019F\_2; 38–39)
- (372) *Efikane=sas na skaftete ta xorafæ=suna.*  
 let.AOR.3PL=OPN.2PL PRT dig.3PL the fields=POSS.3SG  
 ‘They let you dig his fields.’ (Sitaridou 2014a: 125, ex. 18a; glosses modified)
- (373) *Enespala na leyo ti mami ta xaberæ.*  
 forget.AOR.1SG PRT say.1SG the grandmother.ACC the news.ACC  
 ‘I forgot to tell the news to the grandmother.’ (Sitaridou 2014a: 125, ex. 20a)

Note that in a single environment, namely, with potentials with *boro* ‘can’, the connector *tše* ‘and’ was found to be able to take over the function of *na* (374), see Section 4.3.6.5.

- (374) *ti mana=m eporo tš=elepo*  
 ‘I can see my mother.’ (A1)

### 5.3.2.2 Non-finite complementation

#### 5.3.2.2.1 Infinitives

Infinitives have strictly speaking neither person nor tense marking and the subject of the CC is coreferential with that of the matrix clause. However, in Romeyka, the infinitive may have a distinct nominative subject, i.e., inflected and personal infinitives exist as well (Sitaridou 2014a: 130, 2014b; for the morphological forms of all types of infinitives see Section 4.3.7). According to Sitaridou (2014b), the infinitive is licensed in ROF in the following environments (see below), although Mackridge (1999: 102–103) remarks that the syntactic constructions in which the infinitive occurs vary between dialects; for example, RSür only displays inflected infinitives (Sitaridou 2014b: 49, Table 3). In the present Romeyka corpus, (uninflected) infinitives figure only rarely, so the information in this section are mainly based on Sitaridou (2014a) and

(2014b). Note that this section is a concise summary of a more detailed discussion of non-finite structures in the Romeyka corpus in Section 4.3.7.

Morphosyntactic environments which license infinitives in ROF following Sitaridou (2014b):

- (i) negated past tense modals/potentials with *boro* ‘can’ (375); but cf. the inflected infinitive in (376) from the present Romeyka corpus;
- (ii) negated past tense volitionals with *θelo* ‘want’ (377), but not *eyabo* ‘like/love’; but compare the participle-type verb form found in the Romeyka corpus in ex. (378), and a *na*-clause in (379);
- (iii) in counterfactuals such as wishes (380), exclamatives (381), and conditionals (382) as a complement of ‘have’ (Sitaridou 2014a: 136); cf. the inflected infinitive from the present corpus in (383), otherwise, ‘have’ as auxiliary does not figure in the present data;
- (iv) as hapax legomena after a verb of perception (384) and negated past motion verb (385);
- (v) in nominalized infinitives as complements to aspectual and mental perception predicates (see below);
- (vi) in ‘before’ clauses with *prin* (see Section 5.3.3.1.1).

(375) *Utš eporesa tšimiθini.*  
not can.AOR.1SG sleep.INF  
‘I could not sleep.’ (ROF; Sitaridou 2014a: 122; glosses modified)

(376) *utš eborea tšimeθina*  
‘I could not sleep.’ (B1, T1)

(377) *Utš eθelesa mairepsini.*  
not want.AOR.1SG cook.INF  
‘I didn’t want to cook.’ (Sitaridou 2014a: 39, ex. 17; glosses modified)

(378) *utš eθelene buθen (n=)²eβejene*  
‘She didn’t want to go somewhere.’ (01\_04022016F\_1; 105)

(379) *utš eθelena n=andriza*  
‘I did not want to marry.’ (02\_02022015F\_1; 014)

(380) *As išen porpatesini sa rašia!*  
PRT have.IPFV.3SG walk.INF to.the mountains  
‘S/He should have walked in the mountains.’ (Sitaridou 2014a: 136, ex. 72a; glosses modified)

(381) *Na ixame panini xtisini to spit so parxar!*  
PRT have.IPFV.1PL go.INF build.INF the house in.the pastures  
‘I wish we had gone to build the house in the highland pastures.’ (Sitaridou 2014a: 136, ex. 72b; glosses modified)

(382) *An ixa mairepsini, ixame fanini.*  
COND have.IPFV.1SG cook.INF have.IPFV.1PL eat.INF  
‘If I had cooked, we would have eaten.’ (Sitaridou 2014a: 136, ex. 73b; glosses modified)

(383) *[...] eyo na mi ixa škisenda da ksila [...]*  
‘[...] If I had not chopped the wood, [...].’ (04\_01072019F\_13; 53)

- (384) *Eyo tš=iða tus tšopanus almeksini ta xtinæ*  
 I.NOM not=saw.1SG the.ACC shepherds.ACC milk.INF the.ACC animals.ACC  
 ‘I did not see the shepherds milk the animals.’ (Sitaridou 2014b: 40, ex. 22; glosses modified)

- (385) *tš=epies almeksin=ata*  
 not=went.2SG milk.INF=OPN.3PL  
 ‘You didn’t go to milk them.’ (Sitaridou 2014b: 42, ex. 32; glosses modified)

Infinitives are always post-posed after the predicate of the matrix clause. However, in (378) an adverbial element can be inserted between the finite and infinite verb (cf. Sitaridou 2014a: 130, ex. 46 for clause-final adverb placement) and the utterance predicate in (384) allows an object NP in intermediate position. According to (381), infinitival complements allow stacking (see also Sitaridou 2014a: 130, ex. 45a), so the infinitive itself can take either a NP or a predicate as a complement. Furthermore, the infinitive can be coordinated (Sitaridou 2014a: 130).

Nominalized infinitives are complements to (i) past tense aspectuals such as *epiturepsa* ‘I finished’ (386) also *epaşlaepsa* ‘I started’ (see also Section 5.3.2.2.2), and (ii) verbs of mental perception such as *enespala* ‘I forgot’ (387a), alongside *na*-clauses (Sitaridou 2014a) but cf. the deverbal noun strategy in (388) from the Romeyka corpus. According to Sitaridou (2014a: 130), the nominalized infinitive is used with a, possibly obligatory, complex possessive *=(e)muneθe* (for attempts at an explanation of the internal structure, see Section 4.3.7).

Furthermore, nominalization by means of inflected infinitives is attested before predicative adjectives such as in (389) (Sitaridou 2014b: 42); cf. the nominalized infinitive in (390) from the Romeyka corpus. Note that while nominalized infinitives with *=(e)muneθe* in context (i) and (ii) above only occur in past tense, in predicative adjectives, nominalized infinitives are attested in present tense. As a nominalization, the infinitival form carries the neuter article and occurs immediately before the licensing predicate. This structure is clearly a calque of the Turkish right-branching structure (cf. 387a/b), although Sitaridou (2014a: 41) highlights the existence of similar movement mechanisms in Hellenistic times.

- (386) *To tšimithin=emun-eθe epiturepsa.*  
 the sleep.INF=our.its finish.AOR.1SG  
 ‘I finished sleeping (=I woke up).’ (Sitaridou 2014a: 130, ex. 48a; glosses modified)

- (387) a. *To tšimiθin=emun-eθe enespala.*  
 the sleep.INF=our-its forget.AOR.1SG  
 ‘I forgot to sleep.’ (ROf; Sitaridou 2014a: 131, glosses modified)  
 b. *uyuma-yi unuttum*  
 ‘I forgot to sleep.’

- (388) *tin borda do džirin do kliðama enespal*  
 ‘The father’ forgot to lock the door.’ (B1)

- (389) *Afti i ðulian to maθini=mu yola en*  
 this.NOM the.NOM job.NOM the learning=my easy be.3SG  
 ‘It is easy for me to learn how to do this job.’ (Sitaridou 2014b: 42, glosses modified)

- (390) *ena aletero ylossa maθama.. maθenimo bola zor ie*  
 ‘Learning another language is difficult.’ (C1)

## 5.3.2.2.2 Deverbal nouns

Romeyka employs two strategies of nominalization: nominalized infinitives (see Section 5.3.2.2.1 just above) and deverbal nouns (Sitaridou 2014a: 130). While nominalized infinitives are formed by the aorist stem + infinitival ending *-ini/-eni* + nominalization ending *-mo*, e.g., *to piθenimo* ‘the doing’ (see ex. 5a/b below),<sup>233</sup> deverbal nouns are also formed by the aorist stem + nominalization ending *-mo(n)/-ma*, e.g., *to pisimo* ‘the doing’ (cf. also Section 4.1.1.1).<sup>234</sup> Both nominalizations require the neuter article. In deverbal nouns, dependent upon the predicate, a plain nominalized NP or a PP with the locative preposition *s* ‘at, in’ are selected: complementation by a PP is obligatory with aspectuals such as *bašlaevo* ‘start’ but also with *apoliyo* ‘send’, and *eyriko* ‘understand’ (in 396b in the meaning of ‘explain’); but not with volitionals and mental perception verbs such as *anespalo* ‘forget’. Deverbal nouns are selected with several predicates where also other complementation strategies are possible; they seem to be productive and spread also to predicates with traditionally other forms of complements. The spread of non-finite deverbal complements (including the nominalized infinitive) is likely contact-induced by Turkish nominalizations in CCs based on short infinitives in *-mA*, e.g., *çalış-ma* ‘the working’ (cf. with 398a/b below), although Sitaridou (2014a: 52) also notes existing strategies of nominalization in HelGr and MedGr that might play a role as well and might be reinforced by contact with Turkish.

Deverbal nouns are infinite without any TAM marking of their own (391a), but they can have their own subject arguments (392, see also 263 in Chp. 4). Deverbal nouns occur frequently in immediate pre-verbal position (except for complements of *pašlaevo* ‘begin’, see below) which is clearly a calque of the Turkish structure (see 391b). Locational adjuncts can apparently be placed after the nominalization, though (see 392).

- (391) a. *sa rašia to borbatema θelo*  
       b. *dağlarda dolaşmayı seviyorum*  
       ‘I like to walk in the mountains.’ (B1)

- (392) *To peði=m to panimon aso xorion θelo.*  
       the child=POSS.1SG the going from.the village want.1SG  
       ‘I want my child to leave the village.’ (ROf; Sitaridou 2014a: 131, glosses modified)

Deverbal nouns are licensed (i) by volitionals with *θelo* ‘I want’ (xa) and *eyabo* ‘like/love’ and occur also in those environments in which usually (a) *na*-clauses occur, namely with present tense negation (393a vs. 393b; see Section 4.3.6.4), and (b) where the plain infinitive occurs, namely with the negated past tense predicate of *θelo* (394a/b vs. 377 above). Note that instead of the nominalization strategy of deverbal nouns, nominalized infinitives occur in the same environment (395a/b, 396a/b, also 394b). In volitionals, deverbal nouns (or nominalized infinitives) with a meaning of ‘like/love’ seem to occur more frequent with the verb *eyabo*, while deverbal nouns with the modal aspect of ‘want’ tend to be more frequently expressed by *θelo*, e.g., *teso t=erθanimο θelo* ‘I want that you come’ (B1).

<sup>233</sup> Note that Sitaridou (2014a/b) glosses the internal composition of nominalized infinitives as *-ini* + *=mu*.POSS.1SG (for a discussion see Section 4.3.7). Tursun (2019) presents nominalized infinitives in the form *erθen-imo*, Papadopoulos (1955) as *erθæne-mo*.

<sup>234</sup> Note the rare form of the nominalized complement in (vi).

(vi) *sa rašia do dolaneftije bola eyabo*  
       ‘I like walking in the mountains.’ (C1)

- (393) a. *to panimo eyo utš eyabo*  
 ‘I don’t want to go.’ (08\_04072019M\_1; 284)  
 b. *u θelnume na bame*  
 ‘We don’t want to go.’ (01\_04022016F\_1; 141)
- (394) a. *sadedže jad emena faji do psesme utš eθelisa* (B1)  
 b. *sade jad emena faji utš eθelisa pseθinimo* (C1)  
 ‘I didn’t want to cook just for my own.’
- (395) a. *mets arkadašis=mu soxbeti.. n=efteyo.. to pisenimo eyabo* (C1)  
 b. *me di arkadašis=in to.. sohbet to piθimo θelo – parakaθ* (B1)  
 ‘I like chatting with my friends.’
- (396) a. *i aiše tin batsi=nates ebolise diri so piθenimo* (C1)  
 b. *i aiše tin batsi=nades diri so biθimo erlaeps* (B1)  
 ‘Ayşe taught her daughter how to make cheese.’ (in a. lit. ‘send’)

Furthermore, deverbal nouns are (ii) licensed by aspectual predicates like the phase modal *pašlaevo* ‘start’ (397, 398) which requires a deverbal noun introduced by the preposition *so* ‘to the’ (Sitaridou 2014a: 130).<sup>235</sup> But instead of a deverbal noun, a nominalized infinitive is also frequently selected 399, 400; see also Sitaridou 2014a: 45, ex. 40). The word order of complements of *pašlaevo* is frequently post-verbal, although pre-verbal CCs occur as well (e.g., 398a). Object arguments can appear in intermediate position between the verbs (see 400).

- (397) *ebašlaepse so borbatima*  
 ‘She started to walk.’ (04\_01072019F\_13; 30)
- (398) a. *so šeher do θulema bašlaepsa* (B1)  
 b. *bašlaepsa so tšelišema* (C1)  
 ‘I started to work (at the city).’
- (399) *jani gözüz na en tomara na bašlaeps so kseraθinimo adode-džes*  
 ‘When it will be autumn, we start the drying all together around that time.’  
 (03\_07072019F\_1; 44)
- (400) *tšebedži bašlejevum do dovari so piθenimo*  
 ‘Thereafter, we will start building the wall.’ (08\_04072019M\_3; 049)

(iii) Deverbal nouns can also occur with mental perception verbs like ‘forget’ (ex. 388 above) which are usually complemented by *na*-clauses but also occur with nominalized infinitives (ex. 387a above) and an imperfective verb form, e.g., *enespala etroya*.IPFV ‘I forgot to eat’ (Sitaridou 2011). A hapax legomenon with a deverbal noun complement also exists for ‘think’ (401).

- (401) *tin para do mustafa odiıntš to θoθimo düšüinefkume dže kahume*  
 ‘I think about lending Mustafa the money.’ (B1)

<sup>235</sup> Note the lack of the preposition in ex. (398a) which is possibly related to the pre-posed locative adjunct. Otherwise, since the aspectual *bašlaevo* ‘start’ usually requires a PP, one wonders whether contact from Turkish may also play a role in the lack of the preposition.



### 5.3.3 Adjunct clauses

This section is dedicated to various kinds of dependent adjunct (or adverbial) clauses, as opposed to the subordination of complements which is treated in Section 5.3.2 above. While the degree of grammatical integration in complement clauses is higher, adverbial clauses (and relative clauses) show less grammatical integration (Payne 1997: 307). Romyka adjunct clauses are predominantly finite, although some non-finite strategies exist, they are partly inherited (e.g., the infinitive) or have arisen under Turkish influence (nominalizations). The dependent clause is predominantly pre-posed and generally introduced by pre-verbal adverbial subordinators (this also applies to relative clauses and relative markers). However, borrowing of Turkish clause-final subordinators like *dije* is possible. Furthermore, the order of main and dependent clause seems to be variable and post-posed adverbial clauses are possible as well. Stacking of embedded clauses is allowed. Conditionals do not use subordinators but rather modal particles in the verb complex. The frequency of constituent orders in adjunct clauses is in the present data not clearly different from that of basic clauses, where both VO and OV orders are possible, although the latter is clearly reinforced by movement for pragmatic reasons. However, as far as frequency is concerned, it seems that in adjunct clauses OV order is more frequent, especially with pre-posed locations. According to Neocleous (2020: 85), word order of subordinate clauses<sup>236</sup> differs clearly from that of matrix clauses; the unmarked order in subordinate clauses is SOV with OSV attested under left-dislocation; while finite verbs in the subordinate clause trigger SOV order, non-finite subordinate clauses yield SVO order (Neocleous 2020: 118). The data from the present Romyka corpus seem largely to confirm this, although some counter-examples are available (e.g., exs. 411, 433e, this chapter). Given that AG had predominantly VO order in subordinate clauses, OV in most Romyka subordinate clauses is likely due to contact with head-final subordination in TR (Neocleous 2020: 299).

#### 5.3.3.1 Adverbial clauses

Adverbial clauses (ACs) are adjuncts that modify verb phrases or clauses by providing additional information on (i) time, incl. sequential and simultaneous time, (ii) location, (iii) manner, (iv) purpose, (v) reason, (vi) circumstances, (vii) conditionals, and (viii) other types like concessive, substitutive, additive or absolutive (Payne 1997: 316–320). Below strategies of forming adverbial clauses will be described for different kinds of (complex) ACs.<sup>237</sup> Conditional clauses are treated in an individual section (Section 5.3.3.2).

Adverbial clauses classify in most cases as finite subordination strategies (e.g., with adjuncts of time, location, manner and purpose), although some non-finite adverbial subordination strategies exist, often in addition to the finite ones (e.g., infinitives in ‘before’-clauses, participial forms in some adjuncts of time, nominalizations in adjuncts of purpose or in ‘without’-clauses). The complex diachrony of these adverbial clauses makes it often difficult to simply assume influence of Turkish non-finite subordination strategies. Following the Greek model, ACs are principally formed by adverbial subordinators, often subordinating conjunctions like *prin* ‘before’, *os* ‘until’, *omon* ‘like, when’, *anda* ‘when’, *ðae* ‘for’, but also interrogative adverbs like *pote* ‘when’, *pu* ‘where’. While most of these are of Greek origin, there exist also subordination strategies borrowed from Turkish like adjuncts of reason and purpose with *dije*. Furthermore, mere juxtaposition of clauses where the relation between the

<sup>236</sup> In the terminology of Neocleous (2020), it is unfortunately unclear which clause types he refers to as “subordinate clauses”. Based on the examples provided, he uses the term “subordinate clause” for both complement clauses and adjunct clauses such as conditional clauses.

<sup>237</sup> Note that in this chapter, it is only dealt with subordinate adverbial clauses and not with “simple adverbials”. The latter are not treated in an individual section but are briefly addressed in the word order section of declaratives (Section 5.2.1.1).

two clauses is to be inferred on contextual grounds is also frequent, especially in case of uncertainty of the speakers.

As part of the inherited strategy, which is still employed in SMG (Dryer 2013n, Map 94A), subordinating conjunctions seem to occur in principle in clause-initial position of the subordinate clause (402), although in the Romeyka corpus often other elements such as prepositional phrases are fronted so that the subordinator occurs in pre-verbal position, which is due to OV order in subordinate clauses often towards the end of the AC (403a)<sup>238</sup>. According to Neocleous (2020: 149, 169), – although referring to simple adverbials – if an AC is focused, it occurs to the left of the main verb. If this is adapted to subordination, the focus position of ACs is before the main clause. Indeed, in the present data, ACs occur mostly before the main clause (403), although they can also appear post-posed (see ex. 406 in Section 5.3.3.1.1). Both the predominant pre-posed position of the subordinate clause and its internal head-final structure are likely modelled on Turkish ACs, which predominantly consist of an infinite verbform + postposition (403b; see Kornfilt 1997: 67-68).

- (402) [*prin* *đosini* *ton* *paran* *Aiše*] *eyo* *pa* *tši* *pao*  
 before give.INF the money.ACC Ayşe.NOM I.NOM TOP not go.1SG  
 ‘I am not leaving, before Aise gives back the money.’ (Sitaridou 2014b: 130, ex. 44; glosses modified)

- (403) a. [*so pazari brin bane*] *na pao si fatimes* (C1)  
 b. [*Pazar-a gitmekten önce*], *Fatma ’yı ziyaret edeceğim*.  
 ‘Before going to the market, I will visit Fatma.’

#### 5.3.3.1.1 Adjuncts of time

ACs of time, including sequential and simultaneous time, show variation regarding the subordination strategies applied and a range of different subordinators is used. The different forms of temporal adjunct clauses are described in the following (i-x). In principle, however, temporal adjuncts are introduced by an adverbial subordinator which appears at the beginning of the subordinate clause (404; see Sitaridou 2014a; for description of individual subordinators, see Section 3.2.6.2). This is in line with the SMG type of adverbial clauses with clause-initial subordinators and finite verb forms (but cf. the use of the infinitive in *prin*-clauses in Hellenistic and until Medieval Greek time, Sitaridou 2014a). Turkish on the other hand has non-finite left-branching subordinate clauses with adverbial elements following the predicate (Christofaro 2013c, Map 126A; see 403b above).

- (404) [*Anda* *pašlaevis* *so* *fanimon*] [*os* *na* *piturevis* =*ata*],  
 when start.2SG at.the eating.ACC until PRT finish.2SG =OPN.3PL  
  
*u=* *poris* *na* *stetšis*.  
 not= can.2SG PRT stop.2SG  
 ‘Once you start eating, you can’t stop until you finish all of it.’ (ROf; Sitaridou 2014b: 46; glosses modified)

(i) For expressing temporal posteriority, Sitaridou (2014a) discusses the use of the subordinating conjunction *prin* ‘before’ + infinitive verb form (see ex. 402 above) – in the Romeyka corpus, + finite verb (405). Like in the majority of adverbial clauses, the AC is

<sup>238</sup> Note that according to Sitaridou (2014b), ‘before’ clauses with *prin* license the infinitive, which seems however neither to be the case in RSür (Özkan 2013: 149), nor does it occur in ex. (403a).

preposed, the subordinator immediately precedes the verb, in general appearing in clause initial position (but for a post-posed AC, see 406). Stacking of ACs seems to be possible (407).

