Progressives in use and contact

A descriptive, areal and typological study with special focus on selected Iranian languages

Ghazaleh Vafaeian

Academic dissertation for the Degree of Doctor of Philosophy in Linguistics at Stockholm University to be publicly defended on Friday 15 June 2018 at 10.00 in Hörsal 9, hus D, Universitetsvägen 10 D.

Abstract

Progressives are grammatical patterns primarily used to refer to events that are ongoing at a specific time. This thesis investigates uses of such patterns in a number of languages as well as the interaction of a number of progressives in contact. The dissertation includes a typological study of the uses of 89 progressive patterns in two parallel corpora, an investigation of the uses and origin of the Persian *dāštan* progressive and an areal linguistic investigation of 50 Iranian varieties spoken around the Caspian Sea.

The dissertation presents features that increase the likelihood that a progressive is used. Such features are 1) a focalized (punctual) reference point, 2) the engagement or 'busyness' of the agentive subject on the event, 3) an emotive component and 4) the desire to turn the attention of the addressee towards an ongoing event. The significance of these features is expected to weaken as progressives grammaticalize.

There is a cross-linguistic tendency for progressives to occur more often with present time reference than with past time reference. In some cases, they are even restricted to the former. Among the varieties of the Iranian language Taleshi, on the other hand, we find asymmetric temporal paradigms as a consequence of former progressive patterns having expanded and lost their progressive character in the present but not in the past.

The study also shows that progressives are used differently in the present and the past: while events with present time reference often have the features mentioned above in 1-4, events with past time reference are often, although not exclusively, background contexts to other events pushing the narration forward.

The thesis also discusses various peripheral uses of progressives, such as uses in habitual and performative-like contexts, proximative, iterative and futurate uses, uses with stative verbs and temporary and subjective uses. Some of these tend to be found in patterns with higher frequencies and can be regarded as expansions towards the imperfective. Other uses are linked to the type of event to which the progressive applies: the proximative reading is shown to arise with achievements and the iterative use with repeated punctual events.

The data from the varieties of the Iranian languages Mazandarani, Gilaki, Taleshi and Tati, as well as from varieties under the influence of Persian, suggests that the progressive in these varieties is highly borrowable. Among the varieties discussed in Chapter 5, an areal cline is noted where constructional schemas used for ongoing events shift towards the imperfective. In the borrowing process, on occasion, a shift from progressive to proximative is also noted. As expected, the data from Caspian varieties shows that there are more progressive patterns than imperfective patterns.

Keywords: *progressive, use, contact, areal, parallel corpora, typology, grammaticalization, Iranian, Persian, Mazandarani, Gilaki, Taleshi, Tati.*

Stockholm 2018
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-155533

ISBN 978-91-7797-304-1
ISBN 978-91-7797-305-8

Department of Linguistics

Stockholm University, 106 91 Stockholm

PROGRESSIVES IN USE AND CONTACT

Ghazaleh Vafaeian

Stockholm University

Progressives in use and contact

A descriptive, areal and typological study with special focus on selected Iranian languages

Ghazaleh Vafaeian

ISBN print 978-91-7797-304-1
ISBN PDF 978-91-7797-305-8

Printed in Sweden by Universitetsservice US-AB, Stockholm 2018
Distributor: Department of Linguistics, Stockholm University

det föll
lite jord
i min hand

och handen föll
i min mun

och jag föll i glömska och sömn

Khashayar Naderehvandi,
Allting glittrar och ingenting tar slut

Contents

Contents ..... **i**

Acknowledgements ..... **vii**

List of Figures ..... **ix**

List of Tables ..... **xi**

Abbreviations ..... **xiii**

1 Introductory chapter **1**

1.1 Introduction 1

1.1.1 Thesis overview 2

1.1.2 Treated questions 3

1.2 Background 4

1.2.1 Critical notions 4

1.2.1.1 'Gram' and related concepts 4

1.2.1.2 Use, function, reading and context 5

1.2.1.3 Most favorable contexts and peripheral uses 5

1.2.1.4 Vendlerian taxonomy and Smith's binary oppositions 6

1.2.2 The progressive cross-linguistically 7

1.2.2.1 Describing the progressive 8

1.2.2.2 Focality 8

1.2.2.3 Imperfectivity 10

1.2.2.4 Progressives and temporal restrictions 13

1.2.2.5 Grammaticalization leading to progressives 13

1.2.2.6 Peripheral uses 15

1.2.3 Contact-induced change 19

1.2.3.1 Matter and pattern borrowing 20

1.2.3.2 Contact-induced grammatical change 22

1.2.3.3 Borrowability and stability 24

1.3 Glossings and language classifications 28

2 Progressive grams in parallel corpora **29**

2.1 The Parallel Bible Corpus 30

2.1.1 The corpus, gram set, method and methodology 31

2.1.1.1 The corpus 31

2.1.1.2 Identification of progressive grams 32
2.1.1.3 Comparative segment sets and 'direct speech' passages 37
2.1.1.4 Methodological issues 38
2.1.2 Data and analysis 39
2.1.2.1 Top segments in the present 42
2.1.2.2 Top segments in the past 43
2.1.2.3 The red, green and yellow cluster groups 44
2.1.2.4 Temporal restrictions and preferences 45
2.1.2.4.1 No temporal restrictions or preferences 47
2.1.2.4.2 Present progressive grams 48
2.1.2.4.3 Grams with a majority of present uses 50
2.1.2.4.4 Grams with a majority of past uses 51
2.1.2.5 'While' and the progressive 53
2.1.2.6 Grams with peripheral uses 56
2.1.2.6.1 Future segment set 57
2.1.2.6.2 Segment sets with performative-like uses 58
2.1.2.6.3 Habitual segment set 58
2.1.2.6.4 Stative segment set 59
2.1.2.6.5 Conclusion on peripheral uses 61
2.1.3 Summarizing section 2.1 61
2.2 TED corpus 62
2.2.1 The corpus, method and methodology 62
2.2.1.1 The corpus 62
2.2.1.2 Method, gram set and methodological issues 63
2.2.2 Data and analysis 66
2.2.2.1 Present time reference 69
2.2.2.2 Past time reference 72
2.2.2.3 Secondary gram set 74
2.2.3 Summarizing section 2.2 77

3 The *dāštan* progressive in Persian 79
3.1 Introduction 79
3.1.1 Morphology and syntax of the *dāštan* progressive 80
3.1.2 The *dāštan* progressive and other periphrastic patterns 83
3.2 The Progressive Questionnaire (PROGQ) 86
3.2.1 Method and methodological issues 86
3.2.2 Data and analysis 87
3.2.2.1 Most favorable contexts for the *dāštan* construction 92
3.2.2.2 Other patterns with ongoing uses 95
3.3 The *dāštan* construction in five Iranian movies 100
3.3.1 Method 100
3.3.2 Data and analysis 100
3.4 Discussion 102

3.4.1 The *dāštan* construction in present and past tense ... 103
3.4.2 Peripheral uses of the *dāštan* construction ... 106
3.4.2.1 Uses in DUR contexts ... 106
3.4.2.2 Proximative use ... 109
3.4.2.3 Futurate use ... 113
3.4.2.4 Iterative ongoing use ... 115
3.4.2.5 Absentive contexts ... 116
3.5 Concluding Chapter 3 ... 117

4 Cross-linguistic comparison and explanations ... 119
4.1 *Dāštan* construction and event types ... 119
4.2 Typological probe ... 120
4.3 Temporariness and subjective uses ... 128
4.4 The negation restriction of the *dāštan* gram in cross-linguistic comparison ... 136
4.5 Summarizing Chapter 4 ... 138

5 Caspian progressives in contact ... 141
5.1 Introduction ... 141
5.1.1 Two gram familes ... 141
5.1.2 Method and practical issues ... 143
5.2 Mazandarani, Gilaki, Tati and Taleshi ... 144
5.2.1.1 Classification ... 145
5.2.1.2 Sociolinguistic situation ... 147
5.2.1.2.1 Mazandarani ... 148
5.2.1.2.2 Gilaki ... 148
5.2.1.2.3 Taleshi ... 148
5.2.1.2.4 Tati ... 150
5.2.1.3 Verb morphology ... 150
5.3 The DAR and KAR gram families ... 153
5.3.1 The DAR gram family ... 154
5.3.1.1 Type 1: Mazandarani & Central Caspian ... 156
5.3.1.2 Type 2: Gilaki and Tati ... 158
5.3.1.3 Type 3: Taleshi ... 160
5.3.1.4 Type 4: Leriki [TAL] and Shuvi [TAL] ... 162
5.3.2 The KAR gram family ... 163
5.3.2.1 KAR: PROG patterns ... 165
5.3.2.2 KAR: PRS pattern ... 166
5.3.2.3 KAR: PROX pattern ... 166
5.3.3 The DAR & KAR or KAR-DAR patterns ... 167
5.3.3.1 DAR and KAR: PROG patterns ... 168
5.3.3.2 KAR-DAR: PROG patterns ... 168
5.3.4 Turkic, Neo-Aramaic, Nakh-Daghestanian and Tat ... 170
5.3.4.1 Turkic varieties ... 170

5.3.4.2 Neo-Aramaic and Tat ... 171
5.3.4.3 Note on Nakh-Daghestanian ... 175
5.3.5 Discussion and conclusion of section 5.3 ... 175
5.3.5.1 The DAR gram family ... 178
5.3.5.1.1 The areal cline ... 178
5.3.5.1.2 Taleshi DAR patterns ... 179
5.3.5.1.3 Origin of DAR gram family ... 179
5.3.5.2 The KAR gram family ... 181
5.3.5.3 Proximative patterns ... 182
5.3.5.4 Infinitival patterns ... 182
5.4 Progressive vs. present and past imperfective ... 184
5.4.1 Note on stative verbs ... 187
5.4.2 Presentation of patterns within the imperfective domain ... 188
5.4.2.1 Mazandarani ... 188
5.4.2.2 Gilaki ... 190
5.4.2.3 Tati ... 190
5.4.2.4 Taleshi ... 193
5.4.3 Discussion and conclusion on section 5.4 ... 195
5.5 Concluding Chapter 5 ... 198

6 The origin of the *dāštan* construction ... 201
6.1 Persian pattern modeled on Mazandarani pattern ... 203
6.2 Internal grammaticalization in Persian ... 206
6.3 Possessive sources for progressives ... 208
6.4 Concluding Chapter 6 ... 210

7 Summary and concluding discussion ... 213
7.1 Summary of findings ... 213
7.2 Temporal asymmetries ... 214
7.2.1 Temporal preference ... 215
7.2.2 Temporal restrictions ... 215
7.3 Uses of progressives ... 216
7.3.1 Most favorable contexts for progressives ... 216
7.3.2 Peripheral uses of progressive grams ... 218
7.3.3 Refined description of the progressive gram type ... 219
7.4 Grammaticalization of PROG to IPFV ... 220
7.5 Final remarks ... 222

Appendices ... 223
A. Grams in the Bible corpus ... 225
B. PROGQ ... 229
C. English translation of PROGQ ... 237
D. Occurrences of *dāštan* construction in five Iranian movies ... 243
E. List of varieties in Chapter 5 ... 247

F. Location of varieties in Chapter 5 251
G. Constructional schemas in Chapter 5 255

Sammanfattning på svenska 259

Bibliography 263

Acknowledgements

I did not plan on writing a dissertation in linguistics when I applied for the Bachelor's program at the Department of Linguistics at Stockholm University 10 years ago. Throughout the years, I've had most of the researchers at the department as my teachers, friends and later on also colleagues. I cannot imagine a better place to grow as a researcher. I will always be grateful to the people of this department for these years.

It is difficult for me to put into words the gratitude and admiration that I feel for my main supervisor, Östen Dahl, who has been with me throughout this process from the beginning to the end, being strict as well as constructive, supportive as well as demanding and spending countless hours explaining, discussing and answering my questions – always with encouragement. I truly feel privileged to have been Östen's student.

Likewise I am very grateful to my other supervisors, Agnes Korn and Maria Koptjevskaja Tamm. Agnes, who has always read my texts with careful precision and crucial insights. Masja, who has not only been one of the most inspiring teachers but also one of the most constructive critics of my texts. Matti Miestamo and Carina Jahani were also my supervisors for shorter periods in the beginning of the process. Thank you for inspiration and guidance.

I owe many thanks to Bernhard Wälchli who commented on an earlier draft of the thesis in his role as my 'mock-opponent', with many thoughtful suggestions for improvements.

I am thankful to FoSprak, the Special Doctoral Programme in Language and Linguistics, that gave me the opportunity to write this thesis, and to all the fellow doctoral student who have shared and enriched this experience.