- (405) [*so bazar prin bao*] *di fadime zijared n=eftao*  
 ‘Before going to the market, I will visit Fatma.’ (B1)
- (406) *i aiše eplinen ta xapsiæ [prin mairepsini]*  
 the.NOM Ayşe.NOM wash.AOR.3SG the.ACC anchovies.ACC before cook.INF  
 ‘Ayşe washed the anchovies before she cooked them.’ (Neocleous 2020: 61, ex. 60; glosses modified)
- (407) [*Prin pisini fain*], [*prin spudžisini so mandrin*], *tši pao.*  
 before make.INF food before clean.INF at.the barn not go.1SG  
 ‘I am not leaving before I cook and clean the barn.’ (Sitaridou 2014b: 135, ex. 68; glosses modified)

(ii) Clause-initial *os* ‘until’ + finite *na*-clause (Sitaridou 2014b: 135; see 408)

- (408) [*Os na ‘rte o Mehmetis na permeno=se.*  
 until PRT come.3SG the.NOM Mehmet.NOM PRT wait.1SG=OPN.2SG  
 ‘I’ll wait for you until Mehmet comes.’ (ROf; Sitaridou 2014a: 135, ex. 69, glosses modified)

Adverbial clauses with relations of temporal overlap, i.e., ‘when’-clauses, can be introduced by the subordinators *omon*, *anda*, *pote*, (*o*)*sat(i)*. Although their functional distribution is not fully clear (see also Section 3.2.6.2), it seems *anda* is used to express simultaneity, possibly with a generic sense (Drettas 1997: 377), while (*o*)*sat(i)* has a durative meaning of ‘while’; *omon* refers more punctually to a certain simultaneous interval (Drettas 1997: 378) and includes an aspect of manner (also reflected in its lexical meaning *omon* ‘like’); further research is required on the functional distribution of ACs of temporal overlap.

(iii) Clause-initial *omon* ‘when’, literally ‘like’, + *do* (N.DET<sup>239</sup>) + finite verb (409–411). But also note the example in (412) with the conjunction *dž* ‘and’ instead of *do*.

- (409) [*omon d=eroise s=ormi*] *ejendune na-tsurula*  
 ‘When she fell into the water, she got wet.’ (04\_01072019F\_13; 45)
- (410) [*omon d=erakse=me*] *išd ebiyame son doxtorin*  
 ‘When it bit me, we went to the doctor.’ (02\_29062019F\_2; 32–35)
- (411) [*ebedž ümon d=udž evra do džesdani*] *to mural=im ebozulefde*  
 ‘When I could not find the wallet, I was in a bad humour.’ lit. ‘my moral was broken’  
 (08\_04072019M\_1; 173–175)
- (412) [*omon dž=ebiyam*] *oneon dodže šimu*  
 ‘When we went, suddenly it rained a heavy rain.’ (03\_30062019F\_6; 41)

(iv) pre-verbal *anda*<sup>239</sup> ‘when’ + finite verb, occurring predominantly in head-final ACs with SOV order (413, 414). *Anda* has scope over the whole clause, so it does not interfere with

<sup>239</sup> Also in the form *onda*, see ex. (420).

clausal negation (e.g., ex. 418). While most ACs with *anda* are pre-posed, the AC in (415) is post-posed. Importantly, subordinate clauses with *anda* seem to be predominantly head-final (413–415) while VO order in the subordinate clause is possible as well (416–418).<sup>240</sup> Stacking of ACs with *anda* is possible (419), but consider the unclear analysis of (420), where the intermediate clause probably is an AC as well. Note that it is sometimes difficult to distinguish ‘when’-clauses from conditional clauses (see the different strategies provided in 421).

- (413) *[to xora kser anda en] pao so xorafe*  
 ‘When the place is dry, I go to the fields.’ (03\_30062019F\_7; 06)
- (414) *[adin i drana anda xandane] to yarðeli=s düšünev*  
 ‘When the elders die, you<sup>?</sup> think of your child.’ (01\_040202016F\_1; 088)
- (415) *havu dina erodas [d=edžer emen anda utš inanevis]*  
 ‘You [can] ask her, when you don’t believe me.’ (03\_30092019F\_8; 09)
- (416) *[anda erde do tuvaletu=t] na bai etši*  
 ‘When he needs to go to the toilet, he goes there.’ (04\_01072019F\_13; 07–08)
- (417) *[anda n=arxume tomara] ula do didžes idena na im*  
 ‘When we come all together, we will all be there?’ (translation unclear; 03\_07072019F\_1; 42)
- (418) *[so stoma=s anta utš apomenun ðonte], esi pa etote omon emena ylikomalezo na tros*  
 ‘If no teeth remain in your mouth, you will eat polenta like me.’ (Tursun 2019: 221)
- (419) *[anda θelename bola kala insana anda ine.. ersane] panda epejnam etši*  
 ‘When we wanted, when very nice people came, we always went there.’  
 (09\_04072019\_7; 21–23)
- (420) *[temeteron to mahallen onda ertes] [me ti mana=s taništīm]<sup>?</sup> esi etrižes ežis*  
 ‘When you came to our neighbourhood, [when<sup>?</sup>] I met your mother, you run there.’  
 (01\_070702019F\_1; 02–03)
- (421) a. *[onda n=erse] xarume* (H2)  
 b. *[eomon to tš=erxo] bola yabiskume* (H3)  
 c. *[an erse] bola xarenume* (H1)  
 ‘I am happy, if/when you come.’

Note that in the construction ‘When we were small’, the subordinator *anda* seems to be optional (422a/b).<sup>241</sup>

- (422) a. *[mikrina istemunesten] emist pal so mektep epiyame*  
 b. *[mikrina anda emunesten] emist pal so mektep epiyame*  
 ‘In our childhood, we used to go to school.’ lit. ‘when we were small’ (A1)

<sup>240</sup> And not only with non-finite subordinate clauses, as Neocleous (2020: 118) suggests.

<sup>241</sup> Note also the use of *mikri* ‘small’ as temporal adverb in ex. (vii).

(vii) *[mikri] emis ađatšes eporpatesam*  
 ‘When we were small, we used to walk here.’ (08\_04072019M\_1; 012)

(v) Interrogative *pote* ‘when’ + finite verb (423);

- (423) *[fasulijas bote en] erxume bero=na*  
 ‘When there are beans, I come and take it?’ (03\_30062019F\_11; 041)

(vi) Particle *(o)sat(i)* (cf. also Section 3.2.4) + finite verb (424, 425). The AC is frequently pre-posed, but postposition is possible (426).

- (424) *[ta yarðeli=nat osad ebejin] eðokan adona ena suri hedijeðes*  
 ‘When he returned to the city, the children gave him some presents.’ (C1)

- (425) *[sad erxumune] t=aleyo=m ege estrofi*  
 ‘When coming, my horse stumbled (down).’ (08\_04072019M\_2; 100–103)

- (426) *sürükler [öseti kaðe]*  
 ‘It (=a baby bear) drags, when it sits.’ (02\_29062019F\_2; 26)

(vii) Some examples feature modal particles such as *na* (427, 428) and *an* (429) + finite verb to express a ‘when’-clause, while *an* is actually the conditional particle used for ‘if’-clauses.<sup>242</sup> Possibly, these constructions are modelled on Turkish *-(y)Ip* converb constructions which allow combination of several verbs. On the other hand, according to Christofaro (2013c), if the veridical status of a ‘when’-clause is not clear, conditionals or subjunctives may be used in the AC but this does probably not suffice alone to explain the use of modal particles in ‘when’-clauses.

- (427) *jani [gözüz na en] tomara na başlaeps o kseraθinimo adode-džes*  
 ‘Well, when it will be autumn, we will together start the drying at that time.’  
 (03\_07072019F\_1; 44)

- (428) *u=θelume na bame [nde na bam leyome]*  
 ‘We don’t want to go, what do we say when going.’ (translation unclear;  
 01\_04022016F\_1; 141)

- (429) *[am pao so raši] n=epero liyo diri tše na fero*  
 ‘When I go to the mountains, I take a bit of cheese with me.’ (C1)

(viii) Participial verb forms in the pre-posed AC may yield an adverbial meaning without a subordinator, as attested for relations of temporal overlap (‘when’-clause) and temporal anteriority (exs. 430, 431).

- (430) *[...] [vejaxute nejenene to zo] epukan=eθe n=epiname fila*  
 ‘[...] Or when the cows give birth, we put leaves underneath them.’ (04\_01072019F\_1;  
 172–173)

- (431) *[ado ebitjepsenme] endama imestine*  
 ‘After finishing this, we will be together.’ (08\_04072019M\_3; 050)

<sup>242</sup> There is also a hapax legomena (viii) using the particle *ha* whose analysis is unclear, though.

(viii) *[so xor ha pies] kadi pao eftayo kadi fidevo kadi tsakono*  
 ‘When you went to the field, I go [and] make something, I plant something, I break something.’  
 (03\_30062019F\_7; 07)

(ix) Moreover, mere juxtaposition of clauses like in serial verb constructions (Section 5.3.1.2) may suggest some adverbial relation among the clauses like simultaneous time in (432a). Again, Turkish temporal converbs may have an influence here, for example like in (432b) where *-DİğIndA* describes a simultaneously ongoing action (Göksel & Kerslake 2005: 415).

- (432) a. *kati ta malia=s elaiskusanie etrešes*  
 ‘Somehow<sup>?</sup> your hair waved while you were running.’  
 b. *saçların kostuğunda dalgalanıyordu*  
 ‘Your hair was waving while you ran.’ (03\_07072019F\_1; 15)

(x) Finally, some hapax legomena exist mainly in data elicited by means of the translation task from Turkish for relations of temporal anteriority and overlap. The optative particle *as* + *d* (neuter determiner<sup>?</sup>) is used in immediate pre-verbal position as a subordinator in some examples (433a/b). In others, however, *as* alone is used as subordinator (433c) or left out in total (433d). Note that Drettas (1997: 383) describes the use of the ablative preposition *as(o)* (i.e., with merger of the definite article with the preposition *as* ‘from’, although in the majority of his examples only the preposition figures) as temporal subordinator before finite verbs in aorist tense. According to him, this construction expresses a “*terminus post quem*”, roughly corresponding to a meaning of ‘since’, which may as well express a causal relation. (433e) features a subordinator of unclear analysis with the AC placed in rare intermediate position. Lastly, note the use of the adverbial *har(e)* ‘now’ + nominalization, apparently including a Turkish genitive, as temporal subordination strategy in (434). Possibly, the occurrence of *har(e)* in this context is related to the continuance of the AG connective discourse particle *γάρ* in the Romeyka adverbial *har* ‘now’ (see Section 3.2.4.5.2).

- (433) a. *[o tširis ades.. to faji as d=efaje] aso spidi eksev* (C1)  
 b. *[o tširim to fajin as d=efaen tš=arθen] so porpateman epijen* (H2)  
 c. *[to fajin as efoen] o džiri=m epije so porpatema* (H3)  
 d. *[to fai=nat efaen] o tširin tšebedža pijen so porbatin* (H1)  
 e. *o džiris ades [avi d=efaen do fajido] as=osbe (e)kserin* (B1)  
 ‘After having eaten, her father left the house for a walk.’

- (434) *[o har je.. hare jeterun d=arde] o zaman na bame eban*  
 ‘When Yeter comes, we will go up.’ (04\_01072019F\_17; 59–60)

### 5.3.3.1.2 Adjuncts of location

In the present data, only few examples of an AC of location are available using the interrogative *pu* ‘where’ as (indefinite) subordinator at the beginning of the pre-posed subordinate clause (435, 436). Ex. (437) features relativization of a spatial adjunct (see also Section 5.2.3.3.5 on embedded interrogatives).

- (435) *[pudžega na bas] demedera na bas*  
 ‘Wher(ever) you go, you will go to us.’ (06\_03072019M\_2; 04)
- (436) *[erpu en] epejename*  
 ‘Wherever it is, we went [there].’ (09\_04072019\_7; 14)
- (437) *džaka [pu en]*  
 ‘Down there is where she is.’ (translation unclear; 07\_04072019F\_5; 37)

## 5.3.3.1.3 Adjuncts of manner

ACs of manner also use *omon* ‘like’ as a subordinator. In ex. (438) it is used together with the manner adverbial *haets* ‘so, thus’ at the beginning of a pre-posed subordinate clause in intermediate position between two coordinated clauses.

- (438) *eyo esena hats leya ama [haets omon do leyo] yolajin utš edun*  
 ‘I told it you like this, but it was not so easy like I said.’ (02\_29062019F\_2; 47–49)

Furthermore, juxtaposition of clauses like in verb serialization is used indirectly implying a subordinate relation (439).

- (439) *epi utš epiya erθa*  
 ‘I went and I came without taking them.’ lit. ‘I went, did not take [any] and came.’  
 (03\_30062019F\_11; 043)

## 5.3.3.1.4 Adjuncts of purpose

ACs of purpose apply different strategies although none of these seems to be central:

(i) According to Sitaridou (2014b: 41), adjuncts of purpose are formed with the clause-initial subordination conjunction *ðæ* ‘for’ (< HelGr. *ðia* ‘for’) + nominalization, which were used for adjuncts of reason in HelGr (440). This strategy could unfortunately not be attested in the Romeyka corpus. However, according to Tursun (2019: 230), *ðe* is used in ROF as spoken in Uzungöl (Saraxos) to express causality (441).<sup>243</sup> Furthermore, the purpose preposition *ja*, e.g., *ja ta za* ‘for the cows’ (04\_01072019F\_1; 168) can apparently also function as subordinating conjunct (442, 443).

- (440) [*ðæ temon t= erθan=im*] *xavitsin eftes =me.*  
 for my the= coming=POSS.1SG pudding.ACC make.2SG =OPN.1SG  
 ‘For the sake of my coming, you will make me some pudding.’ (ROF; Sitaridou 2014b: 41, ex. 27, glosses modified)

- (441) *ðe t=ata utš erθa*  
 ‘That is why I didn’t come.’ (Tursun 2019: 231)

- (442) *frazo steko to xorafi, jia na min troi=e o arkos*  
 ‘I am hedging the field so that the bear cannot eat it.’ (Tursun 2019: 527)

- (443) [...] *na xuxuvono jia na ðeveno*  
 ‘I have to bend in order to pass.’ (Tursun 2019: 563)

(ii) A subordination strategy including the Turkish subordinator *diye*, which is, among others, used in Turkish for expressing purpose (Göksel & Kerslake 2005: 400–401) is borrowed into Romeyka for adjuncts of purpose and reason (see below; see also Section 3.2.4.3). In (444), the Turkish participle form of *diye*, *diyene* ‘saying’ > regional spoken Tr. *deine* is used as clause-final subordinator in a clause with finite predicate.<sup>244</sup> In subordination with *diye*, the participants of the main and dependent clause may be different (as in 444).

<sup>243</sup> Note also the rare (apparently negated) form *muðe* in the AC in (ix), where it seems to have a meaning of ‘without’.

(ix) *muðe krus tin porta, laya na anizo=se*  
 ‘If you don’t knock the door, how shall I open you?’ (Tursun 2019: 118)

<sup>244</sup> Note that in Cappadocian, a similar construction is used with the particle *deji* (M. Janse, p.c.).

- (444) *anda utš en panda si ðulia=muna epejname [ti ðulian n=eftjame deine]*  
 ‘When there were none, we always went to our work for the work should be done.’  
 (09\_04072019\_7; 24–25)

(iii) The example in (445) shows that the speaker is unsure about the correct production of a Turkish translation prompt and provides first a conditional clause with the particle *an* and then a *na*-clause. Finally, she provides a periphrasis by means of clause coordination.

- (445) *[tipon an ne.. n=ebern] na pao so pazar n=ebero tipo tš=erxume*  
 ‘I went to the market to buy something.’ (C1)

In a non-veridical adjunct of purpose, *na* + negator *mi* are used (446).

- (446) *iben mana daha ndo kahese ipa [na m=eftayo šamada apes]*  
 ‘He said, “Mother, why are you sitting here?”. I said, “In order not to make noise inside.”’ (03\_30062019F\_6; 30–31)

(iv) A nominalization strategy including the preposition *s(e)* ‘at, to’ + deverbal noun is used with motion verbs in (447) and (448). Given the fact that Turkish applies mainly non-finite adverbial clauses (Christofaro 2013c; Map 125A), the nominalization strategy is likely influenced by Turkish.

- (447) *[apsimo s epario me] erθes*  
 ‘Did you come to take fire?’ (Tursun 2019: 199)
- (448) *[si komsidaes=muna so zijareti] na bame*  
 ‘We will go to visit our neighbours.’, lit. ‘We will go to our neighbours to the visit.’  
 (08\_04072019M\_3; 109)

(v) A semantic relation between main and dependent event can be even implied from mere juxtaposition of the clauses, like in (449, 450).

- (449) *ba na bam arajevom do ðromo*  
 ‘We will go [to] seek the way.’ (08\_04072019M\_1; 179)
- (450) *jukesegin en u=bori na landžef*  
 ‘It is high [so that] it cannot jump.’ (02\_29062019F\_2; 17)

#### 5.3.3.1.5 Adjuncts of reason

Causal adjunct clauses are formed in Romeyka by borrowing (matter and pattern borrowing, in the sense of Matras 2009) of the Turkish subordinator *diye*, a converb of *de-* ‘say’, used in Turkish finite adverbial clauses to express reason or purpose (Göksel & Kerslake 2005: 400). Both the form and the syntactic properties of *diye* to appear in clause-final position and to trigger finite subordination are copied in Romeyka adjuncts of reason (451–453); see also Sections 3.2.4.3 and 5.3.3.1.4. In most examples below, the subordinator *dije/dijene* appears in immediate postverbal position like in Turkish, but cf. (453) where word order in the subordinate clause is VO. From a typological point of view, it is not rare that ACs of purpose and reasons are treated alike (Christofaro 2013d). This is also evident in Turkish where adverbials of purpose and reason are both formed by the postposition *icin* ‘for’, albeit with slightly different verbal endings (Kornfilt 1997: 73).



- (451) *avudin ul dinlejevune=me [u-rumdža konušefkume deje]*  
 ‘They are all listening to me for I speak Romeyka.’ (01\_15022015F\_1; 16–17)
- (452) *[emorfo etune deine] epira=do*  
 ‘Since it is beautiful, I took it.’ (A1)
- (453) *bazilarina jardimdže epi inumunestine [ixane sevdaris dije]*  
 ‘We helped some for that they had a lover.’ (04\_01072019F\_2; 171)

Note that data from a translation task from Turkish only yield juxtaposed clauses without subordinator and pre-posed AC, where the relation of clauses needs to be inferred on contextual grounds (454a–e).

- (454) a. *[utše tšelišepse] so sinavi bašarilis utš ejendune* (C1)  
 b. *[bunedž u=tšališepsen] so sinavin utš eborisen kazanepsin* (H1)  
 c. *[utš=eđulepsam bola] abedja da sinave utš eboresen besin* (H2)  
 d. *[Utš eđuilepses] (si)o sinavim tšumur eservivine* (H3)  
 e. *[bola utše na tsališevo] bašlariz utš eboresa inana* (B1)  
 ‘He didn’t succeed in the exam because he didn’t work hard enough.’ (translations partly unclear)

#### 5.3.3.1.6 Circumstantial adjuncts

Inherited ‘without’-clauses are formed by the clause-initial subordinating conjunction *aθiyo(n)* ‘without’ + deverbal noun in a pre-posed adjunct clause (455; Sitaridou 2014b: 135).

- (455) *[Aθiyon tšalisema] paraðes tši porume na ftæme.*  
 without work money not can.1PL PRT make.1PL  
 ‘We cannot make any money without working.’ (Sitaridou 2014b: 135, ex. 71a; glosses modified)

In data translated from Turkish, ‘without’-clauses seem to be merely expressed by juxtaposition (456a), like with other types of ACs above. The word order with intermediate subordinate clause is very likely induced by word order of the Turkish prompt (456b).

- (456) a. *i fadime [tibu tš=efaje] aso spidi eksevi* (C1)  
 b. *Fadime bir şey yemedden evden çıktı.*  
 ‘Fatma left the house without having eaten anything.’

#### 5.3.3.2 Conditional clauses

Conditional clauses are a type of adverbial clauses that specifies under which condition the event of the main clause takes place. Additional information may be given on the likelihood of the condition being fulfilled (hypothetical, counterfactual conditionals; Christofaro 2005: 160). In Romeyka, possible, hypothetical, and counterfactual conditionals exist. The veridicality of a conditional clause is expressed by TAM marking of the if-clause. Conditional clauses in Romeyka are formed by means of the particle *an* + finite verb (457), although the particles *na* (see 458 for the interchangeable use of *an* and *na*) and optative *as* (459) can be used as well. Table 37 provides an overview about the formation of different types of conditional clauses. The conditional clause is always followed by the independent clause.

- (457) [*an* *θelide*] *ebero=sas tše bao adutšam*  
‘If you like, I take you and go up there.’ (08\_04072019M\_4; 03–04)
- (458) [*para na* *ešis saluxe ba an* *ešis*] *ađaga tibu tšališevs do kendis teries emeklis ba na inese na stetšis ađatšeka*  
‘If you have money, if you also have a good health, you work something, you maintain yourself, when you retire, you can stay here.’ (08\_04072019M\_1; 305–309)
- (459) *as ixa enan bats-obon omo hatena hajes treš*  
‘If I only had a little girl like her who runs like this.’ (03\_07072019F\_1; 13–14)

**Table 37:** Non- and antiveridical mood in Romeyka (adapted from Table 3 in Sitaridou 2014a: 122)

		If-clause	Matrix clause	Examples
Nonveridical	Possible	<i>an</i> + present	<i>n(a)</i> + present <i>as</i> + present present imperative	(460, 461) (462) (463) (464, 465)
		<i>n(a)</i> + present	<i>n(a)</i> + present present	(463)
	Irrealis	<i>an</i> + imperfect	<i>n(a)</i> + imperfect imperfect	(466, 467)
		<i>n(a)</i> + imperfect	<i>n(a)</i> + imperfect	(468–470)
Antiveridical	Counterfactual	<i>an</i> + <i>ixa</i> + infinitive	<i>ixa</i> + infinitive	(471)
		<i>n(a)</i> + <i>ixa</i> + infinitive	<i>n(a)</i> + imperfect	(472–475)
		<i>as</i> + <i>ixa</i> + infinitive	<i>n(a)</i> + imperfect	(476)

- (460) *an* *roizune dži na futšiyundane*  
‘When they fall inside, they will drown.’ (04\_01072019F\_13; 05)
- (461) *ena liko an elebo na pao kliđono tin borta tše nbena apes*  
‘When I see a wolf, I go lock the door and enter inside.’ (C1)
- (462) *i sevda=m an evriški as efte=a jakaluk*  
‘When my love finds it, she makes it a collar.’ [folksong] (04\_01072019F\_1; 092)
- (463) *para na* *ešis saluxe ba an* *ešis ađaga tibu tšališevs do kendis teries*  
‘When you have money, when you also have health, you work something hear and you maintain yourself.’ (08\_04072019M\_1; 305–307)
- (464) *an* *elebis enan ligon ayom s=ospidi=s t=ospidi=s kliđa*  
‘When you see a wolf, go home and lock the house.’ (H2)
- (465) *an* *ine na pas apo to har ipé=m=a*  
‘If you will (certainly) come, tell me now already.’ (Tursun 2019: 371)
- (466) *müslüman an* *esane tibu tš=aileyane=mase*  
‘If they were Muslims, they couldn’t tell us anything.’ (04\_01072019F\_2; 292–293)

- (467) *an dž=eyabenis=a abađadžega so xor üts epejnis*  
 ‘When you don’t like it here, you didn’t go to the place.’ (08\_04072019M\_2; 124–125)
- (468) *na ixa para tše exartseva kalo n=edone*  
 ‘If I had money to spend, how nice would that be.’ (Tursun 2019: 371)
- (469) *asinda du mintanis n=jedune da šašdim ona*  
 ‘Actually it would have been ‘of a shirt’, I was mistaken.’ (03\_07072019F\_1; 30)
- (470) *opse n=erxusun*  
 ‘You would/should have come yesterday.’ (Tursun 2019: 371)
- (471) *An ixa mairepsini, ixame fanini.*  
 COND have.IPFV.1SG cook.INF have.IPFV.1PL eat.INF  
 ‘If I had cooked, we would have eaten.’ (Sitaridou 2014b: 136, ex. 73b; glosses modified)
- (472) *na ixa šita piterupsini tas ðulias,*  
 PRT have.IPFV.1SG immediately finish.INF the.ACC chores.ACC  
*xar n=epinamen parakaθin*  
 now PRT=make.IPFV.1PL gathering.ACC  
 ‘If I had finished the chores immediately, we would have been gathered together now.’  
 (Neocleous 2020: 57, ex. 53; glosses modified)
- (473) *[...] eyo na mi ixa kse.. bisena ksila ev.. ešgisa utše n=eborinam dž=efdename don pešgo*  
 ‘When I had not chopped the wood, I could have not lit the oven.’ (04\_01072019F\_13; 49–50)
- (474) *[...] eyo na mi ixa škisenda da ksila dže eftena don pešgo laya n=isayumunese laya n=ixlimunesine*  
 ‘If I had not chopped the wood and made the oven, how would we get dry, how would we get warm?’ (04\_01072019F\_13; 52–55)
- (475) *na mutš ixa kal ipene ti mana=m,*  
 PRT NEG have.IPFV.1SG again say.INF the.ACC mother.ACC=POSS.1SG  
*i mana=m n=estetšen.*  
 the.NOM mother.NOM=POSS.1SG PRT=be.IPFV.3SG  
 ‘If I hadn’t said (that) to my mother again, my mother would have stayed (here).’  
 (Neocleous 2020: 102, ex. 36; glosses modified)
- (476) *as ixa eliyon fanini, xar omorfa n=esteka.*  
 PRT have.IPFV.1SG less eat.INF now well PRT be.IPFV.1SG  
 ‘If only I had eaten less, I would have been well now.’ (Neocleous 2020: 57, ex. 52; glosses modified)

Present tense conditionals are either formed by *an* or *na* + finite present tense verb in the subordinate clause (see also Drettas 1997: 327). The verb of the matrix clause is either in future (460) or present tense (477). It can be also in optative (478) or imperative mood (479). Past tense conditionals that express irrealis mode are formed by *an* or *na* + finite verb in imperfect tense, e.g., *na emune* ‘I would have been’ (Tursun 2019: 243; see also Drettas 1997: 330). The

verb of the matrix clause is in imperfective as well (466 above, but note that this example seems to suggest a counterfactual reading). Interestingly, the example from a Turkish-dominant speaker in (480) uses present tense in the if-clause and the imperfect in the main clause to express past. Antiveridical counterfactuals are in general formed by *an/na* plus an auxiliary *ixa* ‘had’ with governs the infinitive. According to Sitaridou (2014b: 49), there is micro-variation in the formation of counterfactuals: in ROf, the subordinate clause requires a plain infinite (481), while in RSür as spoken in Beşkøy, the counterfactual requires an inflected infinitive (482). Furthermore, Sitaridou (2014a: 55) notes the possible emergence of a new invariant modality marker *ixe* from the former auxiliary *exo* ‘have’/*ixa* ‘I had’, whose invariability cannot be proven at this stage, though, due to the lack of counterfactuals with other than first person singular. An optative of the ‘if only’-type with *as* + *ixa* is possible as well to express counterfactual wishes (476 above; Neocleous 2020: 57); but note that the optative *as* + imperfective in (483) also expresses a counterfactual meaning. The matrix clause has either a *na*-clause in the imperfect or (according to Sitaridou 2014b) again *ixa* governing the infinitive (471 above). It seems perfective aorist forms cannot be used in Romeyka conditional clauses, but Drettas (1997: 329) notes their occurrence in PG.<sup>245</sup> They appear with *an* + aorist in the if-clause and, for example, a *na*-clause in present tense in the main clause (484).

- (477) [*eyo an bayo eksu*] *roizo so nero*  
 ‘If I go outside, I fall into the water.’ (04\_01072019F\_13; 22–23)
- (478) [*an thelide*] *as bam etši*  
 ‘If you like, we can go there.’ (08\_04072019M\_3; 189–190)
- (479) [*i likon an elebis*] *ayome s=ospidin tin borta=s kliða*  
 ‘If you see the wolf, go home and lock your door.’ (H1)
- (480) [*tipu an en*] *eðone=mase anda utš en mo to limo beftename ka*  
 ‘If there was something, she gave it to us, when there was nothing, we went to bed hungry.’ (02\_02022015F\_1; 118–119)
- (481) *An ixa mairepsini, n‘ etroyame.*  
 if had.1SG cook.INF PRT eat.IPFV.1PL  
 ‘If I had cooked, we would have eaten.’ (ROf; Sitaridou 2014b: 45, glosses modified)
- (482) *Na ixa mairepsina etroyamen.*  
 PRT had.1SG cook.INF.1SG eat.IPFV.1PL  
 ‘If I had cooked, we would eat.’ (RSür; Sitaridou 2014b: 49, glosses modified)
- (483) *as elebame denandalo*  
 ‘If we (only) could have seen each other.’ (01\_15022015F\_1; 22)
- (484) *an ipes psemata tha skotonosen*  
 ‘If you (ever) lied, I (will) kill you.’ (Drettas 1997: 329, ex. 65)

In negated conditionals, the negation particle is placed immediately before the verb, preceded by the conditional particle (467 above, 485). Negation of the subordinate clause does not affect

<sup>245</sup> In general, in the present data the aorist seems not to occur frequently with *na* (but see ex. 427 in this chapter and ex. 300 in Chp. 3). Rather, there are many examples of *na* + present indicative where SMG would use the aorist subjunctive which indicate that the distinction between perfective and imperfective is neutralized in such cases (M. Janse, p.c.).

the matrix clause (neither does negation of the matrix clause affect the subordinate clause); both the conditional clause and the matrix clause can be negated (467 above). In the two examples of negated counterfactuals (exs. 473 and 474 above), the VP of the subordinate clause consists of *na + mi +* (inflected) infinitive, and the matrix clause requires *na/n' +* imperfective.

- (485) *an dži θelide zamani an dž=üdz ešide..*  
 ‘If you don’t like, if you don’t have time..’ (08\_04072019M\_4; 05–06)

Stacking of conditional clauses seems to be possible, either by means of the coordinating conjunction *tše* (474) or by means of juxtaposition without repetition of the subordinator (486). Note, however, that in ex. (486) the scope of the subordinate clause is not fully clear. Two subordinate clauses of the same matrix clause and each introduced by their own subordinator can also occur (ex. 463 above, but note again that the exact scope of the subordinate clause is not clear here either).

- (486) *[ađadžeka ilos bal an jen anemo fisai gatsimale sa rašia fisai] avudijega en rahati*  
 ‘If there is sun here, a wind blows, clouds enwrap the mountains, it is comfortable here.’  
 (08\_04072019M\_3; 130–133)

Finally, it needs to be noted that conditional clauses sometimes alternate with adverbial clauses of temporal overlap, i.e., ‘when’-clauses (487; see also Section 5.3.3.1.1).

- (487) *[tipu an en] eđone=mase [anda utš en] mo to limo beftenam eka*  
 ‘If there is<sup>2</sup> something, she gave it to us, when there was nothing, we lay down hungry.’  
 (02\_02022015F\_1; 118–119)

### 5.3.4 Relative clauses<sup>246</sup>

Relative clauses (RCs) in Romeyka are nominal modifiers that are examples of finite subordination. Romeyka has externally headed relative clauses with the head noun outside the RC (cf. Dryer 2013g). RCs in Romeyka are usually prenominal (i.e., in the same position as other nominal modifiers) and in preposed position (488), although they can be inserted in the matrix clause as well (489). Postnominal RCs exist as well (490–492); according to Neocleous (2016: 72), possibly a reflex of AG continuance. Both subject and object relative clauses are allowed. The scarce examples of RCs available do not allow any judgement whether the position of subject RCs and object RCs is different; both seem to exist as well pre- as post-nominally, although in the present corpus, pre-nominal RCs are the vast majority. It seems that under strong contact, zero marking/juxtaposition of relative clauses becomes more likely. While inherited relative markers are diachronically often clause-initial, in the data of the present corpus, they appear strictly in pre-verbal position, which is under OV order in the dependent clause towards the end of the clause. From a typological point of view, VO languages overwhelmingly have post-nominal RCs (Payne 1997: 326), while OV languages have often pre-nominal RCs although post-nominal RCs are also possible (Neocleous 2020: 237). PG has (in line with SMG) predominantly post-nominal RCs (493; Drettas 1997: 347–368), while Turkish as OV language employs pre-nominal RCs with a participial strategy, i.e., the verb in the RC is a participle (494). So, right-branching structures (i.e., postnominal relative clauses) are presumably a reflex of the type widely-attested in AG (but see Bentein & Bağrıacık 2018 for a prenominal “third type of headed relative clause” in Post-classical and early Byzantine Greek), and still maintained in SMG (see Nicholas 1998b for relativization in MedGr). The left-

<sup>246</sup> Note that since relative clauses do usually not figure extensively in naturalistic spoken data, the data basis for this chapter is relatively weak. Further research on relative clauses in Romeyka is certainly needed.

branching word order within the RC, however, where the verbal complex of relativizer + clause-final verb<sup>247</sup> can be preceded by any constituent belonging to the RC, e.g., a PP (488, 489) or a direct object NP (495), is certainly contact-induced by Turkish prenominal and internally head-final relative clauses (494), and provide a good example of how language contact can yield typologically unusual structures. There are also evident similarities to geographically adjacent Laz (Gandon 2016: 221), which has prenominal finite relativization introduced by preverbal subordinators (Lacroix 2018).

- (488) *[aso istanboli ots erte] o mualimis sin džami-tšeka katse*  
 ‘The teacher who came from Istanbul lives opposite the mosque.’ (C1)
- (489) *axmedis [son barxa nde ba] o đormo ediksen=a*  
 ‘Ahmet showed her to road that leads to the pasture.’ (B1)
- (490) *esina [so dolabin do ine] mono džidin ebsindane*  
 ‘Those who are at the repository cooked a bit difficult.’ (03\_30062019F\_11; 089–099)
- (491) *Eyrapses etšinon [do ipamen]?*  
 record.AOR.2SG that.ACC REL say.AOR.1PL  
 ‘Did you record what we have said?’ (Gandon 2016: 221, glosses modified)
- (492) *o peđas [op erθen aso čičenin]*  
 the.NOM child.NOM REL come.AOR.3SG from.the.ACC grocery’s.ACC  
*t=emón t=anépsin en.*  
 the=mine the.NOM=nephew.NOM be.3SG  
 ‘The child who came from the grocery’s is my nephew.’ (Neocleous 2020: 71, ex. 87; glosses modified)
- (493) *tin óran [to k=elép=aten] káθume ka ke kléyo*  
 ‘The moment when I don’t see her, I sit down and cry.’ (Drettas 1997: 354, ex. 15, presentation modified)
- (494) *[kitab-ı al-an] öğrenci*  
 book-ACC buy-PTCP student  
 ‘the student who bought the book’ (Comrie 1998: 82, in Comrie & Kuteva 2013)
- (495) *Opse iđa alis p=epiren ineka.*  
 yesterday saw.1SG Alis.NOM REL =got.married.1SG woman.ACC  
 ‘Yesterday I saw the woman who Ali married.’ (Gandon 2016: 222, glosses modified)

Headless relative clauses (HRCs), frequently with nonspecific head nouns, are possible as well, both for relativization of objects (496, 497) and subjects (498). HRCs occur with all three relativizers, i.e., *op*, *ots*, and *do* (see Section 3.2.2.6). Note that with the object HRC in (496), pronoun retention occurs for the direct object. This is a rare incidence where ‘clitic doubling’ (see Section 5.2.1.3) occurs in Romeyka, that is otherwise common in Christian PG (see, e.g., Drettas 1997: 353, ex. 13a).

<sup>247</sup> But cf. ex. (492) from Neocleous (2016: 71) with an RC internal verb-medial order, which is rare, given Neocleous (2020) attests verb-final orders in (finite) subordinate clauses, otherwise.

- (496) *[aso Katoχor op erθe] iða=na*  
 from.the Katochor REL came.3SG saw.1SG=OPN.3SG  
 ‘I saw the one that came from Katochor.’ (A1)

- (497) *[har to leyo] inšallah ormofa ine*  
 ‘What I say now is hopefully nice.’ (03\_07072019F\_1; 33)

- (498) *i [aso istanbuli ots erte] temo i jenge je*  
 ‘The woman who comes from Istanbul is my aunt.’ (C1)

The relativized head noun is usually expressed in the matrix clause (except for headless RCs). Relative markers (see Section 3.2.2.6) indicate the relativized head in the relative clause, occur in general in immediate pre-verbal position (Gandon 2016: 220), and are invariant to any case-marking that could indicate the grammatical role of the relativized head noun within the RC. Their status in Romeyka could be possibly considered as relative pronouns, at least following the definition of Payne (1997: 326): “If the relativizer reflects some properties of the NP<sub>rel</sub> within the restricting clause (e.g., humanness [...]), then it can be termed a relative pronoun”. In contrast, Drettas (1997: 365) notes that relativizers in PG may be analysed as connectors rather than relative pronouns.

In addition to the relative marker strategy, Romeyka also employs a strategy without relativizer (499a, but cf. 488 above), although these examples seem to be clearly modelled on the Turkish prompt in a translation task (499b). While SMG uses the relative pronoun strategy for subject/object relativization, Turkish does not employ free relativizers (Comrie & Kuteva 2013, Map 122A/123A).

- (499) a. *[aso istanbulin erθen] o mualimis sin džamian-džega gatsenga* (H1)  
 b. *[Istanbul’dan gelen] öğretmen caminin karşısında oturuyor.*  
 ‘The teacher who came from Istanbul lives opposite the mosque.’

Grammatical relations that can be relativized are subjects (e.g., 488 above) and (direct) objects (e.g., 489 above); as well as obliques such as spatial complements (500, 501). All relativization strategies (see Section 3.2.2.6) can be applied for relativization of both subject and object NPs, but note that *oti/ots* is only attested in subject RCs. The Accessibility Hierarchy of Relativization predicts which grammatical relations can be relativized: subject > direct object > indirect object > oblique > possessor (Keenan & Comrie 1977, in Payne 1997: 335). If a certain position is relativizable, all positions to its left are also relativizable. Also note that pronoun retention is more frequent on the lower end of the scale, i.e., for obliques (Comrie & Kuteva 2013).

- (500) *[sin džami etšika ndo en] t=ospı temo en*  
 at.the mosque down\_there REL be.3SG the=house POSS.1SG be.3SG  
 ‘The house at the mosque is mine.’ (A1)

- (501) *atu-merea [do en o džamis] temon en*  
 this-part.ACC REL be.3SG the mosque.NOM POSS.1SG be.3SG  
 ‘That place in which the mosque is is mine.’ (Gandon 2016: 221, translation adapted)

## 6 Language contact

Three research questions have been stated in the introduction in order to disentangle the factors that have contributed to the grammatical shape of Romeyka today: (I.) How strong is the Turkish influence on Romeyka? (II.) Is language shift impacting upon the grammatical structures of Romeyka? and (III.) How “Greek” is Romeyka? These questions will be briefly examined in this chapter to give a preliminary estimation of what further in-depth research needs to prove. Examining language contact and determining the causation of a certain feature is not a simple undertaking, since in order to disentangle the different possible causes – and their interplay – requires in depth diachronical, comparative, typological and sociolinguistic research which goes far beyond the scope of an initial descriptive sketch of the language. The complexity and interplay of factors to be considered in investigating causes and mechanisms of (contact-induced) language change is comprehensively discussed in the literature on language contact and careful examination of the evolution of (morphosyntactic) features in the dichotomy between inheritance and contact with Turkish has been already presented in earlier work on Romeyka (Sitaridou 2014b, 2016; Neocleous & Sitaridou 2022; Michelioudakis & Sitaridou 2020 with a focus on AMG) and also on Cappadocian (Karatsareas 2011, 2013, 2014; dating back till Winford 2005 and Thomason & Kaufmann 1988). Within the scope of the present study, a chapter on language contact can be only an attempt to sketch down some initial observations on the grammatical structures of Romeyka departing from a first glance at the existing grammatical situation of a certain dataset of bilingual ROF speakers and intended as a mere starting point for further in-depth investigation.

In this chapter, the most striking morphosyntactic features of Romeyka are presented, including some notes on the composition of the lexicon and on pragmatic-functional aspects of language contact in bilingual speakers, namely code-alternations, in Section 6.1. The striking morphosyntactic features of Romeyka are grouped together in this section according to their assumed evolution and the primary driving force behind it: namely, language contact (Section 6.2), inheritance and language internal change (Section 6.3) and shift-induced change (Section 6.4). Only further in-depth research can reveal how the presented candidate features evolved exactly.

### 6.1 Pragmatic aspects of language contact: code-switching vs. lexical borrowing

This section intends to briefly shed light on the amount of lexical borrowing in Romeyka (including “matter borrowing” in Matras’ 2009 terminology). However, since it is by no means straightforward to distinguish between more stable lexical borrowings and more spontaneous pragmatically driven instances of code-alternations (involving different layers at word, phrase, and clause level; see, e.g., Poplack 1993: 279; Gardner-Chloros 1995; Matras 2012a: 23–24; i.a.), the present section presents both types of lexical transfer as two ends of the same continuum:

In Matras (2009: 110–111) I define the difference between ‘borrowing’ and ‘codeswitching/codemixing’ as a continuum. At the far end, on the ‘borrowing’ side, we find regular occurrence of single words (rather than phrases), often with grammatical function or representing specific referents (such as names of institutions or other cultural terms), which are integrated morphologically into the recipient language. (Matras 2012b: 381).