Many researchers and informants have made this project possible. I will start by thanking Don Stilo, who has generously shared published as well as unpublished texts on patterns in varieties of Mazandarani, Gilaki, Taleshi and Tati, as well as answering numerous questions. I would also like to thank the twenty-six Persian informants who filled out my questionnaire, some of whom I also consulted for various nuances and readings of patterns in Persian. Many thanks to the people who have answered questions regarding progressives in other languages: Hatice Zora (Turkish), Raphaël Domange (French), Guillermo Montero Melis (Spanish), Francesca Di Garbo (Italian), Calle Börstell (Swedish), Jouko Lindstedt and Matti Miestamo (Finnish), Chingduang Yurayong (Thai), Steven Kaye (English) and Lena Forssén Renner, Richard Kowalik and Martin

Heinze (German). Thank you also René van den Berg (Muna), Guiti Shokri (Mazandarani), Murad Suleymanov (Azerbaijani), Pia Simper-Allen (Swedish Sign Language), Hassan Djamshidpey (Gilaki), Intan Fuji (Indonesian), Zaniar Naghshbandi (Hawraman) and Masud Rad (Central Kurdish) for discussions, explanations and language examples.

Thank you Robert Östling, Östen Dahl and Bernhard Wälchli for the parallel corpora and the innovative methods to use them. Also, thank you Östen for handling the data base and designing the maps in the typological study.

My teachers throughout the years, Ljuba Veselinova, Mikael Parkvall, Tove Gerholm, Henrik Bergqvist, Kina Nilsson Björkenstam, Eva Lindström, Mats Wirén, Emil Perder, Thomas Hörberg and Andrea Kiso. My present and past fellow doctoral students: Francesca Di Garbo, Yvonne Agbetsoamedo, Petter Kallioinen, Patric Klagsbrun Lebenswerd, Anna Jon-And, Hatice Zora, Signe Tonér, David Pagmar, Elisabet Eir Cortes, Richard Kowalik, Lena Forssén Renner, Calle Börstell, Johan Sjons, Robert Östling, Johannes Bjerva, Benjamin Brosig and Josefina Safar. Also Rickard Franzén, Matilda Löfstrand, Hanna Rönnqvist and all of the research assistants. Thank you for making the department such a pleasant work place.

For various administrative issues, thank you Sofia Gustafsson Capkova, Linda Habermann, Nada Djokic, Cilla Nilsson and Liisa Karhapää. Thank you Marcin Wlodarczak for help with the referencing program. Also Lamont Antieau and Ambika Kirkland for careful proofreading.

My roommates, Pernilla Hallonsten Halling and Susanne Vejdemo. Thank you for all of the advice, all of the talks and all of the help enduring the difficult parts of this process. Pernilla, who was there at the challenging period at the end. I don't know what I would have done without you.

Mina finaste vänner, Ellika, David, Petter, John och Jenny. Tack för pratet, för ett annat liv och andra tankar. Och lillebror Pouya, för att du alltid inspirerat.

Älskade Noa, älskade Avan. Ni är ljuset i mitt liv.

Martin. Med ditt lugn under mina mest kaotiska och minst stolta ögonblick. Jag hade aldrig kunnat göra detta om inte du hållit de små muš-muš:arna glada, nöjda, mätta, klädda, varje dag.

Och slutligen, mina föräldrar. För bärandet, dragandet, lyftandet och allt det andra som tagit mig hit. Tack.

List of Figures

Figure 1. Frequency of imperfective vs. perfective grams 12

Figure 2. List of all the segments in which the grams of the gram set occur 34

Figure 3. Areal distribution of the red, yellow and green clusters 44

Figure 4. Ratio of present time reference for progressive grams 46

Figure 5. Ratio of present time reference for progressive grams in direct speech 46

Figure 6. Temporal preference with darker red indicated higher ratio of present over past 46

Figure 7. Provinces of northern Iran 145

Figure 8. The Northwestern Iranian branch according to Glottolog 146

Figure 9. Population density in the area surrounding the Caspian Sea 147

Figure 10. Locations of Types 1, 2, 3 and 4 of the DAR gram family 154

Figure 11. Function and sub-categories in the KAR gram family 164

Figure 12. KAR-DAR or DAR and KAR constructions 167

Figure 13. Infinitival patterns in some Neo-Aramaic and Tat varieties 172

Figure 14. Iranian and non-Iranian patterns and functions 177

List of Tables

Table 1. Greek indicative verb forms in the New Testament (table compiled by Östen Dahl) 12

Table 2. Annotation of the segment 44009011$39 in English 32

Table 3. Annotation of the segment 44009011$39 in Amatlán Zapotec 32

Table 4. Esperanto present and past tense in top segments of the gram set 41

Table 5. Statistical data, at the cluster level and for the whole gram set 44

Table 6. Ratio of present over past 45

Table 7. The progressive grams of the primary gram set 64

Table 8. The grams of the secondary gram set 65

Table 9. Recall and precision for the primary gram set 67

Table 10. Frequencies of grams in the primary gram set in the corpus 69

Table 11. Frequency, recall and precision for the members of the secondary gram set run against the primary gram set 76

Table 12. Temporal, aspectual and modal oppositions of the *mi-* prefix 82

Table 13. Periphrastic patterns in present and past tense, and marking of negation 84

Table 14. Results of the PROGQ 89

Table 15. Uses of the *dāštan* construction in the PROGQ 91

Table 16. Occurrences of the *mašġul* and *dar hāl=e* constructions in PROGQ 98

Table 17. The five Iranian movies used in the studies 100

Table 18. Functional tags of the results from five Iranian movies 101

Table 19. Present or past tense in the five Iranian movies 101

Table 20. Grams included in the typological probe 121

Table 21. Proximative, iterative and futurate uses 128

Table 22. Hypothesis of uses that arise when progressives apply to different types of events 139

Table 23. Function and structure of DAR constructions 155

Table 24. Schemas in the KAR gram family including KAR & DAR schemas ... 164

Table 25. Short version of Table 23 179

Table 26. Infinitival schemas used for imperfective and progressive 183

Table 27. Constructional schemas used in the imperfective domain 196

Table 28. Re-production of table of results from Nematollahi (2014) 203

Table 29. The grammaticalization of PROG to IPFV 221

Abbreviations

-	inflectional boundary
=	clitic boundary
1	first person
2	second person
3	third person
ABL	ablative
ABS	absolutive
ABSN	absentive
ACC	accusative
ADD	additive
ADV	adverbial
AOR	aorist
APPL	applicative
AUG	augment
AUX	auxiliary
ASS	assertive
ASSOC	associative
ART	article
AV	actor voice
AVER	avertive
CAUS	causative
CL	classifier
CLM	clause linkage marker
CM	class marker
CMPL	complementizer
CNJ	conjectural
CONT	continuative
COP	copula
CS	category suffix
DAT	dative
DEF	definite
DEM3	demonstrative root 3
DEMD	demonstrative (distal)
DEMP	demonstrative (proximate)
DET	determiner
DIR	direct object
DIST	distal deixis
DUBIT	dubitative
DUR	DUR
EMPH	emphatic
ERG	ergative
EZ	ezafe
F	feminine
FIN	finite
FNOM	factive nominative
FOC	focalized
FOCS	focus
FOCZ	focalizer
FUT	future
FUTR	futurate
GEN	genitive
GER	gerund
GM	generalizing modality
GRAD	gradual
HAB	habitual
HYP	hyperbolic
I	stem I
II	stem II
IMPERS	impersonal
INC	incompletive
IND	indicative
INDF	indefinite
INF	infinitive

INFM	infinitive marker
INS	instrumental
INT	intensification
INTENT	intentional verbal suffix
IPFV	imperfective
IO	indirect object
IOP	indirect object (proximate)
IRR	irrealis
ITER	iterative
LIG	ligature
LNK	linking item
LOC	locative
M	masculine
MG	my glossing
MO	oblique case marker, masculine
MOD	modal
NC	noun compound
NEG	negation
NEC	necessitative
NOM	nominative
NONS	nonspecific
NON-FIN	non-finite
NMLZ	nominalizer
NPST	non-past
OBJ	object
OBL	oblique case
PASS	passive
PC	pronominal clitic
PFV	perfective
PL	plural
PN	person-number
POSS	possessive
PREF	prefix
PREP	preposition
PRF	perfect
PRN	proper noun
PROG	progressive
PROX	proximative
PRS	present
PRX	proximity
PSNT	presentative
PST	past
PTCP	participle
PV	patient voice
PVB	verbal prefix
Q	interrogative particle
SBJV	subjunctive
SG	singular
SRCE	source
SUB	subordinator
SUBJ	subject
SUF	suffix
SUP	supine
SS	same subject switch reference marker
STAT	stative
STATS	subject of a stative verb
TAM	tense, aspect, mood
TR	transitive
TRI	trial
VPR	vivid present tense

1 Introductory chapter

1.1 Introduction

Some languages of the world have a special pattern indicating that a dynamic event is ongoing at a specific time. These patterns are the focus of this thesis and referred to as *progressives.* For example, English has a progressive pattern formed with the copula followed by the verb in the gerund form, as in the sentence *I am working*. Other languages, such as Persian, Swedish and Spanish, also have such patterns, as illustrated in 1:1. Example 1:1a) presents the Persian progressive formed with the verb *dāštan* 'have', followed by the verb in the imperfective. This pattern is mainly used in spoken language. In Swedish, the progressive is formed with the verb *hålla* 'hold'+ *på* 'on' + infinitive marker and the verb in the infinitive (or, less commonly, the preposition is followed by *och* 'and' and a finite verb), as given in b). In Spanish, the progressive is formed with the verb *estar* and the verb in the gerund form given in c). These patterns are dedicated or "specialized morphosyntactic device[s]" (Bertinetto et al. 2000a:520) that refer to events that are ongoing at the reference time.

1:1 'I am working' in Persian, Swedish and Spanish

a) Persian
dār-am *kār* *mi-kon-am.*
have.PRS-1SG work IPFV-do.PRS-1SG

b) Swedish
Jag *håll-er* *på* *å* *jobb-a.*
I hold-PRS on at work-INF

c) Spanish
Estoy *trabaj-ando.*
be.PRS.1SG work-GER

Not all languages have a dedicated pattern for this function. In Russian, which lacks a progressive, a plain present tense, as presented in 1:2, is the only natural translation for 'I am working' and 'I work'.

1:2 Russian

ja rabotaju.
I work:PRS.1SG
'I work/I am working.'

The overarching aim of this thesis is to look at progressive constructions in a number of languages and investigate the typical and less typical uses of such patterns, as well the borrowing of such patterns in contact situations, with focus on present and past time reference.

1.1.1 Thesis overview

This chapter includes background for some critical notions, a cross-linguistic survey of the progressive and an introduction to contact-induced change. Chapter 2 looks at the uses of 89 progressive patterns cross-linguistically. Although cross-linguistic studies have investigated progressives with regard to structure, grammaticalization and, to some extent, function, no study has investigated these patterns using parallel corpora. Using parallel corpora, the contexts that are shared among progressives, i.e. those in which most progressives occur, can be established and investigated. In addition, differences in use between progressives can also be investigated. As automatic methods for using parallel corpora have been developed at the Department of Linguistics at Stockholm University, this method was used to investigate different uses of progressives in this chapter.

Chapter 3 is an in-depth investigation of the various uses of the Persian *dāštan* progressive. The data for this enquiry was collected using a questionnaire and by selecting occurrences of this pattern in Persian movies. Persian was chosen since data could easily be obtained for this language and since I have some intuition for the uses of this progressive. Also, until now, a thorough investigation of the functions and uses of this pattern has not been done.

Based on the findings of Chapters 2 and 3, Chapter 4 suggests some explanations for the different uses and readings that arise when progressives are applied to different types of events or contexts cross-linguistically.

In Chapter 5, two types of progressive patterns found in 50 Northwestern Iranian varieties spoken around the Caspian Sea are examined. Thanks to the availability of detailed data on dialectal variation in this area, mainly regarding structure and to some extent also function, the behavior of these patterns in contact situations can be investigated. It is shown that these patterns are at different stages of grammaticalization. More specifically, this chapter is concerned with the shift from progressive to imperfective among these varieties.

The findings of Chapter 5 motivated a return to the *dāštan* progressive in Persian with regard to its origin. This is discussed in Chapter 6. Finally, Chapter 7 provides a general summary and discussion of the findings of this thesis.

While Chapters 2, 3 and 4 treat uses of progressive grams, Chapters 5 and 6 are concerned with progressives in contact. One of the main questions in the thesis is the differences in the use of progressives in the present and past, both regarding the contexts in which progressives occur, treated in Chapters 2 and 3, and the grammaticalization from progressive to imperfective, treated in Chapter 5. We will also in different ways be concerned with the typical as well as less typical uses of progressives.

1.1.2 Treated questions

The concrete questions treated in this thesis are given in 1:3.

1:3 Questions and the chapters in which they are treated.