The amount of lexical borrowing in Romeyka is confined here for the sake of feasibility within the scope of this thesis to Turkish/Ottoman lexical influence, and contacts via this pathway, i.e., French, Arabic, Persian, etc. loan words that have entered the Romeyka lexicon via Turkish or Ottoman, e.g., *şoför* ‘driver’ and also notable loan translations like *jer-almasia* (< Tr. *yer elması* ~ *patates* ‘potato’) ‘potato’ vs. *kartofî* ‘potato’. For other lexical contact influences - which are



certain - further research is needed, e.g., Armenian influences on the lexicon like *axpini* ‘animal manure’ from Armenian *agpn*, used in the Çaykara Turkish vernacular (Tursun 2019: 197). Furthermore, it should be noted that some speakers who are competent in SMG (or elderly stages of Greek) tend to introduce words from other Greek varieties or diachronic stages into Romeyka, which may enhance the lexicon of Romeyka but also may lead to levelling of Romeyka with other (mostly standardized/formal) Greek varieties. For this reason, data from speakers with no competence in SMG is preferred in the present thesis and data from speaker with SMG-competence is treated with caution.

As for the spontaneous, pragmatic end of the continuum of lexical insertion, code-switching and other forms of code-alternation are regarded here as mechanisms applied by the speakers in order to exploit their full linguistic repertoires, and be it to balance certain effects of attrition. According to Matras’ (2012) “activity-oriented approach”, contact-induced changes arise in multilingual speakers based on their intention to exploit their full linguistic resources within a specific pragmatic setting to reach certain communicative goals. The analysis of this pragmatic goal as the motivation of speakers to introduce “innovation” into their linguistic system(s) is regarded here as the key to understand the contact-induced change (although different levels of consciousness apply). In the words of Matras (2012: 48), “[c]ontact-induced change is the product of the creativity of speakers seeking new ways to achieve goal-oriented tasks in communicative interaction” (Matras 2012a: 48). This view adds to the understanding of language contact the communicative function of language and allows to explain more spontaneous pragmatic forms of contact as well as structural changes. The focus on the speakers’ pragmatic intentions in their communicative act can also explain the frequent spontaneous insertion of Turkish elements in the present fieldwork situation, which may differ from the amount of spontaneous Turkish interferences in other, Romeyka-dominant communicative settings (see also Section 1.5.3).

### 6.1.1 Lexical borrowing

Lexical borrowings in Romeyka are frequent among both content and function words. Between 10–30% of the nouns in the present Romeyka corpus are loans. Certain lexical items are retained from Ottoman rather than from post-language reform Turkish, e.g., *mektep* ‘school’ instead of Tr. *okul* ‘school’, *mualimina* ‘teacher.F’ vs. *öğretmen* (but also *öretmenis*), *vejaxut(e)* ‘or’ instead of *veja* ‘or’. Nouns denoting technology or concepts/things not in first line associated with traditional village life and work cycles are borrowed from Turkish, e.g., *televizjoni* ‘television’, *tšami* ‘windowpane’, *tufedž* ‘gun’, as are abstract concepts derived in Turkish with the suffix *-lık*, e.g., *kuimaxluk* ‘ingredients for kuymak’, cultural practices like *düyun* ‘wedding’ and some more remote kinship terms like *tejze* ‘aunt’ (see also Tursun’s (2019) dictionary for Turkish loan words, often including derivations). Anthroponyms in Romeyka are generally Turkish/Muslim.

Almost half of the adjectives in the present Romeyka corpus are borrowed, especially predicatively used adjectives but also attributively used ones, despite a relatively limited class of Greek adjectives, e.g., *trano* ‘big’, *kalo* ‘good’, *vari* ‘heavy’, *mikro* ‘small’, *makri* ‘long, far’, *emorfo* ‘nice’ vs. *jengi* ‘new’, *gentš* ‘young’, *ixtijari* ‘old’, *temiz* ‘clean’, *zengin* ‘rich’. Except for the colours ‘black’, ‘red’ and ‘white’, all other colour terms were in elicitation provided in Turkish (this is in line with Berlin & Kay’s 1969 expectations regarding the hierarchy of basic colour terms).

Adverbs of degree, spatial adverbs and most manner adverbs are inherited. Apart from a basic set of Greek temporal adverbs (e.g., *opse/extes* ‘yesterday’, *ifeti* ‘this year’, but *sabale* ‘tomorrow’), especially times of the day (incl. prayer times), e.g., *sabaxtan* ‘in the morning’, all temporal counts including numbers, e.g., *beš alti ai* ‘five, six months’ (08\_04072019M\_3; 86) and composed temporal adverbs are borrowed from Turkish, e.g., *eskiden beri* ‘since long

time’, *ondan sonra* ‘thereafter’, sometimes with a Greek equivalent existing as well, e.g., *olon embron* ‘before’. It is not clear what determines the choice of a Romeyka or Turkish equivalent even within the same speaker, e.g., *so jan=is* ‘to your side’ (03\_30092019F\_7; 51) vs. *s=emena džega* ‘to my side’ (03\_30092019F\_7; 50).

Many Turkish verbs are integrated into Romeyka, including many complex verbs including aspectual information, e.g., *ješ-lan-evo* ‘grow old’. For some frequent verbs also a double set in Romeyka and Turkish is available, e.g., *konuševo* ‘speak’ (vs. *lalo* ‘speak’), *tšališevo* ‘work’ (vs. *đulevo* ‘work’). Many verbs denoting abstract meanings or processes are probably only to be expressed by means of integrated Turkish loans, e.g., *jašamak* ‘live’. It seems that - as well as in nouns - that language shift becomes visible on an idiolectal basis regarding how many inherited words are still known. Furthermore, Turkish verb semantics seem to coincide often with those of equivalent Romeyka words, for example, the collocations of the verb *tero* ‘look’ seem to mirror that of Turkish *bakmak* ‘look’, for example, in the meaning of ‘look after (children, animals, ...)’. Finally, light verb constructions of the Turkish type are calqued on Romeyka verbs, e.g., Tr. *telefon etmek* ‘to call’, lit. ‘do telephone’ vs. *epitšen tšelefoni* ‘she called’, lit. ‘she made telephone’ (07\_04072019F\_5; 15).

All numerals above five are Turkish which is a trait often reported from intense unbalanced language contact (Matras 2011).

Many function words are borrowed from Turkish, for example particles like the question particle *mi*, the negative answer particle *jok*, nearly all discourse particles/discourse connectors, and some interrogative particles like *kaš kiši* ‘how many’ (Sitaridou 2014b). Nearly all conjunctions except for coordinative ones are borrowed, i.e., disjunctive, adversative and correlative conjunctions (see the borrowing hierarchy of conjunctions, Matras 2011). Furthermore, some complementizers/connectives are borrowed from Turkish, especially that of cause/reason, i.e., *tšünkü* ‘because’. All interjections are basically Turkish, e.g., *haidi* ‘Come!’, *vallahi* ‘I swear!’, some are typical for the area, which means they also exist in Trabzon Turkish, e.g., *he he* ‘yes, yes’.

The wide-spread borrowing of functional items in Romeyka seems largely in line with the ERIC type of loan words (Brian Josephs) involving especially those lexical items that are essentially rooted in conversation (i.e., discourse particles, i.a.). Interestingly, nearly an identical set of lexical items are borrowed in Cappadocian (Melissaropoulou & Ralli 2019: 715–717, Table 1), although it seems that the amount of lexical borrowing in Cappadocian exceeds that of borrowing in Romeyka. In sum, it appears that those lexical items related to the discourse/pragmatics interface are extensively borrowed from Turkish (also in other (East) Anatolian languages, Haig 2001: 206), potentially indicating that – dependent on the multilingual profile of the speaker – the mental frame/discourse setting for many speakers is Turkish with the Romeyka mental lexicon embedded. Importantly, this dominance of the Turkish mental frame might be also facilitated by the Turkish-dominant fieldwork setting, rather than necessarily indicating attrition.

### 6.1.2 Code-switching

The present Romeyka corpus contains many instances of code alternations, including code-mixing, code-switching, code-shifting (terminology used here in the general contact linguistic sense, for example as summarized in Auer 2011). Despite extensive literature on forms, functions and grammatical theory of code-switching, only few preliminary observations can be presented here. The function and grammar of code-switching in Romeyka awaits further research.

In the present corpus, five types of code alternations are attested: (i) intra-sentential code-switching of Turkish elements in Romeyka matrix code, (ii) inter-sentential code-switching of Turkish elements in Romeyka matrix code; (iii) intra-sentential code-switching of Romeyka

elements in a Turkish matrix code; (iv) code-mixing with no dominant code; and (v) code-shifting from a Turkish-dominant code to Romeyka (for a definition of “matrix language”, see Section 1.5.2.2). Importantly, the existence and amount of code-switching depends on the speaker and the communicative setting. The fieldwork situation in Turkish played an important role here since speakers would tend more easily to switch to Turkish to elaborate on utterances or interact with the fieldworker(s). Even if recordings were elicited by community members without the presence of the fieldworker, Turkish was the standard means of communication. Analysis of more natural data in a communicative setting where Romeyka is the dominant language would possibly reveal different instances of code-switching. The forms of code alternations attested in the present corpus are briefly summarized below.

(i) Intra-sentential code-switching with the insertion of Turkish elements in Romeyka as matrix language occurs occasionally when a Romeyka word, predominantly a noun, for a certain concept is missing (1), although it is difficult to distinguish between code-switching and lexical borrowing here. Furthermore, fully inflected Turkish verbs showing subject agreement can be inserted (2), in which case, though, Turkish should be considered the matrix language of the clause as the crucial systemic inflectional morphology in the clause is Turkish. Adverbial clauses, especially temporal (3), but also spatial (4), among others (5), are frequently inserted in Turkish. This applies especially to the Turkish temporal adverbial *ondan sonra* ‘thereafter’, which occurs frequently as a filler in elicited Romeyka texts about habitual actions of everyday life. Moreover, within the NP Turkish genitives are inserted in a number of cases (6–8). Turkish *var*-existentials are also prone to code-switching (9),<sup>248</sup> as are Turkish predicatives, both indicative (10) and negated with Tr. *değil* (11). Finally, Turkish subordinating formulations can be inserted (12).

- (1) *bazi hajes epiname ... arabuluculuk*  
‘sometimes we made it like this ... matchmaking’ (04\_01072019F\_2; 172–173)
- (2) *me ti mana=s taniştım esi etrižes ežis*  
‘I met your mother, you are running there.’ (01\_07072019F\_1; 03)
- (3) *istedejim zaman do pükniği n=eftayo*  
‘I make a picnic when I want.’ (08\_04072019M\_1; 261)
- (4) *her jere ferume=sas*  
‘I (will) bring you to all places.’ (08\_04072019M\_4; 34)
- (5) *jane šu mahalle ait ebigam=a*  
‘We made it belonging to this neighbourhood.’ (08\_04072019M\_3; 147–149)
- (6) *do ba d=efte detšinuz-un ešekluxi*  
‘Why did you do their stupidity?’ (04\_01072019F\_2; 211)
- (7) *jane herkes seve seve buranin fasuljesini eber*  
‘Well, everybody likes to buy beans from here.’ (08\_04072019M\_1; 136)

<sup>248</sup> Interestingly, the existential is not the standard Turkish *var*-construction, i.e., Tr. *arkadaşı-mız vardı* ‘we had friends’), but a re-structuring of the Turkish predicative possession along the lines of Romeyka ‘have’-possession which also (partially) resembles predicative possession of the form *var* + POSS in other Turkic languages, e.g., Iran-Turkic.

- (8) *jane **dünja-nin** olon tehlikeli viradža ine tši*  
 ‘The world’s most dangerous road bends are there.’ (08\_04072019M\_2; 084–086)
- (9) *emisdine tše ta xarðela **arkadaši vaiduk***  
 ‘We had friends among the children.’ (04\_01072019F\_2; 028)
- (10) *aðaha so memleket emis **dört divar arasindajuk***  
 ‘Here in our homeland, we are between four walls.’ (01\_28062019F\_3; 38–40)
- (11) *iba esi medžburen na pas eyo **medžbur deulum** m=ila so banemo*  
 ‘I said, “You have to go. I am not forced, I don’t have to go.”’ (03\_30062019F\_6; 06)
- (12) *don dünja=muna ai eðevaksame **sen de ki** dib=udž erlaepsa aso genšlux=im*  
 ‘We spent our world like this. You can say, I have nothing understood from my youth.’  
 (02\_02022015F\_1; 011–012)

(ii) A prominent example of inter-sentential code-switching of Turkish elements in a Romeyka matrix is the insertion of Turkish sayings (08\_04072019M\_1; 050; also 04\_01072019F\_2; 327) or formulaic expressions, e.g., Tr. *o günler kahrolsun* ‘down with these days’ (07\_04072019F\_6; 45–47); for instances of insertion of full Turkish clauses whereby no code seem to be dominant, see (iv) below. The fact that for the speakers of the present corpus sayings are usually Turkish clearly indicates a shift to Turkish as dominant language.

Furthermore, insertion of Turkish clauses seems to appear for the purpose of emphasis (13, 14, also 02\_2906019F\_1; 28). Romeyka clauses may be repeated by Turkish phrases as in the question in (15).

- (13) *batsi-ka=tes ako mikro en akome mikro en **dajanamaz** bori=nata jani*  
 ‘Her little daughter is still small, still small, she does not resist, she cannot.’  
 (07\_04072019F\_5; 25–28)
- (14) *aðahandžeka tip=u=borum n=eberume **para ile jok** met o ðromo na drežis*  
 ‘I cannot by anything around here, not with money, you run with the road.’  
 (01\_28062019F\_3; 55–57)
- (15) *ler mi i mana=s **anlater** mi*  
 ‘Does your mother say this? Does she tell (this)?’ (07\_04072019F\_5; 51)

(iii) Occasionally, code-switching occurs also in the other direction, i.e., with the insertion of Romeyka elements in a Turkish string, such as with the pronominal elements in (16).

- (16) ***ne** abu rumdža verijo **ade ula** aluyr da*  
 ‘Yes, she (lit. ‘this one’) gives Rumca; she records it.’ (04\_01072019F\_5; 64)

(iv) Code-mixing is classified here as the alternation of full clauses in Romeyka and Turkish with no visible dominant code, although in general, Romeyka is the target language in the recordings. Turkish clauses are inserted occasionally to clarify a point (17) although this may be initiated by the struggle to express a certain thing adequately in Romeyka (18).

- (17) *hitš emena utš emnazen etšinus.. etšinus **bir baška adam idi***  
 ‘He didn’t resemble me at all, he was a different man.’ (03\_30062019F\_8; 08)

- (18) *ade so tšedžuk oiret.. šej o juvasinda tšedžukler bakaidiler o o kadara adamlari var idi*  
*o is don ađelfo=m ebare o jalo din ekraba=m ebare*  
 ‘She worked at the kindergarten, say, at this kindergarten they cared for the children.  
 There were so many men: One person [said], “Take my brother!”, the other, “Take my  
 relative!”’ (01\_04022016F\_1; 107–111)

(v) Code-shifting from Turkish to Romeyka occurs either when a previous string of speech has been in Turkish and the conversation is to be continued in Romeyka or when the speaker accidentally starts a clause in Turkish but realizes that Romeyka is the target code. In the first case, the transition may take place intra-sententially as with the subordinate clause in (19) or by (partly) repetition of the Turkish clause into Romeyka (20) or inter-sententially (21), whereby a Turkish form is inserted in the otherwise Romeyka string. In the second case, clauses start often with overt Turkish pronouns which are immediately corrected by the Romeyka form (22, 23, also 02\_02022015F\_1; 180), but also full clauses slip out in Turkish (24). The second phenomenon pictures clearly the disturbing impact of Turkish as the means of communication in the fieldwork setting on the naturality of Romeyka speech, although the accidental shift to Turkish is not limited to the begin of sentences; in the case of (25), it is critically commented on this by the speaker.

- (19) *[...] ama kene bir gin gelur leyune bal kadi leyune*  
 ‘But again a day comes, they will say something.’ (03\_30062019F\_7; 43–44)
- (20) *jane benim aradžim o idi kamjonum o idi, etšino done*  
 ‘Well, this was my vehicle, this was my truck.’ (08\_04072019M\_2; 073)
- (21) *ište eb ordan ebedžeha etšeli edžes*  
 ‘See, from there, from there, he rolled over there.’ (03\_30062019F\_8; 22–23)
- (22) *biz.. eh.. emist ae konušefkumuneste*  
 ‘We used to speak like this.’ (02\_02022015F\_1; 115)
- (23) *belki siz e.. tesetera ta tahadžes mi inane haje bola š[imu]*  
 ‘Maybe you... Did you also have such a heavy rain on your side?’ (03\_30062019F\_6; 49)
- (24) *nenekalar var e.. esane i mamikandi [...]*  
 ‘There are grandmothers..., eh.. there were grandmothers [...].’ (04\_01072019F\_2; 184)
- (25) *i mana=s asin d=ađelfe=s asin tšok sever.. bola ayapena=natin ha janliš dedim*  
 ‘I loved your<sup>2</sup> mother, your<sup>2</sup> siblings much, I said it wrongly.’ (01\_07072019F\_1; 12)

In sum, based on the examples of code-switching provided above it is evident that code alternations between Romeyka and Turkish occur frequently and can be viewed as manifestation of the speakers’ desire to exploit their full linguistic resources in a certain communicative setting (in the sense of Matras’ (2012) “activity-oriented approach”; the issue of whether these code alternations are conscious is not taken up here, and neither is “playful or ‘theatrical’ mixing”, Matras (2012a: 23)). Notably, the bilingual profile of the speaker seems to play an important role for the intensity of code alternations: Beyond certain pragmatic expressions like sayings and other lexicalized phrases which are nearly always borrowed from Turkish, and which are probably no longer available in the Romeyka lexicon, many expressions seem to exist as a “double set” in Turkish and Romeyka. It is assumed that in a Turkish-

dominant speaker, the occurrence of Turkish alternations is more likely, although the concept of “dominant language” itself is not unproblematic. In the case of Romeyka, it seems that all speakers are to a certain degree Turkish-dominant dependent upon different domains of use and based on very different acquisition biographies, whereby it is difficult to define a cut-off point between Turkish- vs. Romeyka-dominance. However, it needs to be noted that the fact that the field work setting of the present study took place in Turkish probably had an impact upon the frequency of Turkish alternations. As the occurrence of both codes in a conversation is pragmatically determined, in Romeyka-dominant communicative settings Turkish alternations are expected to occur much fewer, although further research is in order here.

## 6.2 Contact-induced grammatical change

The research on contact-induced language change deals with the examination of processes and results of linguistic change caused primarily by contact with (an)other language(s). The definition of such a change can be relatively broad: “[...] a particular linguistic change is caused at least in part by language (or dialect) contact if it would have been less likely to occur outside a particular contact situation” (Thomason 2007: 42). Poplack & Levey (2010) present in their “comparative variationist approach” a basic methodology to determine the likelihood of contact influence (see Thomason 2001: 93–94; and Sakel & Matras 2008: 80 for alternative methods): In order to establish that language contact is the cause of a certain linguistic structure, it has to be proven that the structure in doubt (a) is existent (and of some age, i.e., no innovation) in the contact variety; but (b) not existent in any former stage of the recipient language; (c) nor existent in any related language not in contact with the source variety. Thus, not only comparison with the contact variety is crucial but also comparison with diachronic stages of the language in question and with its genetic relatives outside the contact area (for an example of such a study see Schreiber, Haig, Taheri-Ardali & Anonby 2021). Still, such a scrutiny does not rule out the possibility of multiple causation which needs to be borne in mind. An additional factor in examining contact-induced change is to consider the general typological likelihood of a change. According to Haig (2014), not all potential candidate features for contact influence have the same weight with regard to assessing the intensity of language contact: in general, typologically rare and marked features evidence stronger contact as they are less easily targeted by language contact. Also, the genetic relation of the languages in contact plays a role in that contact-induced changes are more likely to occur (and more likely to intensively occur) in closer related languages. In this regard, the contact setting of Indo-European Romeyka with more “Asian-type” Turkic (Haig & Khan 2018) is an interesting constellation.

Although according to Haig (2014: 16) “[t]he prime task in any investigation of language contact is the identification of shared features, followed by a rigorous assessment of their probable genesis”, in the following only some tentative candidates can be presented for contact influence of Turkish on Romeyka and even if only a closer scrutiny can reveal the exact nature of the indicated changes, the option of multiple causation should be always borne in mind.

In compiling a list of candidate features for contact-induced change from Turkish, it is focused here on the most striking phenomena; many smaller candidates for contact-induced change are mentioned in the grammar text, e.g., the similarity of Tr. *bol* with Romeyka *bola* in the function of ‘much’ (Fn. 46, Section 3.1.4.1). Although influences of Turkish on Romeyka are at the focus of this section, contact-induced changes may have occurred/occur also in the other direction and often it is not straightforward to indicate the model for a certain shared construction, e.g., “areal” demonstrative/deictic *ha-* like in Tr. *ha-bu*/ R. *h-avudos* ‘this’; also verb serialization of the same type in Romeyka and Turkish, e.g., Tr. *hep rumdža konusurduk giderduk* ‘We always spoke Romeyka.’ (04\_01072019F\_2; 003).