- What are the most favorable contexts for progressives?	Ch. 2, 3, 7
- What uses other than 'ongoing at the reference time' do progressives have?	Ch. 2, 3, 4
- What uses arise when progressives combine with different event types?	Ch. 3, 4
- Are the uses of progressives in the present different from those in the past? If so, in what way?	Ch. 2, 3, 5, 7
- What temporal restrictions and preferences do we observe for progressives and how might these have arisen?	Ch. 2, 5, 7
- How is the notion of *focalization* related to and relevant for the progressive?	Ch. 2, 3, 4
- How did the progressive patterns in the Caspian area arise?	Ch. 5, 6
- Is the variation of the marking of progressive grams in various varieties in the Caspian area greater than that of present tense and past imperfective grams?	Ch. 5
- What changes do we observe as progressive patterns expand their uses and become more mature?	Ch. 4, 5, 7

1.2 Background

This section includes background and clarifications on how various notions will be used in the thesis. To begin with, some critical notions are presented, followed by a presentation of the progressive cross-linguistically. Section 1.2.3 is a survey of relevant theoretical notions regarding contact-induced change.

1.2.1 Critical notions

1.2.1.1 'Gram' and related concepts

A *gram* is a grammatical item with a specific form and a specific meaning and/or function (Bybee & Dahl 1989). A gram can have different encoding patterns. It can be a morpheme, e.g. the English plural morpheme *-s* is a gram. It can be a syntactic construction, such as the Persian *dāštan* progressive. The notion of *construction* thus partly overlaps with the notion of *gram* (Dahl & Wälchli 2016:328).[1] In the text, I will use the terms *pattern* and *construction* as synonyms for *gram*. I will use *constructional schema* to refer to a generalized representation of the structure (but not function) of one or several grams, e.g. COP V-INF.

Cross-linguistically, grams with similar uses tend to be found in several languages, such groups of grams constitute a *gram type* (Bybee & Dahl 1989:52; Dahl & Wälchli 2016:328). The progressive constructions in English, Persian, Swedish and Spanish given in the introduction are all members of the cross-linguistic *progressive gram type*.

A *gram family* is defined in Dahl (2000a:7–8) as a set of grams that have arisen through the same historical process, either as a result of shared ancestry or language contact.

> 'Gram family' is a somewhat vague term that I use for grams with related functions and diachronic sources that show up in genetically and/or geographically related groups of languages. To take one example, constructions formed with a verb meaning 'to go', with uses sometimes referred to as 'prospective', show up in a number of languages in Western Europe, both in the Germanic and the Romance group. Thus, the usefulness of the term 'gram family' is based on the tendency for grammaticalization processes to cluster areally and genetically. (Dahl 2000b:317)

As an example, the Spanish *estar* + GER progressive can be said to be included in the same gram family as other Romance languages having progressive patterns marked with *estar*/*stare* + GER, such as Italian, Portuguese and Catalan. *Estar* 'to

[1] They are, however, also different in the sense that *gram* refers to elements in grammatical systems, whereas *constructions* also include lexical patterns such as, e.g., concrete verbs and their arguments. There are also entities that are *grams* that may not be classified as *constructions* by all researchers, such as those expressed inflectionally (Östen Dahl, p.c.).

be' originates from Latin *stare* 'stand'. *Stand* is also found elsewhere as a component of progressive patterns, for example, in the Iranian language Tajiki, which has a pattern formed with the past participle of the main verb plus the perfect or pluperfect of the verb *istodan* 'to stand, be standing' (Perry 2005:223, 225). This pattern, however, is not a member of the *estar*/*stare* + GER gram family, since it has not arisen through shared ancestry, nor through contact with those languages.

1.2.1.2 Use, function, reading and context

When discussing the distribution of grams, it is often difficult to disentangle notions such as *use*, *function*, *reading* and *context*. Often, they are interchangeable, but they may also have specific nuances. In this thesis, I will make an effort to use these terms in the following way: *use* will be employed as a neutral way of referring to occurrences of a pattern; I try to restrict the term *function* to the more prominent uses expressing distinct meanings; *reading* focuses on the interpretative perspective and is thus more subjectively oriented; *context* is used to refer to the general environment in which progressives occur, as well as language material other than the grams discussed, i.e. either other elements in the clause or items outside the clause.

Often it is impossible to tell if a certain sense arises due to the meaning of a gram or due to the context in which it occurs or due to particular interpretations of a context. It is also difficult to decide which uses to assign to the function of a gram and which uses are more context dependent, since, as will become clear, all readings are dependent in different ways on something, be it the verbal meaning, the interpretation of the event, or more external contextual items or conditions. In addition, the meaning of a gram and the context in which it occurs influence each other in both directions: on the one hand, the use of a gram with a predicate may change the way we understand the event: on the other hand, contextual information may tint a gram's meaning if the gram occurs often enough in that type of context. So if, e.g., a progressive gram is used often enough in habitual context, the habitual meaning may start to be associated with that gram. This is a process expected to happen as progressives grammaticalize into imperfectives. This means that it is neither easy nor always appropriate to distinguish between a gram's meaning and the context in which it occurs. That being said, in order to better understand progressives, I will at times attempt to pull apart a gram's meaning from the meaning of the predicate or meanings available in the context.

1.2.1.3 Most favorable contexts and peripheral uses

In this thesis, I will refer to the *most favorable contexts* and the *peripheral uses* of progressive grams. The *most favorable contexts* refers to both contexts that are shared among several progressive grams, as in the investigation involving parallel corpora in Chapter 2, and those contexts in which most informants use a

progressive gram, as in Chapter 3, which presents questionnaire answers. *Peripheral uses* refer to uses other than those expressing the typical meaning of 'ongoing at the reference time'.

1.2.1.4 Vendlerian taxonomy and Smith's binary oppositions

In Zeno Vendler's classic article "Verbs and Times" (1957), the combinability of the English progressive with different verbs plays an important role in distinguishing his "time schemata" from each other. Likewise, a modified version of Vendler's taxonomy will prove itself useful in this thesis for the understanding of how progressives are used. Thus, I will be speaking of the predicate types *statives*, *activities*, *accomplishments*, *achievements* and *semelfactives.* In addition, I will use the terms 'state' for the entities denoted by statives and 'events' for the entities denoted by all the other predicate types.

Statives: examples of these are *know*, *love* and *believe*. Predicates of this type are characterized in that they denote states – stable situations with no inherent endpoint (Smith 1997:32). Also, unlike events, states continue as before, unless changed, and typically lack volition on the part of the subject (Vendler 1967:102–103, 106; Comrie 1976:13). Predicates like *being married* and *being healthy* or qualities like *hot* and *yellow* also denote states according to Vendler (1967:108). He distinguishes these predicates from the others by their inability to combine with the English progressive.

Activities: examples are *run* or *push a cart*. These denote events without a natural terminal point. A way of distinguishing these types of predicates from accomplishments is to employ a test: if someone is running or pushing a cart, and stops doing so in the next moment, it is true that (s)he did run or push a cart. The same cannot be said about accomplishments (Vendler 1967:100).

Accomplishments: these are exemplified by *running a mile* and *drawing a circle*, and denote events with an inherent terminal point, which entails that if the event of running a mile or drawing a circle is interrupted, unlike with activities, you cannot say that you ran a mile or drew a circle (Vendler 1967:100–101). Accomplishments "proceed towards a terminus which is logically necessary to their being what they are. Somehow this climax casts its shadow backwards, giving a new color to all that went before" (Vendler 1967:101–102).

Achievements: examples of achievements are *reaching the top* or *winning the race*. These are distinguished from accomplishments in that they occur at a single moment. Smith (1997:30) points out that they may also have a preliminary stage associated with them. Vendler (1967:104) distinguishes between accomplishments and achievements by noting that while *It took me one hour to write a letter* entails that I was writing a letter at every moment of that hour, *It took me one hour to reach the summit* does not entail that the reaching was going on for one hour. Similarly, Dahl (2013:72) distinguishes between achievements and accomplishments by giving the examples *finish a cake*, which is viewed as an

achievement, and *bake a cake*, which is viewed as an accomplishment. Both these phrases have the same result state but can be distinguished in that achievements consist of the final point of the process only, whereas accomplishments include the whole process leading up to the result. Vendler (1957:103) also notes that in English, these verbs in the present tense are most often used as historical presents or to indicate immediate future, as in *Now he finds the treasure*.

Also, inceptive events such as *begin* are often viewed as achievements or, alternatively, accomplishments. Thus, Smith (1997:25) notes that *Mary began to run* is an achievement "which presents a change of state into an [a]ctivity".

Semelfactives such as *knock at the door* and *flap a wing* (Smith 1997:29) refer to events that lack duration and an inherent result-state. They do not include a process leading up to the event. When used with the progressive, they may denote iterative events. (Semelfactives were not part of Vendler's taxonomy.)

Smith (1997:19, 22–35) provides a description of these types by using the binary oppositions *static-dynamic*, where *static* refers to 'an undifferentiated period' and *dynamic* to 'successive stages'; *atelic-telic* where *telic* events have a natural endpoint or goal or outcome, whereas *atelic* events do not; and *instantaneous-durative,* which have to do with whether a predicate is punctual or not. Thus, *activities* are dynamic, atelic and durative; *accomplishments* are dynamic, telic and durative; *achievements* are dynamic, telic and instantaneous; *semelfactives* are dynamic, atelic and instantaneous; and *statives* are static and durative (statives are undefined for telicity) (Smith 1997:3).

It is often noted that statives are different from the other predicate types, e.g. Dahl (2013:59) notes that tense and aspect systems cross-linguistically tend to treat statives differently from the rest. In addition, he notes that languages tend to lack a word that can refer to all four groups in the Vendlerian taxonomy. Both these observations indicate that statives do not create a natural class with the other predicate types.

Finally, it is important to note that the Vendlerian taxonomy, also including semelfactives, is typically used to refer to the interpretation of an event, or how we view, talk of or refer to an event, rather than being defined by a set of predicates or examples. At times, events are vague or ambiguous and can be viewed as, e.g., either (or both) an achievement or an accomplishment.

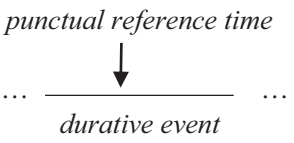
1.2.2 The progressive cross-linguistically

The main cross-linguistic studies on the progressive are Blansitt (1975), Comrie (1976), Dahl (1985), Bybee & Dahl (1989), Bybee et al. (1994) and Bertinetto et al. (2000a). In what follows, the most relevant cross-linguistic findings for this thesis are presented, starting with a description of the progressive gram type.

1.2.2.1 Describing the progressive

Comrie (1976:35) explains "progressiveness" as "the combination of progressive meaning with non-stative meaning". The sentences in which most progressives were given in Dahl's (1985:91) study indicated "what could be labelled an 'on-going activity'. 'To go on' is basically a relationship between a dynamic situation and a point in time". A similar observation is made by Bybee et al. (1994:126): "Progressive views an action as ongoing at reference time". Example 1:4 aims to illustrate 'ongoing at the reference time', where 'reference time' is the point of reference (Reichenbach 1947), in the case of the progressive, this is the point at which we view the ongoing event.

1:4 'Ongoing at the reference time'



I will restrict the term 'progressive' to refer to specific patterns, as in 'progressive gram', or to the gram type as in 'the progressive' or 'progressive gram type'. I will use the phrase 'ongoing at the reference time' to refer to the typical use of progressives, or the shorter but equivalent 'FOC ongoing', which includes the notion of focality, to which we turn below. I will also use the term 'ongoing', which is neutral with respect to the punctual or durative nature of the reference time.

1.2.2.2 Focality[2]

Bertinetto et al. (2000a) and Bertinetto (2000) use *focalized* or *focalized point* to refer to a punctual reference time, i.e. a specific time at which the event referred to is ongoing. For example, *She is working* means that the subject is working at the speech moment. Here the focalized point is implicitly understood. In other examples, it may also be explicitly given, as in *She is working right now* or *She is working at 2 o'clock*. I will abbreviate *focalized* to FOC and refer to such events as FOC *ongoing*. Progressive grams may also occur in what Bertinetto et al. (2000a) call *durative contexts*. In *durative* contexts, the event is related to an extended period of time, e.g. *She is working all day* or *She is working from 2 to 3*.

[2] *Focality* as a term is introduced by Johanson (2000:85): "[f]ocality concerns the concentration (focus) of the psychological interest on the situation obtaining at O" where O is the reference time. Johanson uses *focality* as applying to specific aspecto-temporal categories, i.e. a specific gram in a language, not necessarily a progressive gram, is analyzed as exhibiting high degree of focality or low degree of focality. This use of the term is somewhat different from the use in this thesis.

Relating durative, or DUR, contexts to 1:4, we could say that the reference time is extended in these contexts as opposed to being punctual. That is, while the reference time in *She is working at 2 o'clock* is the punctual '2 o'clock', the reference time in *She is working from 2 to 3* is the interval between 2 and 3. For the sake of simplicity, I will refer to all non-focal contexts as DUR contexts, including habitual contexts such as *She is working every week* as well as gradual processes.

Some progressives, such as the Italian progressive *stare* + GER, the French progressive *être en train de* + INF and the Albanian *po* progressive, are restricted to FOC contexts and cannot occur in DUR contexts (Bertinetto et al. 2000a:539). These are referred to as *focalized progressive constructions* (Bertinetto et al. 2000a:530). Some progressives, referred to as *durative progressive constructions*, "are most naturally interpreted in the durative meaning" (Bertinetto et al. 2000a:530). In Romance languages, for example, patterns formed with motion verbs are used in DUR ongoing contexts only (Bertinetto 2000:567, 577). Many progressive grams, however, occur in both FOC and DUR contexts, as seen in the English examples.

An expression such as *I am smoking* can also be uttered in a focalized sense even though the subject is not involved in smoking as she makes the utterance (perhaps she is just holding the package of cigarettes in her hand and is referring to her being busy or something similar). In discussing such uses, Johanson (2000:86) refers to focality as a relative notion that distinguishes 'presentness' from a broader temporal notion.

It is difficult to know if focality is part of the meaning of progressive grams, a context in which progressives most often occur, or a context that progressives typically require, especially since the FOC point is often implicit. Progressives can be seen as occurring in contexts that are FOC, but also as focalizing an utterance, i.e. disambiguating utterances with regard to focality. Compare, for example, *She reads* to *She is reading*, where the simple present sentence is understood as a general statement, while the progressive sentence is understood as referring to the speech moment. English is, however, different from many other languages that have progressive patterns in that the English simple present and past do not allow for the FOC ongoing meaning, that is, the progressive gram is the typical choice for a FOC reading. In other languages, such as, e.g., Persian, Swedish and Spanish, general presents and pasts may also include the ongoing reading, as they are ambiguous with regard to focality (see 1:5). The difference between the English progressive and progressives in other languages is most probably due to the English progressive being further grammaticalized, in that it has expanded its function and taken over the function of ongoingness in the present and past tenses.

> [The English progressive] is the only device available to express a focalized aspectual view at any temporal location (Past, Present or Future), while all Romance languages, even those where PROG appears to be fairly frequently employed, may freely resort […] to the Present, the Imperfect or the Simple Future, depending on temporal location. (Bertinetto 2000:565)

Most authors acknowledge the notion of focality in one way or another when talking of progressive, often using different terms. Thus, Dahl (1985:91) speaks of typical uses of progressives as involving "a relation between a dynamic situation and **a point in time**" (my emphasis). One illustrative way of capturing the focalizing feature of the English progressive is provided by Mittwoch (1988:233–234), who compares the meaning that arises when using the progressive to "a still from a movie picture". The FOC ongoing meaning is then like a snapshot of an ongoing event viewed at one moment.

> If we compare a stative sentence to an ordinary photograph and an activity sentence like *John run* (semelfactive) to a moving picture, we might say that the nuclear progressive, in capturing a moment of a changing situation, is like a still from a moving picture. (Mittwoch 1988:233–234)

In this dissertation, following Bertinetto et al. (2000a), FOC, i.e. *focalized*, refers to a single point in time, that is, the term FOC is synonymous with a punctual reference time. FOC *context* refers to contexts in which the event is viewed in relation to a punctual reference time, either explicitly given or implicitly understood. Also I will talk of FOC *use* which refers to uses of progressives that are viewed at a point in time. I will not make a distinction between FOC *progressive constructions* and *durative progressive constructions*, other than when I refer to the aforementioned authors. Throughout this thesis, I will often talk of 'FOC ongoing' or 'ongoing at the reference time'. I will, however, leave open the question of whether focalization is part of the meaning of progressive grams or part of the context.

1.2.2.3 Imperfectivity

The term 'imperfective' is ambiguous, as it can be used both of the functional domain, which includes grams such as present tense and past imperfective, as well as of the gram type 'imperfective', the uses of which cover the whole imperfective domain. An example of the latter is the Persian *mi-* verbal prefix. Comrie (1976:24) notes that imperfectivity as a semantic notion is characterized by the feature of viewing a situation from within, as opposed to perfectivity, which views a situation as a whole. In my understanding, this applies to progressives in that when they are used with an event that has a natural endpoint, the endpoint is disregarded, and the starting point of the event is also irrelevant. The imperfective nature of progressives is also evident in the illustration of progressives as a snapshot where a dynamic situation is viewed at a point in time, a FOC point,

resulting in the outcome of the event being irrelevant. Thus, focality is linked to the imperfective nature of progressives: in viewing a dynamic event at one point in time, we turn our attention away from any potential endpoint. This issue is connected to what Dowty (1977) calls the *imperfective paradox*, that is, how to explain that an example such as *John was drawing a circle* "entails that John was engaged in bringing-a-circle-into-existence activity but does not entail that he brought a circle into existence" (Dowty 1977:46).

As previously mentioned, in English, the progressive is the typical choice for events that are ongoing at the reference time, while in Persian, Swedish and Spanish, e.g., the present and past forms may also cover this function. In the latter cases, grammatical patterns that are described as present tense and past imperfective patterns can be seen as ambiguous for the FOC ongoing meaning in that they may or may not entail this reading, whereas a progressive gram, in comparison, disambiguates the utterance in this regard. Thus, a sentence that may have a general reading is narrowed to a reading that entails that something happened at reference time as seen in the Swedish example in 1:5.

1:5 Swedish present and past vs. progressive

a) *Jag rök-er.*
I smoke-PRS
'I smoke/I am smoking.'

c) *Jag rök-te.*
I smoke-PST
'I smoked/I was smoking.'

b) *Jag håll-er på att rök-a.*
I hold-PRS on INFM smoke-INF
'I am smoking.' (i.e. 'right now')

d) *Jag höll på att rök-a.*
I hold.PST on INFM smoke-INF
'I was smoking.' (i.e. 'right then')

The relationship between the present and past progressive examples in 1:5 is complicated by the uses of present and past tenses. First, events in the present are normally incomplete, while things that we speak of in the past are naturally, or more often, completed. Cross-linguistically, in the past, the frequency of perfectives is higher than the frequency of imperfectives. This is, for example, shown in Table 1, which presents Greek indicative verb forms in the New Testament. As can be seen, the two most frequent uses are the present and past perfective with the past imperfective being much less used. Also Janda &

Lyashevskaya (2011:723) show that in Russian past imperfective verb forms are much less common than both non-past imperfective and past perfective verb forms.

Table 1. Greek indicative verb forms in the New Testament (table compiled by Östen Dahl)

VERB FORM	NO. OF OCCURRENCES	PERCENTAGE
future	1605	10
present	5529	35
aorist (past perfective)	5875	38
imperfect (past imperfective)	1682	11
perfect	835	5
pluperfect	86	1

We can also assume that, typically, present time reference is more common in the spoken language than in written sources. The situation may, for example, look as the illustration in Figure 1 does, where the grey area indicates the imperfective domain, i.e. present and past imperfective.

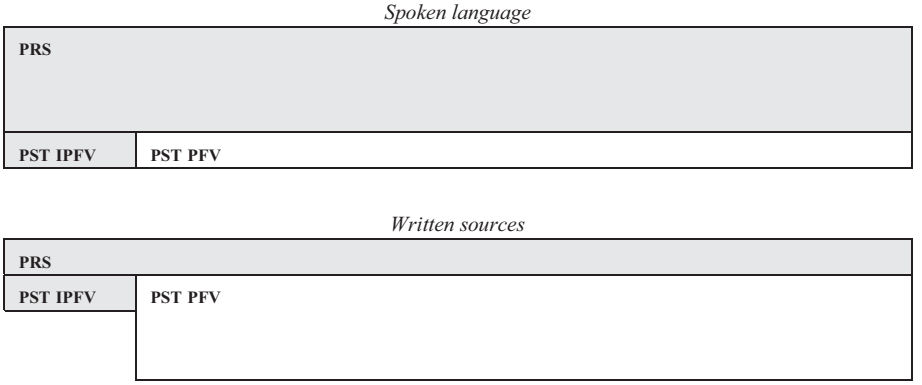


Figure 1. Frequency of imperfective vs. perfective grams

Utterances and sentences with past time reference often involve narratives. In the literature, it is often noted that imperfective grams are incapable of advancing the plot in narration, they are *non-propulsive* (see, e.g., Hopper 1979:239; Johanson 2000:76; Dahl 2013:65). This is because telic events are needed to bring change to the narration. In an example such as *he was reading in his chair, when the door opened*, the progressive clause provides a scene where no change takes place, whereas the telic event of the door opening changes the scene and thereby moves the plot forward. The type of context in which the event of reading is given is typically referred to as *backgrounding* and contrasted to *foregrounding* clauses.

Thus, while foregrounding clauses are narrated, backgrounding clauses support, amplify or comment on the foregrounded clauses (Hopper 1979:215). As change is needed to advance the plot, foregrounding in narration has been shown to be associated with perfectivity, while backgrounding has been shown to be associated with imperfectivity. More specifically, backgrounding contexts have previously been noted to be typical for progressives (Timberlake 2007:288).

> It is evidently a universal of narrative discourse that in any extended texts an overt distinction is made between the language of the actual story line and the language of supportive material which does not itself narrate the main event. I refer to the former – the part of narrative which relate events belonging to the skeletal structure of discourse – as FOREGROUNDING and the latter as BACKGROUNDING. (Hopper 1979:213)

1.2.2.4 Progressives and temporal restrictions

Dahl (1985:92–93) notes that, cross-linguistically, in contrast to perfective and imperfective grams, which have strong correlations with past and non-past time reference, respectively, the progressive tends to be independent of time reference and is used for both present, past and less frequently future time reference. Referring to grams that are restricted temporally, he states that "if there are restrictions, it is rather the present than the past that is favored with progressives" (Dahl 1985:93–94), an example being the Hawaiian (Austronesian) *ke* + V + *nei*, which is restricted to the present. In Bertinetto et al. (2000a:525–526), some progressive grams that are restricted to the past are also mentioned, for instance, in Hungarian, where word order and specific intonation contour of the clauses marks the progressive. Also, in Lithuanian the copula combines with the 'present active participle' to form past and future progressives, while the simple present tense is ambiguous for present and ongoing uses (Blansitt 1975:20).

Blansitt (1975:30) notes several tendencies of morphological biases towards the present tense, one of them being that there are never more tense distinctions in progressive than in non-progressive which is confirmed by Bertinetto et al. (2000a:526).

1.2.2.5 Grammaticalization leading to progressives

Cross-linguistically, progressives are often grammaticalized from locative sources. Bybee et al. (1994:136) propose that the meaning 'the subject is located in the midst of doing something' is the original meaning of progressive patterns. The elements given in 1:6 are explicitly or implicitly given in this original meaning. As patterns grammaticalize, components such as the requirement of an agent as well as the locative meaning weaken, making patterns appropriate for more contexts.

1:6 Elements included in the original meaning of progressives

a. An agent
b. is located spatially
c. in the midst of
d. an activity
e. at reference time

According to Bybee & Dahl (1989), no metaphoric extension is needed to explain the shift from locative to progressive since locative utterances imply temporal location.

> To be located spatially is to also be located temporally in an activity, so that from the beginning the meaning of such constructions has temporal implications. Gradually the locative meaning weakens while the temporal implications stabilize, giving rise eventually to the aspectual meaning of progressive. (Bybee & Dahl 1989:81)

Progressives may however also originate from other sources. A path of grammaticalization of iterative to continuative to progressive is, for example, hypothesized in Bybee et al. (1994:170). Also, elements such as motion verbs, 'hold' verbs, items denoting 'now', 'do' verbs, 'live, exist' verbs, reduplications or tonal differences have been shown to be sources for progressives (Blansitt 1975:9, 14; Bybee et al. 1994:131, 140, 142). Such patterns may then be assumed to have other paths of grammaticalization, some of which have not yet been properly investigated. Notably, although typically incapable to combine with stative predicates, progressive grams are often composed of stative elements themselves, such as locative elements, copula verbs and verbs of existence.

Progressives may further grammaticalize into becoming general imperfective markers. In its maturation, the progressive expands its function and takes over all functions of present and past imperfective constructions, such as occurring in habitual contexts, gaining the ability to have generic meaning and occurring with all types of predicates.

1:7 Typical grammaticalization path of progressives

LOCATIVE → PROGRESSIVE → IMPERFECTIVE

Building on Bybee (1985) and Dahl (1985), Bybee & Dahl (1989:77) note that progressive grams often combine with existing presents, pasts and futures, as well as being periphrastic with transparent lexical sources. Structurally, progressives are often marked periphrastically, as already seen in the given examples. The periphrastic nature of progressive constructions is linked to these constructions

normally being relatively young. Patterns at the beginning of grammaticalization clines have been shown to be more volatile (Dahl 2004:268), i.e. liable to change, as well as more borrowable between languages in contact (Moravcsik 1975:110; Dahl 2004:127). The relatively early stage of grammaticalization for progressives is also seen in that the original meanings of the components making part of progressive patterns are often still transparent, e.g. *dāštan* 'have' in the *dāštan* progressive construction, which is also used as a main verb in Persian.