The potential candidates for contact-induced change from Turkish presented here will be predominantly from the domain of morphosyntax; in the phonology, Romeyka and the Trabzon

Turkish dialect have clearly influenced each other as has been comprehensively described for Trabzon Turkish by Brendemoen (2002, 2006). A detailed investigation of the influence of Trabzon Turkish on Romeyka phonology is beyond the scope of this work and remains a desideratum for further study, although there are some striking features like the kinds of distant vowel assimilation outlined in Section 2.3.2. An overview of the parts-of-speech that are often borrowed from Turkish into Romeyka are presented in Section 6.2. above.

The following morphosyntactic features are candidates for contact-induced change from Turkish, presented in the order of evidential substance starting with the most likely candidates:

- i. borrowing of the Turkish interrogative particle *mI* albeit without vowel harmony
- ii. reduplication of manner adverbs, e.g., *xorja xorja* ‘separately’, *yale yale* ‘slowly’ (< Tr. *yavaş yavaş*) (Tursun 2019: 234)
- iii. similar functions of the particles *ba(l)* and *tše* and Tr. *dA*; an enclitic “clitic additive or ‘Recalled Topic’ marker” is an Anatolian feature (Haig 2017: 407)
- iv. two types of progressive constructions (a) auxiliary *eš(i)* + finite verb, (b) finite verb + *tše* ‘and’ + finite verb *kahume* ‘sit’ or *steko* ‘stay’, whereby the latter is probably modelled on the Turkish continuative construction *-(y)Ip + durmak* ‘stay’
- v. serial verb constructions (partly coordinated with *tše* ‘and’) probably modelled on the Tr. converb *-(y)Ip*; especially, double verb constructions like ‘go and return’
- vi. increase in deverbal nouns/nominalizations based on Tr. short infinitive nominalizations
- vii. productive non-finite complementation strategies under contact with Turkish, e.g., with volitionals: finite subjunctive modal complements after ‘want’ are an Anatolian areal feature (Haig 2017: 414–416); potentially also lack of complementizers in complements of verbs of saying and (in)direct speech
- viii. Turkish-type adjunct clauses of purpose/reason (with Tr. *diye*, *çünkü*)
- ix. clause combination by means of Tr. conjunctions *ama* ‘but’, *hem* ‘as well ... as’, *ne* ‘neither .... nor’, *ve/yoksa* ‘or’ (in line with other Anatolian languages, Haig 2001: 206)
- x. pre-posed temporal adjunct clauses with *anda* resemble word order patterns of Turkish temporal adjuncts with *-DIğI*
- xi. potential changes in frequency: contact-induced higher frequency of full object pronouns instead of clitic pronouns possibly due to Turkish influence; although enclitic object pronouns on verbs are a Western Asia feature (Haig & Khan 2018: 18)
- xii. variation in existentials and predicative possession with *exo* ‘have’ and *ime* ‘be’ caused by model of Tr. *var*-existentials
- xiii. prenominal relative clauses on the Turkish model (although Neocleous 2020 attests also inherited postnominal RCs with a certain frequency; on head-final nominal syntax in AMG, see Michelioudakis & Sitaridou (2020: 258), who argue that negative evidence from Turkish has supported prenominal NP orders)
- xiv. tendency to verb-final copula clauses in Romeyka with deletion of clause-final 3<sup>rd</sup> person singular copula *en* in predicatives; an enclitic copula forming a pan-Anatolian feature (Matras 2009: 270, in Haig 2017: 404–405; Haig & Khan 2018: 18–21); according to Neocleous & Sitaridou (2022) final auxiliaries are indeed contact-induced, but word order drift (cf. some varieties of Cappadocian) has not taken place
- xv. potential shift in word order directionality from SVO (Hellenic) to OV (Turkic) (see Schreiber & Janse, in preparation): OV as default word order in subordinate clauses (Neocleous 2020), but also in copula-complements and ultimately, as the most frequent pattern in declarative clauses due to information structure; (cf. Haig & Khan 2018: 26 for a provisional order of the ease of syntactic word order borrowing: copulas > goal verb order > direct object verb order > adpositional order; esp. post-predicative goals are a candidate for an areal feature, Haig 2017: 408–412)

### 6.3 Language internal change and inheritance

Apart from the more or less obvious cases of Turkish interference and the striking individual differences in language competence and amount of Turkish influence between speakers, there are many structural features in Romeyka which are clearly of Greek descent – not few of them being archaisms preserved from older stages of Greek. When it comes to changes in the grammatical system that are not obviously contact-induced, it is often not easy to discern whether these changes are “internally” motivated, thus following a diachronic change that might be also visible in related varieties outside the contact setting, or whether these changes may point at attrition. While the huge amount of variation in Romeyka may probably point more toward the latter (see Section 6.5), within all this variation, there is a certain core of more or less stable grammatical features that are clearly of Greek inheritance.

Structural features that are likely inherited:

- i. determiner spreading
- ii. diminution
- iii. double set of demonstrative pronouns
- iv. ancient relativizers *pe/pu*
- v. the amount of variation in the pronominal endings reflects probably older stages and internal change
- vi. possession with prenominal genitives is largely intact (no periphrastic PPs, genitive not substituted by accusative)
- vii. AG quantifier like *alos* ‘all’ and *edero* ‘other’ are retained in their ancient distribution
- viii. complex spatial relators
- ix. the tense system (e.g., aorist and imperfective endings)
- x. archaic imperatives
- xi. the wide functions of the modal particle *na* are an internal development and are similar in other modern Greek dialects
- xii. remnants of the infinitive
- xiii. variation in negators
- xiv. comparatives are inherited though possibly undergoing internal simplification (other than in contact literature where comparatives are prone to borrowing; comparatives marking the standard of comparison by a ‘source’ like Romeyka *aso* ‘from’ are an Anatolian feature, Haig 2017: 403–404); but on the other hand limited use of superlatives
- xv. limited use of object clitic doubling (compared to other modern Greek dialects)
- xvi. verb serialization without coordinating conjunction (juxtaposition)

The following striking phenomena (including variation and change) in Romeyka are yet to be explained, although an (at least partly) internal explanation is not unlikely:

- i. immense variation in long and shorter forms of personal and object pronouns, including a possible status of post-verbal pronouns as clitics
- ii. reduction in the nominal inflection of articles, nouns, adjectives, and others like numerals: syncretism in gender inflection (also affecting number and case); instability in some plural forms, e.g., two plural forms like *arko-ðes* vs. *artš-i* ‘bears’
- iii. variation in number of 3<sup>rd</sup> person copula *en/ine* ‘be’, e.g., *i dri en t ospiti i arθob*, ‘the three are the people of the house’ (05\_03072019M\_4; 02)
- iv. progressive aspect with *eš* ‘have.3SG’ and development of *exo* ‘have’ as auxiliary; also, the use of *ixa* ‘had.1SG’ in counterfactuals
- v. reduced verbal morphology like in morphological subjunctives, pluperfect, etc.



- vi. the lack of reflexives and non-periphrastic reciprocals is probably inherited
- vii. few information on voice, passives, causatives due to lack of data or general reduction?
- viii. double set of discourse connectives *ba(l)* and *tše* ‘and’, both probably influenced by Turkish *dA* (see Section 6.3)
- ix. function of *boro* ‘can/cannot’ in potentials as negative polarity item (NPI) (Sitaridou 2014b)
- x. complementation strategies with null-complementizers in verbs of saying and (in)direct speech: internal reduction or contact-induced change?
- xi. demonstrative prefix *h(a)-* and suffix *-ha*

#### 6.4 Language shift-induced change

The grammar of dying languages has become identified with certain structural changes that are common in a situation of language shift: (a) loss or reduction, (b) simplification, regularization, levelling, (c) retention of complexity or innovation, elaboration, (e) interference (based on Tsunoda 2005: 100).<sup>249</sup> In Romeyka, the following features could potentially point towards language attrition induced by an increasing lack of language use:

Personal pronouns are sometimes confused with object (26) and possessive pronouns (27). This is especially frequent in nominalizations of the Turkish type, e.g., *eyo/temo t=erθinimon* ‘my coming’, and in predicative possession modelled on Turkish existentials of the type free possessive + *var* ‘exist’ (Section 5.2.1.5.5). While the confusion of subject with object pronouns might be a sign of attrition, the confusion of possessive pronouns with personal pronouns in have-clauses is likely contact-induced by the Turkish existential construction, though. It needs also to be noted, that confusion of personal and object pronouns is not necessarily a sign of attrition as it takes place in many spoken varieties, among them as vital European languages as English.

(26) *har esena tš=eyo na sindišenum*  
 ‘Now, you and me will talk.’ (04\_01072019F\_5; 06)

(27) *eyo to ḏa.. to šeri=m bola bonise*  
 ‘My finger.. my hand hurt much.’ (02\_2906019F\_1; 29)

A potential sign of simplification could be zero-marked complement clauses as frequent embedding strategy, esp. utterance complements vs. direct speech (Section 5.3.2.1.2), although in this case, Turkish is a model as well.

Furthermore, the gender-sensitive article system in Romeyka is collapsing in various ways, as is the related system of nominal declension and adjective agreement. The reanalysis of a masculine noun as neuter analysing the masculine article as part of the word, e.g., *do o-fengos ekseve* ‘the moon rose’ (04\_01072019F\_12; 39), is just one example. However, as has been pointed out by Karatsareas (2011) for Cappadocian, there are long-term diachronic inherited changes going on that should not be easily classified as attrition; a very nuanced scrutiny is required here to explain these changes in the nominal inflectional system.

Finally, the high amount of inter-speaker variation in all grammatical domains (phonology, morphology, syntax) is clearly a sign of attrition since the different bilingual profiles of speakers clearly show different forms of language competence and attrition. However, in addition to this, the high amount of intra-speaker variation is also possibly pointing at an instability in the grammatical system.

<sup>249</sup> Although recent research has questioned whether “simplification” should be expected a normal outcome in language contact (Kantarovich et al. 2021).

## 7 Conclusion

Despite the pending steps in language description and documentation (Section 1.5.5), this work is: (i) the first comprehensive grammatical description of the endangered indigenous minority language Romeyka as spoken at present in northern Turkey; (ii) a compilation of a fully annotated naturalistic spoken language corpus of a previously under-described endangered language as basis for language description and documentation, and intended as an accessible resource for further scholarly work with theoretical interest in typology, language contact, and ultimately, Greek dialectology – although more field-specific informed work is needed here (see Section 1.5.5); (iii) a synchronic record of an ongoing language shift of a minority language in contact with the nation's majority language, including some preliminary theoretical thoughts on the grammar of language shift. Finally, the present study is intended as a resource for interested speakers; as such it has been attempted to document the language by presenting as many original data as possible in the way the language is actually spoken, i.e., including the full amount of linguistic variation, including the contact influences and aspects of pragmatic language use such as code-switching.

In addition to the grammatical description which is understood to form the basis for any further theoretical investigation, the present work has tried to shed some light on the present grammatical status of Romeyka as (still) spoken in Northern Turkey represented in three research questions:

### I. How strong is the Turkish influence on Romeyka?

It has been found that there is certainly Turkish influence on all grammatical domains of Romeyka, although the present work focused on morphosyntactic features, together with a glimpse at the lexicon. In the lexicon, the Turkish influences are probably comparable to that of Turkish in other Asia Minor Greek varieties such as Cappadocian, although they seem not to be as far reaching as in Cappadocian, which probably indicates a different contact situation with less intense contacts. On the structural domain, the Turkish influence is often in line with what the contact linguistic literature describes as “vulnerable” features like, for example, the borrowing of the Turkish question particle. However, the intensity of Turkish language contact seems to vary according to the bilingual repertoire of the individual speaker and for some structural features more than others: word order appears to be highly susceptible to intra-speaker variation (see Schreiber 2022) and there is certainly more to say about word order in Romeyka in the Western Asian contact zone (Haig & Khan 2018; Schreiber & Janse, in preparation).

### II. Does language shift impact upon the structures of Romeyka?

The structural effects of language shift that are typically associated with language decay and attrition, namely simplification, reduction or loss, are partially visible in Romeyka: traits of regularization exist for some morphosyntactic features like the verbal tense-aspect system and especially in the nominal inflectional morphology. However, these changes cannot be clearly attributed to language shift as they seem to go often hand in hand with language internal changes (potentially furthered by contact with Turkish) that are also reported for other genealogically related varieties outside the immediate contact setting (although at a certain point in time also in contact with Turkish, like Cappadocian). More detailed research on the intertwining of different causations is required. The effects of language shift become indirectly visible at a less consolidated linguistic domain, namely in the individual multilingual repertoires of the speakers, i.e., in the different amounts of more spontaneous, pragmatically-driven bilingual interaction.

### III. How “Greek” is Romeyka in its inherited grammatical structures?

In the light of the list of morphosyntactic candidate features for language inheritance and internal change in Section 6.3, Romeyka can structurally be clearly considered a Greek variety, although considerable influence from Turkish is visible and many structural idiosyncrasies exist that set Romeyka, as member of the Pontic Greek branch, apart from other Modern Greek dialects, not least because of the preservation of many archaic features. However, more informed Greek dialectological research is needed to define the classification of Romeyka within the Hellenic language family and, ultimately, to answer the language vs. dialect discussion based on a profound structural analysis and comparison.

In sum, the central impression of the present thesis is that in this particular scenario of an ongoing language shift, language “decay” does not happen so much at a structural level in all speakers at the same time, but rather by means of immense individual variation regarding language competences and use patterns that shape individual processes of language shift. The individual differences in the multilingual repertoires of speakers appear to affect (a) the amount of Turkish structural borrowings, i.e., the occurrence of a certain contact-induced variant, e.g., the occurrence of the Turkish question particle, (b) the amount of lexical influence from Turkish, and (c) the amount (and type) of spontaneous pragmatically-determined code-alternations. The “activity-oriented approach” by Matras (2012) has been found suitable to explain the actual bilingual behaviour of speakers in a certain communicative setting. Reviewing some types of Turkish structural influences in the present data, like, for example, nominalizations, it needs to be borne in mind that these Turkish influences may be facilitated by the research setting, i.e., either by elicitation by means of a translation task from Turkish and/or by the fact that Turkish was the communicative framework in the interviews with the semi-speaker. Ultimately, it is not clear yet to which extent a mixed acrolect may have formed in Romeyka based on the everyday pragmatic speech settings; here, further research is required as to how Romeyka is actually used in the speakers’ daily lives.

## References

- Alexiadou, Artemis. 2014. *Multiple determiners and the structure of DPs*. Amsterdam, Philadelphia: Benjamins.
- Alexiadou, Artemis. 1997. *Adverb placement. A case study in antisymmetric syntax*. Amsterdam, Philadelphia: Benjamins. <https://doi.org/10.1075/la.18>.
- Andrews, Peter A. 1989. *Ethnic groups in the Republic of Turkey*. Compiled and edited with the assistance of Rüdiger Benninghaus. Wiesbaden: Reichert.
- Andriotis, Nikolaos. 1974. *Lexikon der Archaismen in den neugriechischen Dialekten*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Asan, Ömer. 1996. *Pontos Kültürü*. [Pontus Culture]. Belge Yayınları.
- Auer, Peter. 2011. Code-switching/mixing. In Wodak, Ruth, Johnstone, Barbara, & Kerswill, Paul (eds.), *The SAGE handbook of sociolinguistics*, 460–478. London: SAGE Publications. DOI: 10.4135/9781446200957.
- Bağrıaçık, Metin. 2018. *Pharasiot Greek. Word order and clause structure*. Ghent: University of Ghent dissertation.
- Bağrıaçık, Metin & Mark Janse. 2016. Reduplication with fixed segmentism in Cappadocian and Armenian as a contact-induced phenomenon. In Ralli, Angelika (ed.), *Contact morphology in Modern Greek dialects*, 179–209. Cambridge: Cambridge Scholars Publishing.
- Bentein, Klaas & Metin Bağrıaçık. 2018. On the third type of relative clause in Post-classical and Early Byzantine Greek. *Transactions of the Philological Society* 116. 529–554.
- Berikashvili, Svetlana. 2017. *Morphological aspects of Pontic Greek spoken in Georgia*. Munich: LINCOM.
- Berlin, Brent & Paul Kay. 1969. *Basic color terms: their universality and evolution*. Berkeley, Los Angeles: The University of California Press.
- Bhat, D.N.S. 2013. Third Person Pronouns and Demonstratives. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/43> on Nov 15, 2022.
- Biliçi, Faruk. 2011. Que reste-t-il de la langue et de la culture grecques sur les côtes turques de la mer Noire? *Cahiers Balkaniques* 38-39. 325–342.
- Blevins, Juliette. 2008. Consonant epenthesis: natural and unnatural histories. In Jeff Good (ed.), *Linguistic Universals and Language Change*, 79–107. Oxford: Oxford University Press.
- Bortone, Pietro. 2009. Greek with no history, no standard, no models: Muslim Pontic Greek. In Alexandra Georgakopoulou & Silk, Michael (eds.), *Standard Languages and Language Standards: Greek. Past and Present*, 67–89. London: Ashgate.
- Brendemoen, Bernt. 2006. Aspects of Greek-Turkish language contact in Trabzon. In Hendrik Boeschoten & Johanson, Lars (eds.), *Turkic Languages in Contact*, 63–73. Wiesbaden: Harrassowitz.
- Brendemoen, Bernt. 2002. *The Turkish dialects of Trabzon: Their phonology and historical development*. Wiesbaden: Harrassowitz.
- Chatzopoulou, Katerina & Ioanna Sitaridou. 2014. Negator selection in Romeyka conditionals. Jespersen's cycle for NEG2 and conditional inversion. *Paper presented at the 16<sup>th</sup> Diachronic Generative Syntax conference (DIGS)*, Budapest, 4 July 2014.
- Chelliah, Shobhana L. & Willem J. de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht, Heidelberg, London, New York: Springer. DOI 10.1007/978-90-481-9026-3\_10.
- Chen, Sihwei. 2015. There are many... and more bears!. *Totem Field Storyboards*. Retrieved from <http://www.totemfieldstoryboards.org> on Nov 15, 2022.

- Cristofaro, Sonia. 2013a. Purpose Clauses. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/125> on 15 Nov 2022.
- Cristofaro, Sonia. 2013b. Utterance Complement Clauses. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/128>, on 15 Nov 2022.
- Cristofaro, Sonia. 2013c. 'When' Clauses. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/126> on 15 Nov 2022.
- Cristofaro, Sonia. 2013d. Reason Clauses. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/127> on 15 Nov 2022.
- Christofaro, Sonia. 2005. *Subordination*. Oxford: Oxford University Press.
- Comrie, Bernard. 1998. Rethinking the typology of relative clauses. *Language Design* 1. 59–86.
- Comrie, Bernard & Tania Kuteva. 2013. Relativization on Subjects. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/122> on 21 Mar 2022.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Dawkins, Richard M. 1937. The Pontic dialect of Modern Greek in Asia Minor and Russia. *Transactions of the Philological Society* 36. 15–52.
- Dawkins, Richard. M. 1931. Notes on the study of the Modern Greek of Pontos. *Byzantion* 6. 389–400.
- Dawkins, Richard M. 1916. *Modern Greek in Asia Minor. A study of the dialects of Silli, Cappadocia and Pharasa with grammar, texts, translations and glossary*. Cambridge: Cambridge University Press.
- Deffner, Michael. 1878. Die Infinitive in den pontischen Dialekten und die zusammengesetzten Zeiten im Neugriechischen. *Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 191–230. Berlin: Buchdruckerei der Königl. Akademie der Wissenschaften.
- Dobrushina, Nina, Johan van der Auwera & Valentin Goussev. 2013. The Optative. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/73> on 24 Apr 2022.
- Drettas, Georges. 1999. Το ελληνο-ποντιακό διαλεκτικό σύνολο [The Greek-Pontic dialect group]. In Anastasios-Foivos Christidis, Arapoglou, Maria & Giannouloupoulou, Giannoula (eds.), *Διαλεκτικοί θύλακοί της ελληνικής γλώσσας* [Dialect enclaves of the Greek language], 91–100. Athens: Ministry of National Education and Religious Affairs/Centre for the Greek Language.
- Drettas, Georges. 1997. *Aspects pontiques*. Paris: Association de Recherches Pluridisciplinaires.
- Dryer, Matthew S. 2013a. Order of Adjective and Noun. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/87> on 14 Jan 2021.