In their investigation of progressive grams in European languages, Bertinetto et al. (2000a:530–531, 538–541) suggest a path of grammaticalization by which progressives originally occur in contexts that are DUR and develop to (also) occur in contexts that are FOC. For some grams, such as Italian *stare* + GER, the development has now reached a stage where it is almost exclusively used in FOC contexts. Their main source of historical data comes from Romance languages, but they also note that examples from Old English show the pre-progressive pattern in DUR contexts. For Romance languages, they suggest a five-stage path of development for the progressive grams, starting from locative DUR contexts, to FOC contexts and further to imperfective, where the last stage has not yet been obtained. They note however, that some of the grams that they study, such as French *être en train de* + INF and Albanian particle *po*, probably started out as progressives restricted to FOC contexts directly, without a previous DUR stage. FOC progressives developing into DUR progressives have not been attested.

In the literature, many examples of progressives that have developed or are on their way of developing into imperfectives are given (for some example, see Bybee et al. 1994:140–144). Sometimes, the grammaticalization from progressive to imperfective may lead to temporal asymmetries. In, for example, Kui (Dravidian), the present progressive is formed with the verb *man-* (originating from 'live, exist'). In the past, *man-* covers both habitual and ongoing function.

In the literature, agentivity[3] has been shown to be an important feature of progressive grams related to the different stages of grammaticalization. It is for example, both hypothesized to be a part of their original meaning (Bybee et al. 1994:136), a requirement that certain progressives can have (see, e.g., Bertinetto et al. 2000a:537, 542), and an important criteria in the expanded uses as shown for the English progressive (see, e.g., Ljung 1980).

1.2.2.6 Peripheral uses

In the literature, it has been noted that progressives are often restricted to uses involving dynamic predicates and that they do not normally occur in habitual contexts. Studies have, however, noted that some progressives can occur in less typical contexts, in which case less typical readings may arise. These uses may be somehow restricted and require specific criteria for their use. For example, some progressives can occur with certain stative predicates, as in 1:8a); some

[3] An agent is understood as typically human and volitional subject carrying out the event.

progressives can occur in contexts with an expanded period of time as in b); and some progressives can be used to mark that an event is just about to happen, as in c).

1:8 Less typical uses of progressives

a) English
He is being silly.

b) Persian
tamām=e *ruz* *dār-e* *kār* *mi-kon-e*.
all=EZ day have.PRS-3SG work IPFV-do.PRS-3sg
'S/he is working all day.'

c) Swedish
Han *höll* *på* *att* *raml-a* *ner*.
He hold.PST on INFM fall-INF down
'He was about to fall down.'

As already mentioned, I will refer to uses that do not involve the meaning 'ongoing at the reference time' as *peripheral*. Some peripheral uses are due to further grammaticalizations of progressive grams as they expand towards the imperfective, while others are remains of the original meaning of progressive constructions. In some cases it is uncertain how these uses have come into being. The Mandarin Chinese (Sino-Tibetan) *zài*, for example, marks ongoingness as well as having other meanings such as 'at', 'again (in future)' and 'only then' (Po-Ching & Rimmington 2004:6, 105–106, 141, 174, 237). Po-Ching & Rimmington (2004:106) assume the aspectual meaning to have originated from the locative use 'at'. The Turkish (Turkic) *-Iyor-* is used for ongoingness but combines also with stative predicates. Although referring to *-Iyor-* as a progressive, Kornfilt (1997:357) notes that the term 'continuous' would be more inclusive. Bybee et al. (1994:141) note that, in the spoken language, *-Iyor-* is also used habitually. Similarly, in Quechua *-sa-* (elsewhere given as *-sha-*) is used for ongoingness as well as with stative predicates (Dahl 1985:94). Such uses are rather seen as a further grammaticalization of progressive grams. More examples of patterns that mark ongoingness as well as having other uses are found in e.g. Indonesian (Austronesian), where *lagi* marks ongoingness but also means 'still, again, more, other' (Sneddon 2006), in Korean (Koreanic) and Japanese (Japonic), where the *-ko issta-* and the *-te iru* constructions, respectively, mark ongoingness as well as the resultative function, among other things (Kim 1986:98; Soga 1983:119).

Some peripheral uses depend on the progressive combining with other patterns in the language, such as e.g. the future progressive *will be* V-*ing* in English. Also, the Tajiki (Indo-European) progressive pattern, formed with *istodan* 'to stand, be

standing', has, in addition to a present and past form, also a non-witnessed past progressive form, a present progressive subjunctive form and a past progressive conjectural form (Perry 2005:224–239). All these patterns in turn have separate passive forms (Perry 2005:247–249).

Some progressive grams also cover the *proximative* and *avertive* function. Kuteva (2001:77, 92, 94) discusses *proximative* and *avertive* patterns cross-linguistically, where proximative is given as "a temporal phase located close before the initial boundary of the situation described by the main verb" or "being on the verge of V-ing" (2001:92, 94). The *avertive* is defined as "was on the verge of V-ing but did not V". Looking at descriptions of proximative and avertive, it becomes clear that they overlap in the past, so that a proximative gram can occur in the past or the present, while an avertive gram typically occurs in the past (Kuteva 2001:95). Notably, proximative and avertive uses are FOC.

Kuteva (2001:99–100), who is mainly concerned with avertive grams, notes that avertives combine the aspectual meaning of imminence, the temporal meaning of pastness and the modal meaning of counterfactuality. Often, the grammatical status of avertive patterns vary, so that some patterns are semantically more bleached than others, or some may be restricted to agentive uses only, while others are not. Generally, the avertive appears to be a rather young gram (Kuteva 2001:86). The meanings that grammaticalize into avertive patterns in Kuteva's (2001:86) study are 'be', 'want', 'love', 'sin', 'err', 'fail', 'miss', 'have' and 'go'. Some languages such as e.g. Koasati (Muskogean) have dedicated morphosyntactic devices for proximative and avertive readings. In Koasati, the combination of the intention suffix *a:hi-* and the dubitative marker *-má:m* form a proximative construction, while the suffix *-ápi* is the marker of the avertive (Kuteva 2001:97). Other languages may have a proximative pattern covering both functions, such as Spanish *a punto de*, the Italian *stare per* + INF or *sul punto di* + INF, for example. In Thai (Tai-Kadai), the progressive *kamlaŋ* in combination with the future marker *cà* + main verb marks the proximative (Smyth 2002:69). In Swedish, the rather infrequent *vara på vippen att* + INF pattern exists. In Swedish sign language, the proximative function can be marked with what is referred to as *initial stop*, where only the initial part of the sign denoting the event is given directly followed by a "stop", i.e. the rest of the event is not signed (Bergman 1983:4). Events marked in this way typically have a human agent and are intentional (Pia Simper-Allen, p.c.).

The proximative function may be covered by grams that have other main uses such as the progressive, as seen in 1:8c). Bertinetto et al. (2000a:534) note that this reading, which they name *imminential*, is a general feature of *focalized progressive constructions* in combination with achievement verbs (Bertinetto et al. 2000a:534). Also, Johanson (2000:153–154; 2017:31) mentions the proximative use of progressive grams and comments that progressive grams are often ambiguously denoting progressive and proximative meanings. For Persian, Dehghan (1972) points out that the *dāštan* progressive receives a special reading

with verbs like *mordan* 'to die', *xaste šodan* 'to get tired' and *gorosne šodan* 'to get hungry', namely that of an action that is just about to start. Jahani (2017) also notes this use of the *dāštan* progressive as well as the 'non-imminent'[4] future use. For this reason, she refers to this construction as the progressive/prospective. Progressives covering prospective and proximative uses are found in a number of Iranian and Turkic languages treated in Korn & Nevskaya (2017), e.g. Noorlander (2017:195–198) presents progressives with proximative uses in North Eastern Neo-Aramaic varieties. In Finnish (Uralic), the progressive gram *olla* + INF3 is noted to imply "imminence" with achievements (Heinämäki 1995:149). For Swedish, Blensenius (2015:229) shows that the *hålla på att* also has proximative uses.

I will most often refer to the *proximative* reading of progressive grams in this thesis, restricting the use of the term *avertive* to cases where it is clear and relevant to mention that the event was on the verge of happening but did not.

Progressives have also been noted to have futurate uses, as in *I am lecturing tomorrow*. I will refer to uses of progressive grams with future time reference as *futurate uses*. The term futurate excludes combinations of progressives and futures such as *will be* + GER in English, for example. Progressive grams differ in their ability to have futurate uses. In a comparative study of the English and the French *être en train de* + INF progressives, De Wit et al. (2013:849) show that futurate uses of the English progressive constitute 14% of the uses found in their data, while the French progressive does not have any such uses. At times it is difficult to distinguish between uses of progressive grams with proximative readings and uses that have a future time reference, in addition, the uses with future time reference are of different sorts. I will discuss these issues in more detail in Chapters 3 and 4.

For the English progressive, which is the most well-studied progressive, other peripheral uses have been investigated in various studies. One such use is e.g. *temporariness*. For instance, it is noted that while *Mr. Smith is standing by the Nile* indicates temporariness, the simple tense *The Sphinx stands by the Nile* refers to a more permanent situation (Comrie 1976:37). The English progressive may also connote "greater emotive effect", as in [*S*]*he is always buying far more vegetables than they can possibly eat* in comparison to the simple present [*S*]*he always buys far more vegetables than they can possibly eat* (Comrie 1976:37–38).

Following Comrie (1976), De Wit & Brisard (2014:70) list usage types for the English present progressive such as 'historical present progressive', 'futurate present progressive', 'temporal validity', 'duration', 'iteration', 'repetition', 'habitual' and 'modal', which they adhere to the "core meaning" of "epistemic contingency" for the present progressive in English. This "core meaning" is also

[4] Following Comrie (1976:64), Jahani (2017:275) defines prospective as "a prediction-based or intention-based state that is related to a future event, either imminent or non-imminent, which either takes place (non-avertive) or is averted (avertive)".

suggested for the French present progressive *être en train de* + INF (De Wit et al. 2013) and the German progressive *sein* + *am* + INF (Anthonissen et al. 2016) and possibly also other progressives. With "epistemic contingency" the authors refer to "non-necessity in the speaker's conception of current reality, as opposed to the simple present, which is analyzed as indicating [structural necessity]" (De Wit & Brisard 2014:50). More concretely, when they say that the "core meaning" of progressives is "epistemic contingency", I understand that as including a variety of meanings, such as ongoingness, temporariness, habitual, and various types of subjective readings. In my view, that only excludes generic uses.

The uses mentioned here are notably quite diverse: while some are grammatical forms (as exemplified with Tajiki), others are uses or interpretations of uses. In this thesis, I will restrict the term *peripheral uses* for all uses that are not 'ongoing at the reference time'. The term will also exclude uses that are clearly the remains of original meanings, as with the case of the locative uses of Mandarin Chinese *zài*. Among *peripheral uses*, I will distinguish *extended uses* when I explicitly talk of uses that are part of the further grammaticalization of progressives towards the imperfective. *Subjective uses* may refer to uses that are not 'ongoing at the reference point', in which case they can be seen as peripheral uses. But this notion is also used in the literature to refer to subjective nuances (e.g. irritation, surprise) that arise with the use of the progressive in addition to 'ongoing at the reference time'. These uses are then in principle not peripheral uses.

As is evident, it is not easy to determine what distribution a gram should have to qualify as a progressive and how many peripheral uses are acceptable before we have to view a pattern as something other than a progressive. One way of solving this problem is to view grams that have FOC ongoing as their main, i.e. majority, use as progressive grams. This is done in Chapter 3 for the Persian *dāštan* gram. In the typological investigation in Chapter 2, similarity to a cluster of grams will be used as criterion for being treated as a progressive. This will necessitate a discussion of borderline cases.

1.2.3 Contact-induced change

This section summarizes some relevant theoretical notions on contact-induced change. General issues related to borrowability in contact situations as well as more specific characteristics of periphrastic patterns and grammatical maturation are presented. These issues are intertwined but an attempt is made here to disentangle them in the following order: first, matter and pattern borrowing will be presented in section 1.2.3.1, then issues on contact-induced grammaticalization are presented in section 1.2.3.2 and finally the question of borrowability and stability is presented in section 1.2.3.3.

1.2.3.1 Matter and pattern borrowing

In studies on contact, Weinreich's (1953) terms *model* and *recipient language* are often adopted: model language refers to the language from which material or pattern is borrowed, and recipient language refers to the language that adopts the material or pattern from the model language. In their discussion of borrowings between languages, Matras & Sakel (2004) introduce the terms *matter* and *pattern borrowing*, where *matter borrowing* refers to the borrowing of morphological or phonological material and *pattern borrowing* refers to the borrowing of "the organization, distribution and mapping of grammatical or semantic meaning, while the form itself is not borrowed" (Sakel 2007:15).

Pattern borrowing is also referred to as *calquing* or *structural borrowing*. Johanson (2002:9) distinguishes between *global* versus *selective copying*, where *global copying* refers to situations when a language borrows an entire block of material and structural properties and *selective copying* when a language borrows certain features of the model structure. The function or structure of the borrowed item does not have to be identical to the original item in the source language. Otomi (Oto-Manguean), for example, borrowed *ko* from Spanish *con* 'with' but extended its function to also include 'made of', a function not covered by the Spanish *con* (Hekking & Bakker 2007:450).

Sometimes, one type of borrowing is more prominent than the other. For example, in Biak (Austronesian), the main type of borrowings from local Malay/Indonesian (Austronesian) are matter loans. Sakel (2007:19) assumes this to be due to the strict word order of Biak. Also, the Mandarin Chinese (Sino-Tibetan) influence on Vietnamese (Austro-Asiatic) mainly concerns matter borrowings, this is assumed to be related to written materials being the main source of influence. In other cases, pattern borrowing is more prevalent. It is, for example, assumed that if speakers attribute high status to two languages and wish to keep them apart, borrowing of patterns is mainly noted, and those matter loans that are found primarily consist of functional words (Sakel 2007:18–21). Sakel (2007:16, 24–25) notes that the hierarchical relations between languages as well as bilingualism have an impact on the occurrence of matter and pattern borrowings, so that the direction of matter loans goes from dominant language to dominated language and that pattern loans are generally not possible without bilingualism.

The situation becomes more complicated if varieties are related in which case it may not be possible to distinguish between matter and pattern borrowing. For example, the Swedish *kommer att* future is borrowed from the Danish *kommer til at* future, where it originally included the same constructional schema with both the preposition *till* 'to' and the infinitive marker *att*. Today, *till* has been dropped and the infinitival marker is optional (Dahl 2000b:320).

1:9 Danish, Indo-European (Davidsen-Nielsen 1990:122, mg)

Vi	*komm-er*	*til*	*at*	*synke.*
we	come-PRS	to	INFM	sink.INF

'We are going to sink.'

1:10 Swedish, Indo-European

Vi	*komm-er*	(*att*)	*sjunka*.
we	come-PRS	INFM	sink.INF

'We are going to sink.'

Finnish (Uralic) has created a future construction inspired by the Swedish one using the verb *tulla* 'become/come' with the illative case of the 3rd infinitive *-maan* (Dahl 2000b:320).

1:11 Finnish, Uralic (Karlsson 2015:222)

Tule-n	*palaa-ma-an.*
come-1SG.PRS	return-INF-ILL

'I will return.'

In the case of the Finnish pattern, it is clear that it is a calque on the Swedish pattern, i.e. a case of pattern borrowing. When it comes to borrowing from Danish to Swedish the two processes matter and pattern borrowing would in principle yield the same result before the more recent drop of *till* 'to'. In this case, however, it is natural to assume that we are dealing with pattern borrowing as well, since we can assume that Swedish did not borrow the words *komma*, *till* and *att*.

Heine & Kuteva (2005:220) describe replicated structures as structures that are perceived by the speaker as somewhat equivalent to structures in another language. Similarly, Matras & Sakel (2007a:835) view pattern loans as giving advantage to speakers who in this way match the organization of structures in separate languages. This, however, does not mean that the replica and source patterns have a one-to-one form-function relation, in fact, often the replica pattern will differ from the source pattern.

In discussing matter and pattern borrowing, the terms *Sprachbund* or *linguistic area* are often used, sometimes with slightly different meaning. What is most commonly meant by a linguistic area is that varieties within this area exhibit features that cannot be explained through shared ancestry but are assumed to have arisen due to language contact (e.g. Heine & Kuteva 2005:chap. 5; Wiemer & Wälchli 2012:14–18).

The notion of *areal cline* or *areal grammaticalization cline* refers to traces of diffusion forming clines that show degrees of grammaticalization synchronically

with regard to a specific feature in a restricted area: "[t]he gradual character of grammaticalization as a diachronic process is reflected in synchronic patterns of different kinds" (Dahl 2001:1468). Wälchli (2012:233) discusses *areal clines* as a special type of *areal pattern*.

> An *areal pattern* is a spatial constellation of linguistic features across languages which is significantly different from a random distribution and which cannot be fully explained by other factors than areality such as genealogic relatedness or universal principles. (Wälchli 2012:233–234)

Wälchli (2012:233) points out that in larger areas the involved languages may be expanding or disappearing, or speakers may be migrating. In addition, it can be difficult telling apart *exceptions* where diffusion did not take place from *destructions* where the result of diffusion is removed and *obstructions* where other processes interfere with diffusion. Also, information regarding the situation before the spread is needed in order to rule out genealogical effects as well as information regarding migration patterns. He concludes that, due to these complications, no ideal areal cline exists in reality (2012:235–236).

Studies on pattern borrowing often concentrate on cases where the languages in question are unrelated, but, as Dahl (2001:1457) points out, although borrowing or spread of a linguistic feature between two unrelated languages is in a sense more spectacular, studies of areal linguistics cannot restrict themselves to this 'tip of the ice-berg'-phenomenon. Often, features shared by related languages are due to later diffusion after the initial split of the languages. Also, contact-induced change is more likely to occur if languages are related than if they are not. In discussing unusual features in cross-linguistic studies, Koptjevskaja-Tamm (2010) notes that "[m]ost contact-induced change is […] not particularly spectacular, and most isoglosses are probably neither unique to an area or skewed in their distribution so much that they will 'betray' the area in a large-scale sample".

1.2.3.2 Contact-induced grammatical change

Contact-induced grammaticalization is a central notion in the literature on contact-induced change. Heine & Kuteva (2003) distinguish between *ordinary contact-induced grammaticalization* and *replica grammaticalization*. The former is the process whereby speakers of one language, after hearing a pattern in another language, grammaticalize a pattern with similar function using material in their own language. In the latter case, on the other hand, "the model language provides speakers of the replica language with guidelines as to how to replicate a grammatical category via grammaticalization" (Heine & Kuteva 2003:540).

Heine & Kuteva (2003:534) give an example of ordinary contact-induced grammaticalization from Tayo, a French-based creole language. In Tayo, the French numeral *deux* 'two' was grammaticalized as a pronominal dual *-de* under

the influence of Drubéa (Austronesian) and Cèmuhî (Austronesian), which have obligatory pronominal dual markers. Cross-linguistically, the numeral 'two' is a common source for creating dual markers (Heine & Kuteva 2003:534). Thus, when Tayo speakers grammaticalized the pronominal dual marker *-de*, they drew upon universal strategies of grammaticalization.

For Replica grammaticalization, Heine & Kuteva (2003:556) give an example from Basque as spoken in southwestern France, which has grammaticalized an indefinite article modeled on Gascon and French. Noting that the Romance languages have an indefinite article that is identical to the numeral 'one', Basque speakers developed their numeral one *bat* into an indefinite marker "using a grammaticalization process that they may have assumed had taken place in the model languages" (Heine & Kuteva 2003:556).

The notion of replica grammaticalization has been criticized for assuming that speakers have access to meta-linguistic information such as historical knowledge of grammaticalization processes (see, e.g., Gast & van der Auwera 2012:393). In this dissertation, it is assumed that speakers can only replicate processes that are transparent, i.e. where previous stages in the grammaticalization process are still in use or apparent – as in the Basque example. To give another example: a variety can only replicate the English future construction *gonna* making use of the verb *go* and *to* in their own language, as long as the *going to* future construction in English is still in use in a variety of English known by the speakers, or perceived as the original form of *gonna* (Östen Dahl, p.c.). Once a pattern has been incorporated, however, it may continue to grammaticalize, following the universal rules of grammaticalization.

Previous studies have shown that replica patterns may be less grammaticalized, less frequent, associated with a smaller range of contexts, and/or exhibit a lower degree of morphosyntactic obligatoriness than the model structure on which the replica is based (Heine & Kuteva 2006:17, 22–25). Also, Aikhenvald (2013:27) notes that contact-induced change often results in a copied pattern having more limited use in the recipient language than in the model language. She gives an example from Manambu (Ndu), spoken in Papua New Guinea, which has been in close contact with and under the influence of Kwoma (Sepik). In Kwoma, a possessive verb *ta* 'have', also meaning 'be, exist, stay, live', is used with possession of any kind. In Manambu, a possessive verb *tə-* 'have', also 'stand, exist', is copied from Kwoma but is used for temporary possession only. No other Ndu variety has a possessive verb. The Manambu possessive verb construction thus has a more limited use than the Kwoma construction. Examples 1:12 and 1:13 show the possessive verb construction in Kwoma and Manambu, respectively.

1:12 Kwoma, Sepik (Aikhenvald 2013:27)

Mii *him* *ta-wa*?
you.M tobacco have-PRS
'Do you have tobacco?'

1:13 Manambu, Ndu (Aikhenvald 2013:27)

dəy *samasa:ma* *wapi* *tə-na-di*.
They many bird have-ACTION.FOCUS[5]-3PL
'They have/possess many birds.'

There are also examples of contact situations where the replica pattern has the same functional span or is further along the maturation path than the model pattern, but this is not very common.

Wiemer & Wälchli (2012:41–44) point out that in some cases languages take turns in being the model versus the recipient language, a scenario that they refer to as *cross-transfer*. There are also instances of contact-induced grammatical change that do not involve a recipient and a model language. In these cases, structures develop simultaneously in several languages, which is referred to as *parallel grammaticalization*. An instance of parallel grammaticalization is the grammaticalization of the second singular imperative of the verb 'let' in Latvian (*ladi* > *lai*) and Livonian (*lask* > *laz*) into a marker of purposive clauses, as illustrated in 1:14 (Wälchli 2000).

1:14 Parallel grammaticalization of 'let!' in Latvian and Livonian

'let!' > adhortative > indirect command > purposive

1.2.3.3 Borrowability and stability

Not every grammatical category has the same likelihood of being borrowed, i.e. the same degree of *borrowability*. Aikhenvald (2006) summarizes features that facilitate diffusion of forms and patterns (referring to Heath 1978; Moravcsik 1978; Matras 1998; Matras 2000; Dalton-Puffer 1996). The features relevant for this study are given in 1:15.

1:15 Features which facilitate diffusion (Aikhenvald 2006:26–32)

- Constructions used for pragmatic functions such as focus, topic, backgrounding and foregrounding are most easily diffused.

[5] The author does not further explain the glossing.

- The more frequent a category is in one language the more likely it is to be diffused.

- Markers that have greater impact on cultural conventions are more likely to diffuse than those who have less.

- Gaps in the recipient language increase likelihood of diffusion.

- Typologically common phenomena are more likely to be diffused.

- If two varieties are structurally similar this increases the likelihood of diffusion.

It is also commonly assumed that the probability of borrowing increases if the varieties in question are typologically close (Koptjevskaja-Tamm 2010). Thus, "if languages in contact share a category or a construction, language contact may increase its frequency or its productivity" (Aikhenvald 2006:22). The outcome of language contact depends on linguistic factors, like the ones just given, as well as on non-linguistic factors, such as the aforementioned hierarchical relationship between languages in contact, the level of bilingualism within the speech community, the speaker's attitude toward the languages, the time span and intensity of contact and the characteristics of contact (Matras & Sakel 2007b:2).

Many typological studies on contact-induced change, however, have a linguistic focus rather than a non-linguistic one as they aim to examine what linguistic factors are affected once contact is established. Or as Matras (2007:34) puts it, "what is it that makes one category (or category value) a more attractive candidate for 'system conflating' than another?". According to Matras, the borrowing susceptibility of one category over the other cannot be explained by features such as social acceptability, prestige or gaps in the recipient language, rather, he sees borrowing of categories as a strategic compromise adopted in conversation and explains the hierarchical structure by the "need to reduce the cognitive load when handling a complex linguistic repertoire" (Matras 2007:67).

Moravcsik (1975:110) suggests that lexical items are borrowed before non-lexical items, that free morphemes are borrowed before bound morphemes, and derivational items before inflectional ones. Borrowability also interacts with the types matter versus pattern so that borrowings of certain grammatical categories are restricted to material borrowings, while others mainly involve pattern borrowing. Matras (2007:46), which provides an overview of 27 sample languages, presents a hierarchical relation of TAM categories and their borrowability, as seen in 1:16. The hierarchy reflects both frequency and implicational relationships.

1:16 Borrowability hierarchy for TAM markers (Matras 2007:46).

Modality > aspect/aktionsart > future tense > (other tenses)

Matras' study shows that borrowing of tense, aspect and mood markers is relatively rare, although not unattested. Modality markers are most often borrowed, followed by aspect/aktionsart markers. Tense markers are less often borrowed than modality and aspect/aktionsart, in fact, the borrowed tense markers that Matras found most often involved future tense. In Matras' study, borrowings of modality markers mainly involved matter borrowings, while borrowing of future tense markers involved pattern borrowings. Borrowings of aspect or aktionsart markers involved both material and pattern borrowings. It should be noted that Matras does not comment on the way these markers are expressed, i.e. if they are free markers or their degree of inflectionality. We may then suspect that the hierarchy could be partly explained by the boundedness of markers in view of the higher borrowability of free markers (Östen Dahl, p.c.).