- Dryer, Matthew, S. 2013b. Definite Articles. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/37> on 26 Apr 2021.
- Dryer, Matthew S. 2013c. Position of Pronominal Possessive Affixes. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/57> on 05 May 2022.
- Dryer, Matthew S. 2013d. Position of Polar Question Particles. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/92> on 14 Feb 2022.
- Dryer, Matthew S. 2013e. Order of Adposition and Noun Phrase. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/85> 27 May 2022.
- Dryer, Matthew S. 2013f. Order of Genitive and Noun. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/86> on 07 Feb 2022.
- Dryer, Matthew S. 2013g. Order of Relative Clause and Noun. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/90> on 16 mar 2022.
- Dryer, Matthew S. 2013h. Relationship between the Order of Object and Verb and the Order of Relative Clause and Noun. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/96> on 16 March 2022.
- Dryer, Matthew S. 2013i. Relationship between the Order of Object and Verb and the Order of Adposition and Noun Phrase. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/95> on 15 Nov 2022.
- Dryer, Matthew S. 2013j. Relationship between the Order of Object and Verb and the Order of Adjective and Noun. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/97> on 15 Nov 2022.
- Dryer, Matthew S. 2013k. Order of Subject, Object and Verb. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/81> on 15 Nov 2022.
- Dryer, Matthew S. 2013l. Order of Object and Verb. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/83> on 15 Nov 2022.
- Dryer, Matthew S. 2013m. Position of Interrogative Phrases in Content Questions. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/93> on 15 Nov 2022.

- Dryer, Matthew S. 2013n. Order of Adverbial Subordinator and Clause. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/94> on 22 Mar 2022.
- Dryer, Matthew S. 2013o. Determining Dominant Word Order. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/s6> on 15 Nov 2022.
- Dryer, Matthew S. 1991. SVO Languages and the OV:VO Typology. *Journal of Linguistics* 27. 443–482.
- Dryer, Matthew S. & Orin D. Gensler. 2013. Order of Object, Oblique, and Verb. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/84> on 10 Dec 2021.
- Dwyer, Arianne M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Gippert, Jost, Himmelmann, Nikolaus P. & Mosel, Ulrike (eds.), *Essentials of Language Documentation*, 31–66. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110197730.31
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. <http://www.ethnologue.com> (accessed 30 Jan 2023).
- Emde Boas, Evert van, Albert Rijksbaron, Luuk Huitink & Mathieu de Bakker. 2019. *Cambridge Grammar of Classical Greek*. Cambridge: Cambridge University Press.
- Gandon, Ophélie. 2016. La relativisation dans une perspective aréale: l'aire Caucase - Anatolie de l'Est - Iran de l'Ouest. Paris: Université Sorbonne Nouvelle dissertation.
- Gardner-Chloros Penelope 1995. Code-switching in community, regional and national repertoires: the myth of the discreteness of linguistic systems. In Lesley Milroy & Muysken, Pieter (eds.), *One speaker, two languages. Cross-disciplinary perspectives on code-switching*, 68–90. Cambridge: Cambridge University Press.
- Genç, Tuğçe. 2021. *Trabzon örneğinde Rumca-Türkçe dil ilişkisi* [Rumca-Turkish language contact on the example of Trabzon]. Unpublished MA thesis. Ankara: Hacettepe Üniversitesi.
- Giannakou, Aretousa & Ioanna Sitaridou. 2020. Microparametric variation in the syntax of Spanish and Greek pronominal subjects. *Glossa: A journal of general linguistics* 5(1): 75, 1–42. DOI: <https://doi.org/10.5334/gjgl.960>.
- Gil, David. 2013a. Numeral Classifiers. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/55> on 03 Mar 2022.
- Gil, David. 2013b. Conjunctions and Universal Quantifiers. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/56> on 17 Nov 2022.
- Gil, David. 2013c. Genitives, Adjectives and Relative Clauses. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/60> on 16 Mar 2022.
- Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel. 2006. *Essentials of language documentation*. Berlin, New York: Mouton de Gruyter.
- Göksel, Aslı & Celia Kerslake. 2005. *Turkish. A comprehensive grammar*. London, New York: Routledge.

- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (ed.), *Universals of Human Language*, 73–113. Cambridge, Mass: MIT Press.
- Grinevald, Collette & Michel Bert. 2011. Speakers and Communities. In Austin, Peter K. & Sallabank, Julia (eds.), *Cambridge Handbook of Endangered Languages*, 45–65. Cambridge University Press.
- Hadjidas, Harris & Vollmer, Maria. 2015. Multi-CAST Cypriot Greek. In Haig, Geoffrey & Schnell, Stefan (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. Bamberg: University of Bamberg. Retrieved from <https://multicast.aspra.uni-bamberg.de/#cypgreek> on 16 Nov 2022.
- Haig, Geoffrey. 2017. Western Asia. East Anatolia as a Transition Zone. In Hickey, Raymond (ed.), *The Cambridge Handbook of Areal Linguistics*, 396–423. Cambridge: Cambridge University Press.
- Haig, Geoffrey. 2014. East Anatolia as a linguistic area? Empirical and conceptual issues. In Behzadi, Lale, Franke, Patrick, Haig, Geoffrey, Herzog, Christoph, Hoffmann, Birgitt, Korn, Lorenz & Talabardon, Susanne (eds.), *Bamberger Orient Studien*, Vol. 1., 13–36. Bamberg: Bamberg University Press.
- Haig, Geoffrey. 2001. Linguistic diffusion in present-day East Anatolia: From top to bottom. In Aikhenvald, Alexandra Y. & Dixon, R.M.W. (eds.), *Areal diffusion and genetic inheritance. Problems in comparative linguistics*. 195–224. Oxford: Oxford University Press.
- Haig, Geoffrey & Geoffrey Khan. 2018. Introduction. In: Haig, Geoffrey & Khan, Geoffrey (eds.), *The languages and linguistics of Western Asia. An areal perspective*, 1–29. Berlin, Boston: De Gruyter Mouton.
- Haig, Geoffrey, Paul Noorlander & Nils N. Schiborr. Under review. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian Transition Zone. In Darquennes, Jeroen, Salmons, Joe & Vandenbussche, Wim (eds.), *Language contact. An international handbook*, Vol. 2. Berlin, Boston: De Gruyter Mouton.
- Haig, Geoffrey, Dorota Molin & Paul Noorlander. 2022. Tracing the VO-to-OV shift across Northeastern Neo-Aramaic: quantitative and qualitative analyses. *Paper presented at the 1<sup>st</sup> International Conference on the Languages of the Anatolia-Caucasus-Iran-Mesopotamia area (LACIM)*, Paris, 16–18 November 2022.
- Haig, Geoffrey, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils N. Schiborr (eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology*. Berlin: Language Science Press.
- Haspelmath, Martin. 2013a. Indefinite Pronouns. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/46> on 10 May 2021.
- Haspelmath, Martin. 2013b. 'Want' Complement Subjects. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/124> on 01 Mar 2022.
- Haspelmath, Martin. 2013c. Ditransitive Constructions. The Verb 'Give'. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/105> on 05 Jan 2022.
- Haspelmath, Martin. 2001. The European linguistic area. Standard Average European. In Haspelmath, Martin, König, Ekkehard & Raible, Wolfgang (Eds.), *Language typology and language universals*. Vol. 2, 1492–1510. Berlin, Boston: De Gruyter Mouton.



- Holton, David, Geoffrey Horrocks, Marjolijne Janssen, Tina Lendari, Io Manolessou & Notis Tournafexis. 2019. *The Cambridge grammar of medieval and early modern Greek*. Cambridge: Cambridge University Press.
- Horrocks, Geoffrey. 2010. *Greek. A history of the language and its speakers*. Malden, MA, Oxford: Wiley-Blackwell.
- Janse, Mark. 2022. Η καππαδοκική διάλεκτος [The Cappadocian dialect]. In Tsitsilis, Christos (Ed.), *Νεοελληνικές διάλεκτοι* [Modern Greek dialects]. Thessaloniki: Institute of Modern Greek Studies.
- Janse, Mark. 2009. Greek-Turkish language contact in Asia Minor. In Tamis, Tassos (ed.), *Etudes Helleniques – Hellenic Studies* 17(1). 37–54.
- Janse, Mark. 2008. Clitic doubling from Ancient to Asia Minor Greek. In Dalina Kallulli & Tasmowski, Liliane (eds.), *Clitic Doubling in the Balkan Languages*, 165–202. Amsterdam, Philadelphia: Benjamins.
- Janse, Mark. 2004. Animacy, definiteness, and case in Cappadocian and other Asia Minor Greek dialects. *Journal of Greek Linguistics* 5. 3–26.
- Janse, Mark. 2002. Aspects of Pontic grammar. *Journal of Greek Linguistics* 3. 203–231.
- Janse, Mark. 1999. Les besoins de la Cause: causatifs hébraïques/causatifs grecs. In Alain Blanc & Christol, Alain (eds.), *Langues en contact dans l'Antiquité. Aspects lexicaux*, 131–149. Nancy: de Boccard.
- Jessica Kantarovich, Jessica, Lenore A. Grenoble, Antonina Vinokurova & Elena Nesterova. 2021. Complexity and Simplification in Language Shift. *Frontiers in Communication* 6: 638118. DOI: 10.3389/fcomm.2021.638118.
- Karatsareas, Petros. 2016. The Asia Minor Greek adpositional cycle. *Journal of Greek Linguistics* 16(1). 47–86. DOI: <https://doi.org/10.1163/15699846-01601001>.
- Karatsareas, Petros. 2014. On the diachrony of gender in Asia Minor Greek. The development of semantic agreement in Pontic. *Language Sciences* 43. 77–101.
- Karatsareas, Petros. 2013. Understanding diachronic change in Cappadocian Greek. The dialectological perspective. *Journal of Historical Linguistics* 3(2). 192–229. DOI: 10.1075/jhl.3.2.02kar.
- Karatsareas, Petros. 2011. A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective. Cambridge: University of Cambridge dissertation.
- Karatsareas, Petros & Marika Lekakou. 2016. Marking definiteness multiply. Evidence from two varieties of Greek. *Studies in Greek Linguistics* 36. 189–204.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63–99.
- Kornfilt, Jaklin. 1997. *Turkish*. London, New York: Routledge.
- Lacroix, René. 2018. Laz. In Haig, Geoffrey & Khan, Geoffrey (eds.), *The languages and linguistics of Western Asia. An areal perspective*, 830–861. Berlin, Boston: De Gruyter Mouton.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig. 2016. *Ethnologue. Languages of the World*, 19<sup>th</sup> edn. Dallas: SIL International. <http://www.ethnologue.com> (accessed on 29 March 2016).
- Lowry, Heath. 1977. The Ottoman tahrir defters as a source for urban demographic history. The case of Trabzon (ca. 1463–1583). Los Angeles: UCLA dissertation.
- Lüpke, Friederike. 2011. Orthography development. In Austin, Peter & Sallabank, Julia (Eds.), *The Cambridge Handbook of Endangered Languages*, 312–336. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511975981.016.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511605789.

- Mackridge, Peter. 1999. The Greek spoken in the Region of Pontus. In Christidis, Anastasios-Foivos, Arapopoulou, Maria & Giannouloupoulou, Giannoula (eds.), *Διαλεκτικοί θύλακοί της ελληνικής γλώσσας* [Dialect enclaves of the Greek language], 101–105. Athens: Ministry of National Education and Religious Affairs/Centre for the Greek Language.
- Mackridge, Peter. 1995. Pontic in contemporary Turkey. Ancient features in the dialect of Ofis. *Αρχαίον Πόντου* 46. 133–161.
- Mackridge, Peter. 1987. Greek-speaking moslems of North-East Turkey. Prolegomena to a study of the Ophitic sub-dialect of Pontic. *Byzantine and Modern Greek Studies* 11. 115–137.
- Mackridge, Peter. n.d. Pontus and the Pontic infinitive [Draft]. Extract from Peter Mackridge's unpublished memoirs, 13/04–02/06/2022.
- Manolessou, Io. 2005. From participles to gerunds. In Stavrou, Melita & Trezi, Arhonto (eds.), *Advances in Greek generative syntax*, 241–283. Amsterdam, Philadelphia: Benjamins. DOI: 10.1075/la.76.10man.
- Matras, Yaron. 2012a. An activity-oriented approach to contact-induced language change. In Chamoreau, Claudine & Léglise, Isabelle (eds.), *Dynamics of contact-induced language change*, 17–52. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110271430.17.
- Matras, Yaron. 2012b. *A grammar of Domari*. Berlin, Boston: De Gruyter Mouton.
- Matras, Yaron. 2011. Universals of structural borrowing. In Siemund, Peter (ed.), *Linguistic universals and language variation*, 204–234. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110238068.204.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Matras, Yaron & Jeanette Sakel 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.
- Meakins, Felicity, Jennifer Green & Myfany Turpin. 2018. *Understanding linguistic fieldwork*. London, New York: Routledge.
- Melissaropoulou, Dimitra & Angelika Ralli. 2019. Revisiting the borrowability scale(s) of free grammatical elements. Evidence from Modern Greek contact induced varieties. *Journal of Language Contact* 12. 707–736.
- Michelioudakis, Dimitris & Ioanna Sitaridou. 2020. Towards a formal model of transfer under contact. Contrasting Asia Minor Greek to mainland Greek and Turkish in search of syntactic borrowings. In Georgiafentis, Michalis, Giannouloupoulou, Giannoula, Koliopoulou, Maria & Tsokoglou, Angeliki (eds.), *Contrastive studies in morphology and syntax*, 245–261. London, New York: Bloomsbury.
- Michelioudakis, Dimitris & Ioanna Sitaridou. 2016. Recasting the typology of multiple wh-fronting. Evidence from Pontic Greek. *Glossa: A journal of general linguistics* 1(1): 40. 1–33. DOI: 10.5334/gjgl.72.
- Michelioudakis, Dimitris & Ioanna Sitaridou. 2013. Multiple wh-fronting across Pontic Greek dialects. In Ralli, Angela, Joseph, Brian, Janse, Mark & Karasimos, Athanasios (eds.), *Online Proceedings of the 5<sup>th</sup> International Conference of Modern Greek Dialects and Linguistic Theory* 5(1). 353–378.
- Michelioudakis, Dimitris & Ioanna Sitaridou. 2012. Syntactic microvariation. Dative constructions in Greek. In Etxepare, Ricardo & Fernández, Beatriz (eds.), *Datives in variation. A micro-comparative perspective*, 212–255. Oxford: Oxford University Press.
- Michelioudakis, Dimitris & Ioanna Sitaridou. 2010. Syntactic micro-variation in Pontic Greek. Dative constructions. In Ralli, Angela, Joseph, Brian, Janse, Mark & Karasimos, Athanasios (eds.), *Online Proceedings of the 4<sup>th</sup> International Conference of Modern Greek Dialects and Linguistic Theory (MGDLT4), Chios, 11-14 June 2009*. 131–146.

- Moseley, Christopher. 2010. *Atlas of the World's Languages in Danger*, 3<sup>rd</sup> edn. Paris: UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000187026> (accessed on 17 November 2022).
- Moseley, Christopher. 2007. *Encyclopedia of the World's Endangered Languages*. London, New York: Routledge.
- Munro, Pamela. 1982. On the transitivity of “say” verbs. In Hopper, Paul J. & Thompson, Sandra A. (eds.), *Studies in Transitivity*, 301–318. New York: Academic Press.
- Myers-Scotton, Carol. 2002. *Contact Linguistics. Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
- Németh, Gyula. 1962/2020. *Turkish grammar*. (ed. Halasi-Kun, Tibor). Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783112317150.
- Neocleous, Nicolaos. 2022. The evolution of VO and OV alternation in Romeyka. In Asadpour, Hiwa & Jügel, Thomas (eds.), *Word order variation. Semitic, Turkic and Indo-European Languages in Contact*, 141–161. Berlin, Boston: De Gruyter Mouton.
- Neocleous, Nicolaos. 2020. Word order and information structure in Romeyka. A syntax and semantics interface account of order in a minimalist system. University of Cambridge dissertation. DOI: 10.17863/CAM.50619.
- Neocleous, Nicolaos & Ioanna Sitaridou. 2022. Never just contact. The rise of final auxiliaries in Asia Minor Greek. *Diachronica* 39(3), 369–408. DOI: 10.1075/dia.17048.
- Nicholas, Nick. 1998a. The Story of pu. The Grammaticalisation in Space and Time of a Modern Greek Complementiser. Melbourne: University of Melbourne dissertation.
- Nicholas, Nick. 1998b. TO APER and O OPOIOS. Untangling Mediaeval Greek relativisation. In Joseph, Brian, Horrocks, Geoffrey & Philippaki-Warbuton, Irene (eds.), *Themes in Greek Linguistics II*, 283–323. Amsterdam, Philadelphia: Benjamins.
- Oikonomidis, Demosthenes H. 1958. *Γραμματική της ελληνικής διαλέκτου του Πόντου* [Grammar of the Greek dialect of Pontos]. Athens: Ακαδημία Αθηνών.
- Oikonomidis, Demosthenes H. 1908. *Laulehre des Pontischen*. Leipzig: A. Deichert'sche Verlagsbuchhandlung Nachf.
- Özkan, Hakan. 2013. The Pontic Greek spoken by Muslims in the villages of Beşkøy in the province of present-day Trabzon. *Byzantine and Modern Greek Studies* 37(1). 130–150.
- Özkan, Hakan. n.d. Grammar sketch of Romeyka. Unpublished draft.
- Papadopoulos, Anthimos A. 1961. *Ἱστορικὸν Λεξικὸν τῆς Ποντικῆς Διαλέκτου, Τόμος Δεύτερος* [Historical dictionary of the Pontic Dialect, Vol. 2]. Archeion Pontou. Athens: Epitropi Pontiakon Meleton.
- Papadopoulos, Anthimos A. 1958. Γλωσσικά ποικίλα. Archeion Pontou 22. 191–198.
- Papadopoulos, Anthimos A. 1955. *Ἱστορικὴ γραμματικὴ τῆς ποντικῆς διαλέκτου* [Historical Grammar of the Pontic dialect]. Athens: Epitropi Pontiakon Meleton.
- Papadopoulos, Anthimos A. 1933. Συμβολή εις την έρευναν της ποντικῆς διαλέκτου (ήτοι γραμματικά και συντακτικά ζητήματα) [A contribution to the study of Pontic Dialect. Grammatical and syntactic issues]. *Αθηνά* 45. 15–48.
- Parcharidis, Ioannis. 1880. Γραμματική της διαλέκτου Τραπεζουντος. Περί επιθέτων, αντωνυμιών και βαρυτόνων ρημάτων [Grammar of the Trabzon dialect. On adjectives, pronouns and baritone verbs]. *Χειρόγραφο* 335, Κέντρο Ερεύνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων [Research Center of Modern Greek Dialects and Idioms], Academy of Athens.
- Payne, Thomas E. 1997. *Describing morphosyntax. A guide for field linguists*. Cambridge: Cambridge University Press.
- “Pontic”. *Catalogue of Endangered Languages*. 2017. University of Hawaii at Manoa. <http://www.endangeredlanguages.com/lang/3265> (accessed on 17 Nov 2022).

- Poplack, Shana. 1993. Variation theory and language contact. In Preston, Denis R. (ed.), *American dialect research. An anthology celebrating the 100<sup>th</sup> anniversary of the American Dialect Society*, 251–286. Amsterdam, Philadelphia: Benjamins.
- Poplack, Shana & Stephen Levey. 2010. Contact-induced grammatical change: a cautionary tale. In Auer, Peter & Schmidt, Jürgen E. (eds.), *Language and Space. An International Handbook of Linguistic Variation*, 391–419. Berlin, New York: Mouton de Gruyter.
- Prince, Kilu von & Sebastian Nordhoff. 2020. An empirical evaluation of annotation practices in corpora from language documentation. *Proceedings of the 12<sup>th</sup> Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, 11–16 May 2020. 2778–2787.
- Ralli, Angela. 2020. Matter versus pattern borrowing in compounding: Evidence from the Asia Minor Greek dialectal variety. *Morphology* 30. 423–446. <https://doi.org/10.1007/s11525-020-09362-6>.
- Ralli, Angela. 2008. Compound markers and parametric variation. *Language Typology and Universals (STUF)* 61. 19–38.
- Revithiadou, Anthi & Vassilis Spyropoulos. 2021. In search of the Pontic Koine and the factors that have led to its development. *Paper presented at the 2nd International Conference on the Koine, koines, and the formation of Standard Modern Greek*, November 5–6, 2021 at the Aristotle University Thessaloniki.
- Revithiadou, Anthi & Vassilis Spyropoulos. 2012. *Οφίτικα: πτυχές της γραμματικής δομής μίας ποντιακής διαλέκτου* [Ophitic. Aspects of the Grammatical Structure of a Pontic Dialect]. Thessaloniki: Kyriakidis.
- Rice, Keren. 2011. Ethical Issues in Linguistic Fieldwork. In Thieberger, Nicholas (ed.), *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press.
- Sağlam, Erol. 2017. Constitutive ambiguities. Subjectivities and Memory in the Case of Romeika-Speaking Communities of Trabzon, Turkey. London: Birbeck University of London dissertation.
- Sakel, Jeanette & Yaron Matras. 2008. Modelling contact-induced change in grammar. In Stolz, Thomas, Bakker, Dik & Salas Palomo, Rosa (eds.), *Aspects of language contact. New theoretical, methodological and empirical findings with special focus on romancisation processes*, 63–88. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110206043.63.
- Schlücker, Barbara & Tanja Ackermann. 2017. The morphosyntax of proper names. An overview. *Folia Linguistica* 51(2). 309–339. DOI: 10.1515/flin-2017-0011.
- Schreiber, Laurentia. In preparation. The Romeyka corpus.
- Schreiber, Laurentia. 2022. Word order & post-predicate elements in Romeyka. *Paper presented at the Workshop on Post-predicate elements across the languages of Western Asia: Theoretical and empirical approaches*, University of Bamberg, 22–23 September, 2022.
- Schreiber, Laurentia. 2021. Pontic Greek (Romeyka). In Haig, Geoffrey, Stilo, Donald, Doğan, Mahir C. & Schiborr, Nils N. (eds.), *WOWA – Word Order in Western Asia. A spoken-language-based corpus for investigating areal effects in word order variation*. Bamberg: University of Bamberg. Retrieved from <https://multicast.aspra.uni-bamberg.de/resources/wowa/> on 01 Feb 2022.
- Schreiber, Laurentia. 2018. Romeyka. In Haig, Geoffrey & Khan, Geoffrey (eds.), *The languages and linguistics of Western Asia. An areal perspective*, 892–933. Berlin, Boston: De Gruyter Mouton.