Interestingly, when aspect is mentioned, a couple of examples of progressives are given: in Nahuatl (Uzo-Aztecan) a progressive is found that is based on the Spanish model.[6] Similarly, in Sekpele (Niger-Congo) a periphrastic present progressive is found that is similar to Ewe (Niger-Congo)[7] (Matras 2007:44–45).[8] In addition, the increase in frequency and productivity in the progressive marker in Pennsylvania German under the influence of English has been noted (Aikhenvald 2006:22).

In the literature, it has been noted that "grammaticalization processes are highly sensitive to contact influence" and that periphrastic constructions are more prone to being borrowed than inflectional constructions. (Dahl 2004:127)

> [T]he empirical evidence at hand […] suggests that borrowing is constrained in a way that reflects the maturity of the borrowed patterns. Thus, Field (2002:38) suggests the following 'Hierarchy of Borrowability' […]:
>
> CONTENT ITEM > FUNCTION WORD > AGGLUTINATING AFFIX > FUSIONAL AFFIX
>
> Thus, what is borrowed, or calqued (i.e. translated), in grammar will most frequently be periphrastic constructions or free markers, and less often affixes, although the latter is also observed to happen (Dahl 2004:127).

There is a greater volatility in the earlier stages of grammaticalization than later ones; "[f]or many grammatical phenomena, it may well be that they either

[6] There is no information on which Nahuatl language is being referred to.

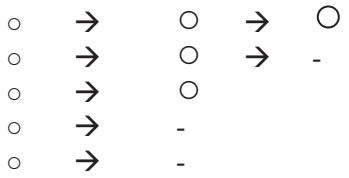
[7] The situation of Sekpele and Ewe turns out to be of interest for the discussion of the origin of the *dāštan* construction in Chapter 6.

[8] Matras (2007:45, 69) also mentions a progressive construction in Neo-Aramaic, which probably refers to the pattern discussed in section 5.3.4.2.

disappear (a majority) or make it to a more mature stage (a minority)" (Dahl 2004:268). Features may also be easily obtained but rarely lost (for example vowel nasalization) or rarely obtained and rarely lost (vowel harmony) (Greenberg 1978:75–76; Dahl 2004:275).

The perfect gram type is interesting for comparison since it is also often expressed periphrastically: "the perfect is a gram type that is frequent, that is to say, likely to appear in different languages, but unstable, as it often tends to be lost" (Lindstedt 2000:366). Also future grams are often marked periphrastically, as seen in Bybee et al. (1994:252–253). In their study, they note that it is not uncommon for languages to have more than one future pattern, in fact, 49 out of the 70 languages in their sample had more than one future gram, some even as many as six. They regard such duplications as "a consequence of the independent development of grams from distinct sources and from similar sources at different periods" (Bybee et al. 1994:243). Progressive, perfect and future constructions, as well as other patterns, may grammaticalize into something with a more general function. If these patterns are marked periphrastically, in this process, the pattern tends to become inflectional. Not all patterns will follow or survive every step of a grammaticalization process, some may stay where they are or perhaps disappear. As a result, inflectional constructions, typically being further along grammaticalization paths than periphrastic ones, have a lower probability of existing than periphrastic patterns at the beginning of a maturation process (Dahl 2004:137–140). Example 1:17 illustrates this by depicting patterns as circles whereas dashes illustrate the disappearance of a pattern. The lack of an arrow illustrates that the pattern is stable with no further grammaticalization at the point illustrated. The increase in size of the circles represents their increase in frequency.

1:17 Surviving and disappearance in the process of maturation



In addition, it may well be that later stages of grammaticalization have a slightly lower probability of taking place (Östen Dahl, p.c). Looking at future gram families in Europe, Dahl (2004:267–268) notes that the majority are periphrastically marked and that these are all more recent patterns. He concludes that "[t]his is consistent with the claim that the early steps in a grammaticalization process are more easily taken than the later ones" (Dahl 2004:268). Thus, we expect to find more competing constructional forms for the same function in the beginning of a maturation process compared to later stages.

1.3 Glossings and language classifications

Throughout the dissertation, the abbreviation *mg* will refer to examples from unglossed sources that have been glossed by me. In cases where existing glossings are changed, the original glossing is given in a footnote. When examples are from my own data collection, like most of the examples in Chapter 3, examples are glossed by me, in which case *mg* is not specified. Throughout the dissertation, obvious overlaps in glossing abbreviations have been adjusted, so that, for example, PT for 'past' is changed to PST, IMPF for 'imperfective' is changed to IPFV, and so on. In Chapter 2, language examples may be given without any glossing in cases where no glossing is provided by the author of the language description. When sources contain translations into languages other than English, these are provided in footnotes.

Language names are generally followed by the name of the top-level family the language belongs to, following the classification in WALS, except for Turkic languages, which are classified as such rather than Altaic. In Chapter 5, further sub-classification of the Iranian languages is provided. When citing authors, the classification provided by the author is kept. Well-known languages will be given without classification.

2 Progressive grams in parallel corpora

This chapter investigates the uses of the progressive gram type by looking at the distribution of a number of progressive grams in two parallel corpora. The patterns discussed here are grams that have a distribution similar to patterns that are generally viewed as progressives, such as the English *be* + GER progressive pattern. Using parallel corpora presents a set of problems, but it also enables large-scale cross-linguistic studies of the distributions of grammatical patterns. As noted in Dahl (forthcoming a.), an important assumption in corpus studies is that the distribution of items reflects the semantic and pragmatic properties of the items. Thus, in cross-linguistic corpus studies, items with similar meanings/functions, such as, e.g., progressives, are expected to have similar distributions, that is, they are *distributional equivalents*. Using parallel corpora, we may evaluate whether two grams that are referred to by the same name in the literature have similar distributions, and how they differ.

In this section, the Bible corpus, which consists of translations of the New Testament from the Bible, and the TED corpus, which includes transcriptions of subtitles of TED (Technology, Entertainment, Design) and TEDx (local TED-like) Talks, are used in order to investigate the distribution of progressive grams. The Bible corpus has previously been used for studies such as Dahl (2014), which looks at the perfect gram type, or Dahl & Wälchli (2016), which examines perfect and iamitives.[9] Dahl & Wälchli (2016:330) discuss the issue of assuming that texts are 'parallel' in the following way:

> It is easy to see that the use of parallel corpora presupposes that the texts are "parallel" enough semantically and structurally – in other words, the quality and faithfulness of the translations, as well as the approach of the translators, are crucial […]. In addition, we usually do not know how well a translation represents the language of the community for which it was made. Still, translations are samples of how humans use language, and at least as long as we see them as representing themselves (i.e. particular doculects of a language rather than the language they are written in in its entirety) they are valid objects

[9] The use of parallel corpora for cross-linguistic studies is not limited to Bible translations, often however, such studies use corpora consisting of texts in two languages. Cysouw & Wälchli (2007:95) introduce the term 'massively parallel texts' for texts that have translations in many languages. For an overview of some massively parallel texts the reader is referred to this article.

> of study. Any generalization to other language varieties, however, has to be treated with great caution. (Dahl & Wälchli 2016:330)

The Bible corpus and the TED corpus are quite different from one another: while the Bible corpus includes 1107 languages, is annotated, (fully) parallel and also includes languages other than so-called 'standard languages', the TED corpus includes 99 languages with an uneven representation of texts, meaning that it is only partially parallel and heavily skewed towards standard languages with official status and many speakers. Unlike the Bible corpus, it is not annotated. However, the texts in the TED corpus can be expected to be much closer to modern everyday language than the Bible translations. Also, the TED corpus is much larger, as it may contain up to 3.5 million words for a single language, while a translation of the New Testament typically contains around 200 000 words. Since the TED corpus includes a much smaller and less representative set of languages, it is mainly used as a complement to the Bible corpus, more specifically to further examine the occurrences of the progressive in present and past.

In the Bible corpus, 89 grams are examined, while the TED corpus is mainly concerned with 14 grams, some of which are not included in the Bible corpus. Grams belonging to English, Indonesian, Portuguese, Spanish, Vietnamese and Turkish are included in both corpora.

In this chapter, the methods and results using the Bible corpus and the TED corpus are presented in turn. The investigations have been conducted in collaboration with Östen Dahl, who has been responsible for handling and searching the data base. Two different methods were employed in the two corpora. In the Bible corpus, progressive grams were automatically identified, adopting the methods used in Dahl & Wälchli (2016). In the TED corpus, on the other hand, grams that are traditionally assumed to be progressive are investigated. The reason for the different approaches has to do with the nature of the two corpora, both the lower number of grams and the texts being partially parallel make the approach used for the Bible corpus to be unsuitable for the TED corpus. Besides, for most of the languages in the TED corpus, the finding of progressive grams is easily obtained through descriptions.

2.1 The Parallel Bible Corpus

In this section, the investigation using the Bible corpus is presented. In section 2.1.1, the method used in this section is presented, including information about the corpus, methods of identifying and grouping progressive grams and methodological issues. In section 2.1.2, data and analysis are presented and section 2.1.3 summarizes the findings.

2.1.1 The corpus, gram set, method and methodology

2.1.1.1 The corpus

The Bible corpus consists of 1267 New Testament translations from 1107 languages that have been automatically annotated and aligned at the word level using methods developed by Östling (2015). Here, *language* refers to a variety with an ISO 639-3 code. The corpus thus includes approximately 15% of the world's languages. The text in the Bible has the advantage of being divided into numbered verses. Consequently, translations of the same verses can easily be identified using these numberings. The automatic alignment includes annotation of part of speech categories in all translations as well as linkage at the word level from all texts to the translations in English, French, German and Swedish; the latter versions are also fully parsed syntactically. The automatic annotation is not perfect and the accuracy of the alignment depends on how close the source and target languages are. The alignment enables further division of the verses into 'segments' using automatic methods. The automatic segmentation of the corpus has been done by Östen Dahl. As these methods are automatic, columns may contain errors.

In English, a 'segment' always contains a main verb or an infinitive, the infinitive is included since other languages may use finite forms for similar functions. If a language has no copula verb, segments that include copula verbs in other languages will be verb-less in that language. Roughly speaking, then, segments consist of a verb and all its syntactic dependents. In the other translations, segments are defined in two steps: words that are linked to words in English are given the same segment number, and the numbers are extended to unlinked words next to the linked ones.

An example of the segment *For behold, he is praying* is given for English in Table 2 and for Amatlán Zapotec (Oto-Manguean) in Table 3, with the 6th column showing the alignment between the two languages as produced by Östling's (2015) algorithms. The first column in both tables gives the verse code; the second column gives a numbering of the words in the verse; the third column gives the word in the corpus; the fourth column gives an automatically assigned identifier of the lemma; the fifth column gives the part of speech annotation; the sixth column gives the alignment with English, French, German and Swedish, for example in this segment *nel* in Amatlán Zapotec is linked to *for* in English and so on; and finally the seventh column gives the number of the segment.

Table 2. Annotation of the segment 44009011$39 in English

VERSE ID	WORD ID	FORM	LEMMA	POS	ALIGNMENT	SEGMENT
44009011	34	For	for	ADP	34	39
44009011	35	behold	behold	NOUN	35	39
44009011	36	,	,	.	36	0
44009011	37	he	he	PRON	37	39
44009011	38	is	be	VERB	38	39
44009011	39	praying	pray	VERB	39	39
44009011	40	,	,	.	40	0

Table 3. Annotation of the segment 44009011$39 in Amatlán Zapotec

VERSE ID	WORD ID	FORM	LEMMA	POS	ALIGNMENT	SEGMENT
44009011	30	nel	nel	CONJ	34	39
44009011	31	lee	lee	DET	35	39
44009011	32	xaa	xaa	PRON	37	39
44009011	33	kawdizhno	wdizhno	VERB	39	39
44009011	34	na	na	PRON	_	39
44009011	35	.	.	.	40	39

The segment *For behold, he is praying* has the code 44009011$39 which can be read as [44][009][011][39], where 44 refers to the books in the Bible, 009 refers to the chapter and 011 to the verse. The number following the $ sign identifies the segment by its head word number in English. The numbering of the books in the New Testament starts with 40, which refers to the book of Matthew, followed by 41 which refers to the book of Mark, 42 to the book of Luke and so on.

Since the inclusion of a language in the Bible corpus depends on the availability of a translation of the New Testament for that language, the coverage of the languages of the world is uneven. For example, the Bible corpus contains relatively few languages from North America since there are fewer available Bible translations for these languages. The corpus then aims at completeness rather than genealogical and areal representativity.