- Schreiber, Laurentia. 2016. Assessing sociolinguistic vitality. An attitudinal study of Rumca (Romeyka). In Wiese, Heike, Marten, Heiko F. & Bunk, Oliver (eds.), *Arbeitspapiere "Sprache, Variation und Migration"*. *Studentische Arbeiten*, Papier Nr. 4. Potsdam: Universität Potsdam.  
[https://www.uni-potsdam.de/fileadmin/projects/svm/Arbeitspapiere/No4\\_Schreiber\\_2016.pdf](https://www.uni-potsdam.de/fileadmin/projects/svm/Arbeitspapiere/No4_Schreiber_2016.pdf) (accessed on 17 Nov 2022).
- Schreiber, Laurentia, Geoffrey Haig, Mortaza Taheri-Ardali & Erik Anonby. 2021. Contact-induced change in the morphosyntax of Turkic in Boldaji, Chahar Mahal va Bakhtiari Province, Iran. *Turkic Languages* 25. 210–242. DOI: 10.13173/TL.25.2.210.
- Schreiber, Laurentia & Mark Janse. In preparation. Word order and post-predicate elements in Romeyka. In Haig, Geoffrey, Rasekh-Mahand, Mohammad, Stilo, Donald, Schreiber, Laurentia & Schiborr, Nils N. (eds.), *Post-predicate elements in the Western Asian Transition Zone. A corpus-based approach to areal typology*. Berlin: Language Science Press.
- Schreiber, Laurentia & Ioanna Sitaridou. 2017. Assessing the sociolinguistic vitality of Istanbulite Romeyka. An attitudinal study. *Journal of Multilingual and Multicultural Development* 39(1). 1–16. DOI: 10.1080/01434632.2017.1301944.
- Scontras, Gregory, Zuzanna Fuchs & Maria Polinsky. 2015. Heritage language and linguistic theory. *Frontiers in Psychology* 6: 1545. 1–20. DOI: 10.3389/fpsyg.2015.01545.
- Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/44> on 26 Apr 2022.
- Sitaridou, Ioanna. 2016. Reframing the phylogeny of Asia Minor Greek: The view from Pontic Greek. *CHS Research Bulletin* 4(1). 1–17.
- Sitaridou, Ioanna. 2014a. Modality, antiveridicality, and complementation: The Romeyka infinitive as a negative polarity item. *Lingua* 148. 118–146.
- Sitaridou, Ioanna. 2014b. The Romeyka infinitive: Continuity, contact and change in the Hellenic varieties of Pontus. *Diachronica* 31(1). 23–73.
- Sitaridou, Ioanna. 2013. Greek-speaking enclaves in Pontus today: The documentation and revitalization of Romeyka. In Jones, Mari & Ogilvie, Sarah (eds.), *Keeping languages alive. Documentation, pedagogy, and revitalization*, 98–112. Cambridge: Cambridge University Press.
- Sitaridou, Ioanna. 2011. Syntactic microvariation in complementation strategies in Pontic Greek varieties. Abstract submitted to the 5<sup>th</sup> European Dialect Syntax Workshop, 25<sup>th</sup> June 2011, University of Amsterdam.
- Sitaridou, Ioanna & Maria Kaltsa. 2014. Contrastivity in Pontic Greek. *Lingua* 146, 1–27.
- Skopeteas, Stavros. 2016. The Caucasian Urums and the Urum language/Kafkasya Urumları ve Urum Dili. In Eker, Süer & Ülkü Çelik (Eds.), *Handbook of Endangered Turkic Languages*. Ankara: Astana.
- Stassen, Leon. 2013a. Zero Copula for Predicate Nominals. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/120> on 31 Jan 2022.
- Stassen, Leon. 2013b. Predicative Possession. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/117> on 07 Feb 2022.

- Stassen, Leon. 2013c. Comparative Constructions. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/121> on 28 Mar 2022.
- Stassen, Leon. 2013d. Noun Phrase Conjunction. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/63> on 18 Feb 2022.
- Stassen, Leon. 1997. *Intransitive Predication*. Oxford: Clarendon Press.
- TFS Working Group (2011). The Woodchopper. *Totem Field Storyboards*. Retrieved from <http://www.totemfieldstoryboards.org> on Nov 15, 2022.
- Thomason, Sarah G. 2007. Language contact and deliberate change. *Journal of Language Contact* 1(1). 41–62. DOI: 10.1163/000000007792548387.
- Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Tiersma, Peter Meijes. 1982. Local and General Markedness. *Language* 58(4). 832–849.
- Tombaidis, Dimitrios E. 1996. Η ποντιακή διάλεκτος [The Pontic dialect], In Tombaidis, Dimitrios (ed.), *Μελετήματα ποντιακής διαλέκτου* [Studies on the Pontic Dialect], 222–233. Thessaloniki: Kothikas.
- Tombaidis, Dimitrios E. 1992. Η τύχη των μικρασιατικών ιδιωμάτων στον ελληνικό χώρο [The fate of the Asia Minor dialects in Greece], *Centre for Asia Minor Studies Bulletin* 9. 241–250.
- Tsunoda, Tasaku. 2005. *Language endangerment and language revitalization*. Berlin, New York: Mouton De Gruyter.
- Tursun, Vahit. 2019. *Romeika – Türkçe sözlük. Trabzon Rumcası* [Romeika – Turkish dictionary]. Istanbul: Heyamola Yayınları.
- Urban, Matthias. 2020. Mountain linguistics. *Language and Linguistics Compass* 14(9). 1–23. DOI: 10.1111/lnc3.12393.
- Vryonis, Speros. 1986. *The decline of medieval Hellenism in Asia Minor. And the process of Islamization from the eleventh through the fifteenth century*. (Publications of the Center for Medieval and Renaissance Studies). Berkeley: University of California Press.
- Winford, Donald. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22(2). 373–427.
- Woodbury, Anthony C. 2011. Language documentation. In Austin, Peter K. & Sallabank, Julia (eds.), *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.
- Yarbay Duman, Tuba, Nermin Altınok, Neşe Özgirgin & Roelin Bastiaanse. 2011. Sentence comprehension in Turkish Broca’s aphasia: An integration problem. *Aphasiology* 25(8). 908–926.

## Appendix A: Corpus overview

**Table A.1:** Speaker information

Speaker no.	Speaker code(s) as part of the recording codes	Gender	Approx. age (in years)	Place of birth	Place of residence
1	01_28062019F	F	72	Çaykara, Uzungöl (Saráchos), Sóros/Kastanós	Çaykara (vicinity)
2	03_30062019F	F	90	Dernekpazarı (Kondú)	Dernekpazarı (Kondú)
3	04_01072019F	F	45	Çaykara, Karaçam/Ogene (Ótšena)	Çaykara (vicinity)
4	07_04072019F	F	75	Çaykara, Karaçam/Ogene (Ótšena)	Çaykara (vicinity)
5	05_03072019M	M	40	Sürmene, Beşköy	Sürmene, Beşköy
6	06_03072019M	M	50	Çaykara, Karaçam/Ogene (Ótšena)	Çaykara (vicinity)
7	08_04072019M	M	55	Çaykara, Karaçam/Ogene (Ótšena)	Çaykara (vicinity)
8	01_07072019F 03_07072019F	F	50	Çaykara, Köknar/Ogene (Ótšena) <sup>?</sup>	Çaykara (vicinity)
9	09_04072019F 01_06042017F 01_04022016F 01_15022015F 01_02022015F 02_02022015F 01_14012015F	F	85	Çaykara, Ataköy, Soğanlı (Zíhon)	Istanbul
10 (H1)	03_21042018M	M	97	Çaykara, Uzungöl (Saráchos), Köseli	Germany
11 (H2)	02_21042018M	M	52	Çaykara, Uzungöl (Saráchos), Köseli	Germany
12 (H3)	01_21042018M	M	47	Çaykara, Uzungöl (Saráchos), Köseli	Germany

**Table A.2:** Recording catalogue of the Romeyka corpus

Trabzon 2019 Recording catalogue per type of elicitation					
(Light grey shading = transcribed in questionnaire, not annotated)					
No	Date	Code	Type	Topic	Length of Recording
1	28.06.2019	01 28062019F 2	Free speech	Monologue	02:07min
2	28.06.2019	01 28062019F 3	Free speech	Monologue	01:48min
3	29.06.2019	02 29062019F 1	Free speech	Monologue: bear story	01:54min
4	29.06.2019	02 29062019F 2	Free speech	Monologue: bear story continuation	02:02min
5	30.06.2019	03 30062019F 1	Free speech	Monologue: pasture works	01:06min
6	30.06.2019	03 30062019F 2	Free speech	Monologue: recipe cheese	01:59min
7	30.06.2019	03 30062019F 6	Free speech	Monologue: lack of rain	03:17min
8	30.06.2019	03 30062019F 7	Free speech	Monologue	02:17min
9	30.06.2019	03 30062019F 8	Free speech	Monologue: death of husband	01:10min
10	30.06.2019	03 30062019F 11	Free speech	Monologue: recipe	03:44min
11	01.07.2019	04 01072019F 1	Free speech	Monologue: pasture works	07:42min
12	01.07.2019	04 01072019F 2	Free speech	Monologue: youth	10:59min
13	01.07.2019	04 01072019F 5	Free speech	Monologue/Conversation	02:27min
14	01.07.2019	04 01072019F 17	Free speech	Conversation	02:31min
15	03.07.2019	06 03072019M 2	Free speech	Monologue: family	03:18min
16	04.07.2019	07 04072019F 5	Free speech	Monologue	02:40min
17	04.07.2019	07 04072019F 6	Free speech	Monologue: experiences	03:08min
18	04.07.2019	07 04072019F 8	Free speech	Monologue: bear in garden story	01:34min
19	04.07.2019	07 04072019F 10	Free speech	Monologue: a long way	01:12min
20	04.07.2019	08 04072019M 1	Free speech	Monologue	12:01min
21	04.07.2019	08 04072019M 2	Free speech	Monologue	05:17min
22	04.07.2019	08 04072019M 3	Free speech	Conversation	06:03min
23	04.07.2019	08 04072019M 4	Free speech	Conversation	01:12min
24	04.07.2019	09 04072019F 7	Free speech	Monologue: old times at the village	02:38min
25	05.07.2019	01 05072019F 1	Free speech	Monologue/WhatsApp message	00:47min
26	05.07.2019	02 05072019F 1	Free speech	Video conversation	01:09min
27	07.07.2019	03 07072019F 1	Free speech	Monologue/WhatsApp message	02:58min
28	06.04.2017	01 06042017F 4	Free speech	Monologue: last kina you attended	09:27min



29	04.02.2016	01_04022016F_1	Free speech	Dialogue: phone call	06:25min
30	15.02.2015	01_15022015F_1	Free speech/telephone	Personal greetings	01:25min
31	02.02.2015	01_02022015F_1	Free speech/telephone	Personal greetings	00:35min
32	14.01.2015	01_14012015F_1	Free speech/telephone	Personal greetings	00:26min
33	02.02.2015	02_02022015F_1	Free speech	Monologue	09:15min
34	01.07.2019	04_01072019F_12	Semi-naturalistic elicitation	Description picture prompt bear story	02:35min
35	01.07.2019	04_01072019F_13	Semi-naturalistic elicitation	Description picture prompt wood story	02:39min
36	03.07.2019	05_03072019M_3	Semi-naturalistic elicitation	Description picture prompt bear story_1	02:18min
37	03.07.2019	05_03072019M_4	Semi-naturalistic elicitation	Description picture prompt bear story_2	01:55min
					<b>Total: 116min</b>
38	28.06.2019	01_28062019F_4	Elicitation/Free speech	Q_lexicon/monologue	05:07min
39	30.06.2019	03_30062019F_13	Elicitation	Q_morphsynt	18:20min
40	30.06.2019	03_30062019F_14	Elicitation	Q_morphsynt: verb paradigms	16:54min
41	01.07.2019	04_01072019F_6	Elicitation	Q_lexicon_1	00:41min
42	01.07.2019	04_01072019F_7	Elicitation	Q_lexicon_2	12:44min
43	01.07.2019	04_01072019F_8	Elicitation	Q_lexicon_3	01:30min
44	01.07.2019	04_01072019F_9	Elicitation	Q_morphsynt_1	02:15min
45	01.07.2019	04_01072019F_10	Elicitation	Q_morphsynt_2	01:08min
46	01.07.2019	04_01072019F_11	Elicitation	Q_morphsynt_3	22:42min
47	03.07.2019	05_03072019M_1	Elicitation	Q_lexicon_1	28:26min
48	03.07.2019	05_03072019M_2	Elicitation	Q_lexicon_2	01:39min
49	04.07.2019	09_04072019F_1	Elicitation	Q_lexicon	37:15min
50	21.04.2018	01_21042018M_1	Elicitation	Q_morphosynt/monologue	21:34min
51	21.04.2018	02_21042018M_2	Elicitation	Q_morphosynt/monologue	21:08min
52	21.04.2018	03_21042018M_3	Elicitation	Q_morphosynt/monologue	17:48min
53	21.04.2018	04_21042018M_4	Elicitation	Q_lexicon	10:24min
54	06.04.2017	01_06042017F_0	Elicitation	Q_lexicon	48:14min
55	06.04.2017	01_06042017F_1	Elicitation	Q_morphosynt	31:01min
56	06.04.2017	01_06042017F_2	Elicitation	Q_morphosynt	20:51min
57	06.04.2017	01_06042017F_3	Elicitation	Q_morphosynt	12:51min

**Table A.3:** Glosses of the Romeyka corpus in ELAN

abl	Ablative (Turkish)	loc	Locative (Turkish)
add	Addressee of verb of speech	m	Masculine
adj.a	Adjective (attributive)	n	Noun
adj.p	Adjective (predicative)	n	Neuter
adv.m	Adverb of manner	n.indef	Indefinite noun
adv.p	Adverb of place	neg	Negation particle
adv.t	Adverb of time	nmz	Nominalization
aor	Aorist	nn	Proper noun
ben	Beneficial	nq	Numeral classifier
com	Comitative	num	Numeral
comp	Comparative	opn	Object pronoun
cond	Conditional particle	opt	Optative particle
conj	Conjunction	pl	Plural
dat	Dative (Turkish)	pn	Personal pronoun
dem	Demonstrative	poss	Possessive pronoun
det	Definite article	prep	Preposition
dp	Discourse particle	pres	Present tense
f	Feminine	pron	Pronoun
foc	Focus particle	prt	Particle
fut	Future particle	pst	Simple Past (Turkish)
gen	Genitive	ptcp	Participle
imp	Imperative	quant	Quantifier
indet	Indefinite article	reflx	Reflexive
instr	Instrumentative	rel	Relativizer
int	Interjection	v	Verb
interr	Interrogative	1,2,3	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> person
ipfv	Imperfective		

## Appendix B: Phonological analyses

**Table B.1A:** Distributional patterning of vowels

<b>i</b>	<b>e</b>	<b>æ</b>	<b>a</b>	<b>æ</b>	<b>o</b>	<b>u</b>
‘#_l	‘#_l	d_#	‘#_n	‘g_ŋ	‘#_d	‘#_l
‘ð_a	‘#_m	d_#	‘#_r	‘g_ŋ	‘#_n	‘r_ɣ
‘f_l	‘#_p	r_#	‘ð_r	‘tʃ_r	‘b_l	‘t_ð
‘l_tʃ	‘ð_l	s_#	‘ð_x	g_’l	‘ð_m	#_t
‘m_d	‘f_ŋ	d_#	‘k_d		‘ð_n	b_g
‘r_ð	‘f_s	f_#	‘l_t		‘f_#	f_r
‘r_z	‘j_m	f_#	‘m_d		‘i_#	h_m
‘t_r	‘l_tʃ	l_#	‘m_m		‘l_s	p_#
‘tʃ_r	‘l_v		‘m_n		‘ps_m	s_’b
‘z_j	‘n_k		‘ps_r		‘r_v	t_#
#_n	‘ʃ_r		‘s_r		‘st_m	x_l
d_#	#_b		‘t_m		‘x_m	
d_#	#_k		‘ɣ_l		‘ɣ_n	
ð_#	ð_#		#_’l		#_’m	
j_n	f_#		#_ð		#_’ps	
l_θ	l_#		b_’t		#_’s	
n_#	m_#		j_#		#_’v	
n_#	m_#		k_#		d_(n)	
s_m	m_l		k_r		g_(s)	
v_z	n_#		l_s		g_#	
z_#	n_m		m_’l		k_’l	
	n_m		m_#		k_’s	

	n_r		m_#		l_#	
	p_đ		m_#		n_#	
	ps_#		m_#		p_’đ	
	r_#		n_a		r_s	
	r_#		s_#		t_’k	
			z_#		t_#	
					x_’r	
					z_#	
					‘n_#	

**Table B.1B:** Distributional patterning of vowels with examples

environm ents	/i/	environm ents	/e/	environm ents	/æ/	environm ents	/a/	environm ents	/œ/	environm ents	/o/	environm ents	/u/
‘tš_r	tširi	‘š_r	šeri	d_#	omadæ	‘m_d	omadi	‘g_ŋ	gæŋksæ	#_’m	omadi	‘t_ð	stuði
d_#	omadi	‘j_m	jema	d_#	ðondæ	‘t_m	potami	‘g_ŋ	gæni	‘#_d	odi	f_r	furno
d_#	odi	‘n_k	ineka	r_#	poðaræ	m_#	stoma	‘tš_r	tšærado	‘st_m	stoma	s_’b	tsubaði
‘m_d	midi	p_ð	peði	s_#	gönksæ	s_#	γlossa	g_’l	gælimbe ma	‘l_s	γlossa	x_l	xuljeri
v_z	vizija	‘ð_l	γarðeli	d_#	ospidæ	‘ð_x	ðaxtila			‘ð_n	ðondæ	‘r_γ	kruyo
‘z_j	vizija	f_#	nife	f_#	xorafæ	n_a	γonado			‘γ_n	γonado	h_m	kahume
‘ð_a	karðia	n_m	onema	f_#	aðelfæ	‘ð_r	poðari			p_’ð	poðari	b_g	ebuğa
n_#	göni	‘l_tš	skoletši	l_#	γarðelæ	j_#	vizija			t_#	to	p_#	pu
ð_#	stuði	‘l_v	alevre			k_r	karðia			‘b_l	bola	#_t	utš
#_n	ineka	l_#	katsimale			m_’l	malia			r_s	auros	t_#	katu
j_n	jinega	m_l	xameleda			m_#	jema			‘f_#	aðelfo	‘#_l	ula
‘t_r	fitira	n_r	nero			k_#	ineka			‘#_n	onema		
‘l_tš	γlitši	n_m	anemo			‘m_n	mana			‘r_v	provado		
‘r_ð	kariði	‘f_ŋ	fengo			‘m_m	mami			d_(n)	provadon		
‘r_z	rizi	m_#	ayome			#_ð	aðelfo			k_’s	kosarin		
z_#	rizi	‘#_l	ela			b_’t	batsi			z_#	zo		
‘f_l	fila	‘f_s	feso			‘#_n	andra			l_#	škilo		
s_m	apsimo	m_#	ðome			m_#	onema			n_#	furno		
l_θ	liθari	‘#_p	epare			z_#	za			g_#	ligo		
‘#_l	ilos	r_#	epare			‘s_r	kosarin			g_(s)	ligos		
n_#	soni	#_k	ekoliya			‘k_d	kada			#_’ps	opsari		
		ps_#	opse			‘#_r	argos			#_’v	ova		
		#_b	eban			‘ps_r	opsari			‘ps_m	psomi		

		‘#_m	<u>e</u> ma			#_’l	<u>a</u> levre			k_’l	sko <u>l</u> etši		
		r_#	alevr <u>e</u>			l_s	al <u>a</u> s			t_’k	li <u>f</u> tokari		
		ð_#	stuð <u>e</u>			‘y_l	y <u>a</u> la			#_’s	<u>o</u> spidi		
		n_#	šon <u>e</u>			‘l_t	al <u>a</u> ti			x_’r	xo <u>r</u> io		
						m_#	xom <u>a</u>			‘i_#	xori <u>o</u>		
										‘ð_m	rð <u>o</u> mo		
										‘x_m	x <u>o</u> ma		
										‘n_#	kamn <u>o</u>		

**Table B.2A:** Distributional patterning of stops

[p]	[b]	[b]	[t]	[t]	[d]	[d]	[k]	[k]	[g]	[g]
‘#_i	‘#_a	‘#_i	‘f_i	x_e	‘#_i	r_a	‘#_a	‘#_s	‘#_o	r_o
‘#_r	‘#_o	‘#_o	a_u		‘a_a	e_a	#_a	‘#_r	‘š_i	a_’r
‘#_s	‘i_a	e_u	o’_a		#_r		e_o	‘#_a	e_a	
‘a_s	#_a		u_e		a_a		o’_a		i_o	
#_o	#_e				a_i		s_o		ŋ_o	
a’_e					a_o		u’_a			
a’_l					u_e					
e’_a					e_i					
e’_e					f_o					
e’_i					i_i					
o’_i					n_r					
o’_s										
s’_i										

**Table B.2B:** Distributional patterning of stops with examples

environm ents	/p/	environm ents	/b/	environm ents	/t/	environm ents	/d/	environm ents	/k/	environm ents	/g/
#_o	<i>p̣oðari</i>	‘#_a	<i>ḡatsi</i>	‘f_i	<i>f̣tira</i>	n_r	<i>anḡra</i>	‘#_a	<i>ḡada</i>	ṇ_o	<i>f̣eṇḡo</i>
o’_s	<i>op̣sari</i>	#_e	<i>ḡedino</i>	o’_a	<i>poṭami</i>	f_o	<i>lifḡokari</i>	s_o	<i>sḡoletsi</i>	e_a	<i>ineḡa</i>
‘#_s	<i>p̣somin</i>	‘#_o	<i>ḡorda</i>	u_e	<i>havuṭe</i>	a_a	<i>provaḡa</i>	#_a	<i>ḡariḡi</i>	‘š_i	<i>šḡilo</i>
s’_i	<i>osp̣idi</i>	#_a	<i>ḡarxari</i>	a_u	<i>kaṭu</i>	e_i	<i>beḡino</i>	‘#_s	<i>ḡsilo</i>	i_o	<i>ligos</i>
‘a_s	<i>ap̣simo</i>	‘#_o	<i>ḡola</i>	x_e	<i>eyanaxṭesa</i>	‘a_a	<i>kaḡa</i>	o’_a	<i>liftoḡari</i>	r_o	<i>argos</i>
#_o	<i>p̣otami</i>	e_u	<i>ḡbuga</i>			a_o	<i>ṭšöraḡo</i>	‘#_a	<i>ḡatu</i>	a_’r	<i>magra</i>
‘#_r	<i>p̣rovada</i>	‘#_a	<i>ḡaleo</i>			‘#_i	<i>ḡiri</i>	u’_a	<i>ebuḡa</i>	‘#_o	<i>ḡolizo</i>
‘#_i	<i>p̣ios</i>	‘i_a	<i>iḡa</i>			i_i	<i>osp̣iḡi</i>	e_o	<i>eḡoliya</i>		
o’_i	<i>opi</i>	‘#_i	<i>ḡiso</i>			r_a	<i>borḍa</i>	‘#_a	<i>ḡahume</i>		
a’_e	<i>ap̣es</i>					a_i	<i>alaḡi</i>	‘#_r	<i>ḡruyo</i>		
e’_a	<i>eḡan</i>					e_a	<i>xameleḡa</i>				
e’_e	<i>eḡeran</i>					a_o	<i>ḡomaḡo</i>				
a’_l	<i>haṭsaplan</i>					#_r	<i>ḡrano</i>				
e’_i	<i>eḡinasa</i>					u_e	<i>vuḡero</i>				



**Table B.3A:** Distributional patterning of nasals

[n]	[ŋ]	[m]
‘#_i	e_g	‘#_a
‘a_d	e_g	#_a
‘a_e	æ_k	a_’n
‘o_e	i_g	a_e
#_a	#_g	a’_i
#_d		e_a
#_e		e_o
a_a		i_o
i’_e		o_a
m’_o		o_o
o_#		o’_i
o_a		r’_i
o_d		
æ_i		

**Table B.3B:** Distributional patterning of nasals with examples

environments	/n/	environments	/ŋ/	environments	/m/
æ_i	<i>gæni</i>	æ_k	<i>gæŋksæ</i>	#_a	<i>malia</i>
a_a	<i>mana</i>	e_g	<i>fɛŋgo</i>	e_a	<i>jema</i>
‘#_i	<i>nife</i>	e_g	<i>jɛŋgi</i>	‘#_a	<i>mana</i>
‘o_e	<i>onema</i>	i_g	<i>niŋgorona</i>	a’_i	<i>mami</i>
m’_o	<i>kamno</i>	#_g	<i>ŋgozo</i>	e_a	<i>onema</i>
#_e	<i>nero</i>			o’_i	<i>psomin</i>
‘a_e	<i>anemo</i>			o_o	<i>rðomo</i>
o_d	<i>ðondæ</i>			i_o	<i>apsimo</i>
o_a	<i>ɣonado</i>			a_’n	<i>kamno</i>
#_a	<i>na</i>			o_a	<i>xoma</i>
i’_e	<i>ineka</i>			r’_i	<i>ormi</i>
‘a_d	<i>andra</i>			e_o	<i>anemo</i>
o_#	<i>provadon</i>			a_e	<i>xameleda</i>
#_d	<i>ndo</i>				

**Table B.4A:** Distributional patterning of the tap and approximants

[r]	[l]	[j]
‘#_ð	‘#_i	‘#_e
‘au_o	‘a_a	‘t_a
#_a	‘i_o	‘t_e
#_i	#_i	(#_a)
a_’ð	a_#	(#_o)
a_g	a_a	#_i
a_i	a’_a	a’_i
a’_i	a’_e	i_a
d_a	a’_i	i_e
e_i	e_’f	l’_e
e’_o	e_i	n_e
i_a	i_a	
i_i	i_o	
o_’m	o’_e	
o_d	u_’j	
o’_a	v_e	
v_e	ʏ_i	
	ʏ_o	

**Table B.4B:** Distributional patterning of the tap and approximants with examples

environments	/r/	environments	/l/	environments	/j/
v_e	<i>alevre</i>	a'_i	<i>malia</i>	'#_e	<i>jema</i>
e'_o	<i>nero</i>	e_i	<i>yarðeli</i>	i_a	<i>vizija</i>
o'_m	<i>ormi</i>	e'_f	<i>aðelfo</i>	#_i	<i>jinega</i>
a_i	<i>poðari</i>	i_o	<i>skilo</i>	l'_e	<i>xuljeri</i>
a'_ð	<i>karðia</i>	'#_i	<i>ligos</i>	i_a	<i>skilija</i>
'au_o	<i>auros</i>	o'_e	<i>skoletši</i>	a'_i	<i>faji</i>
a'_ð	<i>yarðeli</i>	a'_e	<i>alevre</i>	i_e	<i>epije</i>
i_i	<i>tsiri</i>	'a_a	<i>alas</i>	#_a	<i>ja to pio</i>
d_a	<i>andra</i>	a_a	<i>yala</i>	#_o	<i>jok</i>
a_g	<i>argos</i>	#_i	<i>liftokari</i>	't_e	<i>eftijej</i>
i_a	<i>ftira</i>	u'_j	<i>xuljeri</i>	't_a	<i>eftjayo</i>
i_i	<i>diri</i>	a'_a	<i>aladi</i>	#_a	<i>jan</i>
a'_i	<i>kariði</i>	i_o	<i>ksilo</i>	'n_e	<i>njezma</i>
'#_ð	<i>rðomo</i>	i_a	<i>fila</i>		
o_d	<i>borda</i>	'i_o	<i>ilos</i>		
o'_a	<i>xorafi</i>	a_#	<i>bal</i>		
e_i	<i>xuljeri</i>	y_o	<i>ylossa</i>		
#_i	<i>rizi</i>	v_e	<i>avlea</i>		
#_a	<i>rasi</i>	y_i	<i>ylitsi</i>		

**Table B.5A:** Distributional patterning of fricatives [f], [v], [θ], [ð]

[f]	[v]	[θ]	[ð]
#_a	#_a	#_a	‘#_a
‘#_e	#_i	‘#_e	‘#_
‘#_i	‘#_r	a_#	‘#_o
‘#_t	‘#_u	a’_a	‘#_r
#_u	a_’l	a_i	a_#
a_i	a_u	a_o	a’_a
i_e	e’_r	n_r	a_e
i_t	o’_a	o_i	a_i
l’_o		r_e	e’_i
o’_a		l’_a	i_i
			o’_a
			r’_e
			r’_i
			u_i

**Table B.5B:** Distributional patterning of fricatives [f], [v], [θ], [ð] with examples

Environment	/f/	Environment	/v/	Environment	/θ/	Environment	/ð/
‘#_e	<i>feso</i>	#_i	<i>vizija</i>	l’_a	<i>liθari</i>	‘#_i	<i>ðio</i>
‘#_e	<i>fengo</i>	o_a	<i>provada</i>	a’_a	<i>maθano</i>	o’_a	<i>poðari</i>
l’_o	<i>aðelfo</i>	o’_a	<i>ova</i>	‘#_e	<i>θelo</i>	r’_i	<i>karðia</i>
i_e	<i>nife</i>	e_r	<i>alevre</i>	o_i	<i>eyomoθi</i>	r’_e	<i>yarðeli</i>
‘#_t	<i>ftira</i>	#_r	<i>vroši</i>	r_e	<i>erθe</i>	e’_i	<i>peði</i>
i_t	<i>liftokari</i>	a_u	<i>avude</i>	#_a	<i>θadia</i>	a_e	<i>aðelfo</i>
a_i	<i>xorafî</i>	#_a	<i>vari</i>	a_i	<i>maθizo</i>	a_i	<i>tsubaði</i>
‘#_i	<i>fila</i>	a_’l	<i>avlea</i>	‘#_e	<i>θeno</i>	i_i	<i>kariði</i>
#_u	<i>furno</i>	‘#_r	<i>vrað</i>	n_r	<i>anθrobo</i>	‘#_r	<i>ðromo</i>
o’_a	<i>tšofal</i>	‘#_u	<i>vudero</i>	a_#	<i>parakaθ</i>	‘#_o	<i>ðondæ</i>
#_a	<i>faji</i>	e’_r	<i>evriško</i>	a_o	<i>kalaθo</i>	‘#_a	<i>ðaxtila</i>
						u_i	<i>stuði</i>
						‘#_i	<i>ðiyo</i>
						a_#	<i>vrað</i>
						a’_a	<i>haða</i>

**Table B.6A:** Distributional patterning of fricatives [s], [z], [ʃ], [ʒ]

[s]	[z]	[ʃ]	[ʒ]
‘#_t	#_o	‘#_e	a_a
‘k_i	i_i	‘#_k	e_i
‘p_a	i_o	‘#_o	
‘p_o	i’_i	‘t_i	
‘t_i		a_a	
#_k		a’_i	
a_#		e_i	
k_ae		o’_i	
o_#		t_#	
o_a		t_o	
o’_a			

**Table B.6B:** Distributional patterning of fricatives [s], [z], [ʃ], [ʒ] with examples

Environment	/s/	Environment	/z/	Environment	/ʃ/	Environment	/ʒ/
o'_a	<i>kosara</i>	#_o	<i>zo</i>	'#_o	<i>šoni</i>	a_a	<i>haʒa</i>
a_#	<i>kadas</i>	i'_i	<i>rizi</i>	o'_i	<i>vroši</i>	e_i	<i>dreʒis</i>
'p_a	<i>opsari</i>	i_o	<i>kolizo</i>	'#_k	<i>škilo</i>		
#_k	<i>skoletši</i>	i_i	<i>vizija</i>	'#_k	<i>škini</i>		
'p_o	<i>psomin</i>	i_o	<i>spalizo</i>	a'_i	<i>raši</i>		
a_#	<i>alas</i>	i_o	<i>anizo</i>	a_a	<i>raša</i>		
'k_i	<i>ksilo</i>	i_o	<i>spundžizo</i>	'#_e	<i>šeri</i>		
o_#	<i>ilos</i>			e_i	<i>eši</i>		
o_a	<i>ylossa</i>			e_i	<i>inetši</i>		
'#_t	<i>stoma</i>			't_i	<i>tširi</i>		
k_ae	<i>gönksæ</i>			t_o	<i>tšofal</i>		
o_#	<i>auros</i>			't_i	<i>eiši</i>		
't_i	<i>batsi</i>			t_#	<i>utš</i>		



**Table B.7A:** Distributional patterning of fricatives [x], [ɣ], [h]

[x]	[ɣ]	[h]
‘#_r	‘#_a	‘#_a
‘e_o	‘#_l	#_a
#_a	‘#_o	a_u
#_o	‘a_o	a’_a
#_u	#_a	
a_t	a_o	
e_’t	e_o	
e’_a	e’_o	
o_n	u_o	
r’_a		

**Table B.7B:** Distributional patterning of fricatives [x], [ɣ], [h] with examples

Environments	/x/	Environments	/ɣ/	Environments	/h/
#_o	<i>xorio</i>	‘#_a	<i>ɣala</i>	a_u	<i>kahume</i>
#_u	<i>xuljeri</i>	#_a	<i>ɣarðeli</i>	‘#_a	<i>hadžaplan</i>
#_o	<i>xoraŋi</i>	‘#_l	<i>ɣlitši</i>	‘#_a	<i>havudo</i>
#_o	<i>xoma</i>	‘#_o	<i>ɣonado</i>	#_a	<i>haða</i>
r’_a	<i>barxari</i>	‘#_o	<i>ɣomado</i>	‘#_a	<i>hais</i>
#_a	<i>xameleda</i>	‘#_l	<i>ɣlossa</i>	a’_a	<i>aðaha</i>
a_t	<i>ðaxtila</i>	‘a_o	<i>aɣome</i>		
e’_a	<i>exaθe</i>	e’_o	<i>eɣo</i>		
‘#_r	<i>xrono</i>	u_o	<i>kruɣo</i>		
o_n	<i>doxna</i>	e_o	<i>leyo</i>		
‘e_o	<i>exo</i>	a_o	<i>layo</i>		
e_’t	<i>extes</i>				

**Table B.8A:** Distributional patterning of affricates

[ts]	[tʃ]	[dʒ]	[ks]	[ps]
‘e_#	‘#_i	i_e	‘#_e	‘#_i
#_a	‘#_oe	n’_i	‘#_i	‘#_o
#_u	#_e	u_e	a_i	‘a_i
a’_i	#_i		n_ae	i_o
	#_o			o_e
	a_e			o’_a
	e_#			
	e_i			
	e’_i			
	i_i			
	o_i			
	u_#			

**Table B.8B:** Distributional patterning of affricates

Environmen ts	/ts/	Environmen ts	/tʃ/	Environmen ts	/dʒ/	Environmen ts	/ks/	Environmen ts	/ps/
a'_i	<i>batsi</i>	e_i	<i>inetši</i>	n'_i	<i>spundžizo</i>	‘#_i	<i>ksilo</i>	‘#_o	<i>psomi</i>
#_u	<i>tsubadi</i>	‘#_i	<i>tširi</i>	u_e	<i>pudžega</i>	a_i	<i>aksinari</i>	‘a_i	<i>apsimo</i>
a'_i	<i>katsimale</i>	i_i	<i>litši</i>	i_e	<i>epidže</i>	‘#_e	<i>ksero</i>	o'_a	<i>opsari</i>
#_a	<i>tsakono</i>	e_i	<i>skoletši</i>			n_æ	<i>gænksæ</i>	i_o	<i>ðipso</i>
‘e_#	<i>aets</i>	‘#_æ	<i>tšærado</i>					o_e	<i>opse</i>
		i_i	<i>ylitši</i>					‘#_i	<i>psilo</i>
		#_i	<i>tšimume</i>						
		#_o	<i>tšofal</i>						
		e'_i	<i>etši</i>						
		a_e	<i>haðatšeka</i>						
		u_#	<i>utš</i>						
		#_e	<i>tšelia</i>						
		o_i	<i>kotšino</i>						
		u_#	<i>mutš</i>						
		e_#	<i>stetš</i>						

## Appendix C: Token frequencies

**Table C.1:** Token frequencies per word class in the Romeyka corpus

Word class	Sub-categorization	Gloss in ELAN corpus	Token number	Comments	Total token frequency per word class
Adjectives	Attributive adjectives	adj.a	55		230
	Predicative adjectives	adj.p	175		
Adverbs	Adverbs of manner	adv.m	354		1,193
	Adverbs of place	adv.p	424		
	Adverbs of time	adv.t	415		
Pronouns	Personal pronouns	pn	264		1,382
	Possessive pronouns	poss	318	clitic and full	
	Object pronoun	opn	343	clitic and full	
	Demonstrative pronouns	dem	111		
	Interrogatives	interr	189	incl. particle <i>mi</i>	
	Relativizer	rel	36		
	Other pronouns	pron	121		
Quantifier	Quantifier	quant	93		93
	Quantifying noun	nq	60	counted as nouns	
Numeral		num	147		147
Articles	Definite articles	det	1,314	in total	1,397
	Fem. definite article	f.det	191	exclusively F marked, unspecified for number, case	
	Masc. definite article	m.det	116	exclusively M marked, unspecified for number, case	
	Neut. definite article	n.det	298	exclusively N marked, unspecified for number, case	
	Pl. definite articles	det.pl	330	unspecified for gender, case	
	Indefinite articles	indet	83	unspecified for gender, case	

Nouns	Feminine nouns	f.n	273	unspecified for number, case	1,326
	Masculine nouns	m.n	170	unspecified for number, case	
	Neuter nouns	n.n	643	unspecified for number, case	
	Singular forms	n	826	unspecified for gender, case	
	Plural forms	n.pl	500	unspecified for gender, case	
	Diminutive forms	dim	6		
	Proper names	nn	158	unspecified for gender, case	
	Nominalizations	nmz	36		
Verbs	Present tense forms	v.pres	1,221		2,811
	Aorist past tense forms	v.aor	1,048		
	Imperfective past tense forms	v.ipfv	376		
	Imperatives	imp	71		
	Other Verbforms	v	95	often unclear cases	
Particles	Modal particle	fut	269	here glossed PRT	1,032
	Conditional particle	cond	20		
	Optative particle	subj	35	here glossed OPT	
	Topicalizer	foc	170	here glossed TOP	
	Negation particle	neg	305		
	Discourse particle	dp	125		
	Other particles (e.g., answer particle)	prt	108	could be largely subsumed under discourse particles	
	Interjections	interj	111	could be largely subsumed under discourse particles	
Conjunctions		conj	369		369
Prepositions	Instrumental	instr	53		656
	Beneficial	ben	5		
	Comitative	com	25		
	All other prepositions	prep	573		

## Appendix D: Spatial adverbials

**Table D.1:** Tentative overview of spatial adverbials in the Romeyka corpus

	singular	plural	ha-	-ha	with preposition	with special localizers					
						-tšega			-tšes		-merea
							with h(a)-	with -ha		with h(a)-	
‘here’	aða		h-aða h-aža	(a)ða-ha eða-ha	(h-)ap-aða h-aða-bahen ap-aða-dž-an h-ab-aða-ha h-aða-beškes daha-buko/ðaha-buko	ada-dž(ega) aða-džega a-džega	h-aða-džega ha-džega	ada-ha-džega aða-ha(n)-džega	aða-tšes aða-bahen-djes	h-aða-tšes	
					ab-aða-džega						
					h-ab-aða-džega						
‘yonder’	adu				adu-tšan (h-)ab-adu-džan	adu-džega	h-adu-džega				
‘there’	(e)tši adž(i)	etši(e)s	h-atši	(a)dža(-ha) edža(-ha) adže-ha (e)dže-ha	eb-edži h-ab-adži ab-adža(n) eb-edž(a) (h-)ab-aržan ha-tši-bukahen adži-bera(n) h-atši-peran ab-adž-ablan eb-edža-ha eb-edž-ega eb-edž-eberan adžijan	adža-džega e-džega etšiega itšiega	ha-tšijega	adžes-edžes			
								h-ab-a-džega h-ab-adža-han-džega			

'this'	<i>avudo(s)</i>	<i>(h)avu- da (h)avu- des</i>	<i>havu-</i>		<i>avu-dijega avu-džega</i>	<i>havu-džega hau-džega</i>
others					<i>apahan-džega apohon-džega tšabadrižga</i>	
				<i>dika doika</i>	<i>di-džega</i>	
				<i>aža alatška ebahen</i>		
simple adverbs				<i>(a)ka/(e)ka epuka-hen aban/(e)ban(a) opis ap-opis ebeban eksu ep-eksu katu(-haž) (((e)(b)e)buka/bugo tšan (a)peran/epera(n) (h-)apes abeškes</i>	<i>epan-džeka</i>	
interrogative				<i>pu-θen budžedža bešejeva beškes kapoθen</i>	<i>pu-džega buθen-džega</i>	

*havu-de-džes*

*di-džes  
do-tšes  
eyava-džes*

*(e)ban-džes  
buga-džes  
(e)buka-han-džes*

*eksu-  
merea*