2.1.1.2 Identification of progressive grams

Using the methods of Dahl & Wälchli (2016), Dahl (forthcoming a.) and Dahl (forthcoming b.), a cluster of patterns is identified in the Bible corpus, the members of which will be treated as representatives of the progressive gram type. This was done by searching for grams with a similar distribution to that of known

progressive grams such as the English progressive. A simplified explanation of the procedure is given below in which the method is divided into two main steps.[10]

The first step involves the collection of candidates for progressive grams, and the second step involves the trimming down of the number of grams in order to exclude grams that differ greatly in their distribution from the others. The first step can be rather inclusive, because the second step is restrictive, which means that one need not worry if a clearly non-progressive gram is erroneously included in the first step. These steps are explained in turn. Similar to Dahl (forthcoming b.), I will refer to the final 89 grams discussed here as a *gram set*.

Collection. The first step is an exploratory one where grams are collected in different ways: grams may be included if they have been described as progressives in the literature, or they have been collected using a group of *seed grams*. The use of seed grams is an automatic method in which a number of potential progressive grams are used to search for other grams with similar distribution. Basically the distribution of seed grams in a number of languages is compared against any string of letters that occurs with a similar distribution in other languages. Chi-square tests are used in order to rule out strings that occur in the same segments by chance. To start with, the English (present and past) progressive gram was chosen to be a seed gram in the collection of other grams with similar distribution. The grams that were collected in this search were then used as seed grams in a second search. Thanks to the second search, then, the bias toward English is reduced. The collection of grams resulted in 130 potential progressive grams.

Trimming down. The second step involves the trimming down of the group of 130 potential grams to a gram set consisting of 89 grams that are viewed as belonging to the progressive gram type.[11] The results of this investigation will, however, also include a small group of grams that have both ongoing and other uses, these are discussed in section 2.1.2.6.

Trimming down is done using two statistical criteria for relevance, namely *recall* and *precision*. Both recall and precision make use of top lists from the corpus. A top list includes the top segments in which our grams occur, ranked from the segments where most grams occur to the segment were least grams occur. For recall, the occurrence of a gram in the top 20 segments is calculated. That is, the top 20 segments for all grams are listed, and for each gram the percentage of occurrence in that list is checked. The top list used for calculating precision was generated in the following way: a list of all the segments in which the grams of the gram set occurred was created. From this list, the occurrences in the segments

[10] In reality, the process has not been a straight forward one, rather different attempts were made in order to locate the grams, i.e. the collection and trimming down were done several times, meaning that the steps were repeated.

[11] A rather large group of grams consisting of locative markers were excluded. In these languages, the locative markers are used as such as well as being part of a progressive pattern (when occurring with a verb or similar). Attempts at improving the search strings failed, likely due to errors in the automatic part of speech annotation for these translations.

in which the fewest grams occurred were removed one by one until 15% of the occurrences had been distanced. The remaining occurrences constitute 85% of all the occurrences of the members of the gram set. This is illustrated in Figure 2. For each gram, the percentage of occurrences in the precision list is calculated.

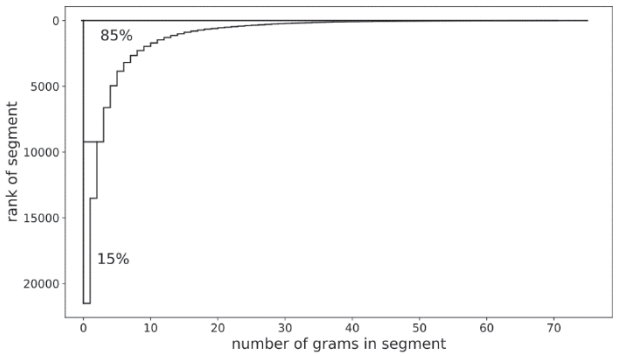


Figure 2. List of all the segments in which the grams of the gram set occur

Recall and precision were calculated for the whole gram set, as well as for three subsets of that set obtained by the clustering algorithm PAM (Partitioning Around Medoids).[12] Clustering is a statistical tool applied to the data, and the results need to be further analyzed, meaning that it is up to the researcher to figure out what type of uses it is that puts grams in the same cluster. In section 2.1.2, the three cluster groups will be shown to include grams with specific uses differentiating them from the other groups.

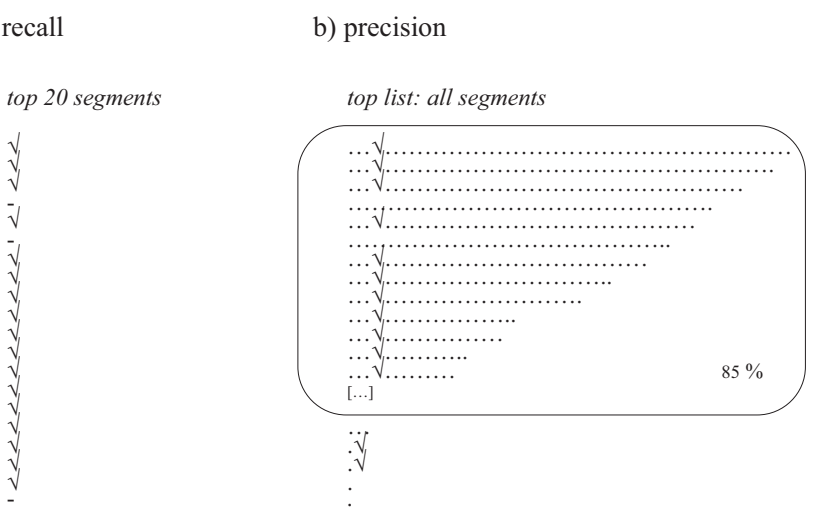
Grams that at the cluster level either had a recall value higher than 65% and a precision value higher than 50% *or* a precision value higher than 80% were included.[13] The values 65% and 80% were chosen so that the grams with the lowest percentages would be excluded.

Example 2:1 aims at illustrating recall and precision for a hypothetical gram included in the gram set, where the check mark illustrates the occurrences of the gram in the top lists. In a), then, the gram occurs in 17 of the segments in the top 20 list. The list in b) aims to illustrate the occurrences of our hypothetical gram in the top list presented in Figure 2, where each dot represents an occurrence of a gram in the gram set and the check-mark represents the occurrences of our hypothetical gram. […] cuts off the middle of the top list. The figure intends to show that most of the occurrences of our gram lie within the top 85% occurrences.

[12] The algorithm is `pam` in the R `cluster` package (https://cran.r-project.org/web/packages/cluster/cluster.pdf).

[13] I have made two exceptions for Sekpele and Makaa as they instead fulfil the requirements of recall and precision at the gram set level.

2:1 Illustration of *recall* and *precision* for a hypothetical gram



There are two reasons for using both precision and recall on our data. The first is that it is a problem to compare grams with different number of occurrences in the corpus. The extracted pattern *kampound-* in Central Khmer (Austro-Asiatic), for example, occurs only 174 times in the Bible corpus while *yac-* in Tzeltal (Mayan) occurs 2796 times. Tzeltal then has a clear advantage when calculating its occurrence in the top 20 list, i.e. its recall value. It is as if Tzeltal would have more shots to take than Central Khmer in a game of shooting down clay pigeons. The precision test, instead, looks at all the occurrences of the gram and whether they appear in the segments that other potential progressive grams do. The precision test benefits grams with lower occurrences. Progressive grams with many occurrences in the Bible corpus can then be expected to have a high recall value, while progressive grams with fewer occurrences can be expected to have a high precision value.

The second reason for using the precision test is that it aims to exclude those grams that are used with the ongoing function, but which also occur to a great extent in other functions. An imperfective gram, for example, may occur in most of the segments of the top 20 list, since it also covers the ongoing function, but will have low precision, since it also occurs in other segments in which progressive grams typically do not occur. The precision value then essentially answers this question: Out of all the occurrences of a gram in our corpus, how many fall within and outside a distribution we assume to be typical for progressive grams? In, for example, Wolof (Niger-Congo), the gram *ngi*[14] has a recall value

[14] Attempts were made to also include the forms *nga*, *ngay* and *ngiy* but in all cases the precision value decreased, which suggests that these forms also have other uses. In addition, *ngi* + VERB did not increase the values, which could be due to errors in the automatic annotation.

of 5% for the gram set, which is very low, but has a precision value of 90%. This means that it has only one occurrence in the general top 20 list, but almost all of its occurrences fall within what we assume is a typically progressive distribution. When looking at the cluster level, however, we see that its recall value rises to 55% for one of the groups. It turns out that this gram is restricted to present time reference, an issue that is further discussed in section 2.1.2.4.

The precision value is also interesting when looking at the peripheral uses of grams, since progressive grams that have additional uses are expected to have a lower precision value than grams that mainly occur with ongoing uses. Notably, it is difficult to draw an exact line between progressive grams with extended uses and grams that are better described as something else.

Once the exclusion of grams was completed the list was further cleansed so that each gram is represented only once for those languages with several translations of the Bible, which is the case for English, Spanish, Portuguese, Indonesian and a few other languages. Here, the version with the best recall and precision values was simply chosen.[15] Also, in cases where grammatical descriptions have been available and easily accessible, a swift check was done, and inaccuracies detected in the forms yielded by the automated searches were corrected. The forms of the 89 grams are presented in Appendix A.

To summarize, in the first step the aim is to throw a wide net, that is, to obtain many constructional candidates that may be progressive. In the second step, grams that differ in their distribution to a great extent in comparison to the other grams are excluded. This means that we can allow ourselves to be rather inclusive in the first step. Hypothetically, we could even include a gram that we know is not progressive, such as, e.g., a perfect gram, since this gram will be excluded in the second step. This enables a view of the progressive based on distribution in the Bible corpus, so that a gram with a high recall and precision value can be discussed as a progressive, and a gram with lower recall and precision values can be discussed as, e.g., a progressive with peripheral uses, as is done in section 2.1.2.6. There is no claim in this investigation that all progressive grams in the Bible corpus have been collected. In fact, as will be evident in section 2.1.1.4, there are a number of problematic issues using this method, some of which make it difficult to capture grams marked in certain ways.

The Venda and Ewe grams illustrate the outcome of the method. In Venda, a Bantu language spoken at the border of Mozambique and South Africa, a freestanding *khou* marker was extracted. As seen in example 2:2, the marker may occur in negated clauses.

[15] One language is represented by two grams. In Vietnamese (Austro-Asiatic), *đang* and *đương* were extracted from two different Bible translations. Phan (2013:4) refers to both *đương* and *đang* as "durative". Although both translations contain both markers, there is a clear preference in that the translation from 1934 has 86 occurrences of *đang* and 204 occurrences of *đương,* while the 'easy-to-read' version from 2011 contains 485 occurrences of *đang* and 4 occurrences of *đương*. I will discuss these grams again in section 2.2.

2:2 Venda, Niger-Congo (Poulos 1990:258)

Musadzi ha khou ri bikela.
The woman is not cooking for us.

Since the Bible corpus includes automated part of speech annotations, these can be used to make searches more precise, although some caution is necessary in view of the variation in quality of the part of speech tags. In Ewe, a Kwa (Niger-Congo) language spoken in Ghana, the extracted marker is a suffix *-m*[16], which attaches to an item annotated as a verb, in other words a verbal suffix. This marker occurs with an auxiliary verb *le* in the present and an auxiliary verb *nɔ* in the past to form the progressive (Dzablu-Kumah 2006:58). Dzablu-Kumah (2006:58) notes that the *-ḿ* suffix on verbs originates from *me* 'in', thus *éle yiyiḿ* '(s)he is going' is construed from a pattern that literally meant '(s)he is in going'.[17]

2.1.1.3 Comparative segment sets and 'direct speech' passages

In order to check for temporal preferences or peripheral uses of the captured grams, their distributions were compared to sets of relevant segments. To begin, the distributions were compared to the segments in which the Esperanto translation uses present or past tenses in order to see whether any grams are restricted or have preferences for present or past. The simplicity of tense marking in Esperanto makes it possible to use the present and past tenses as proxies for present and past time reference, respectively. In the Esperanto present tense, all verbs take the *-as* suffix and in the past tense all verbs take the *-is* suffix. There is a progressive in the languages that builds on the present and past tenses, i.e. *-antas* and *-antis*, respectively, but this marker is barely used in the Bible translation. Regardless, the present and past endings will also include any uses of the progressive.

In order to check to what extent the grams have habitual or future readings, their distributions were compared to those of habitual and future grams, more specifically, to the top 20 lists of sets of such grams.[18] Performative utterances are another peripheral use of interest. A performative sentence can be described as an utterance by which the speaker performs a speech act by saying that (s)he does, e.g. *I* (*hereby*) *promise to pay you ten euros*. Due to their meaning, these utterances are restricted to present time reference. In order to check whether members of the gram set occur in such contexts, a set of segments of this character was identified. Since it is sometimes difficult to know if the segments were

[16] In the PBC, *-m* is written without a tone mark. Also, the search was for V-*m* without the auxiliaries determining tense, since V-*m* captures both present and past forms.
[17] Ameka (2006:131) refers to *-ḿ* as a pattern found in certain dialects.
[18] The habitual and future grams were identified in ongoing research by Östen Dahl.

actually intended to have a performative interpretation, they will be referred to here as "performative-like". Finally, the distributions of the grams were run against the distributions of the stative verbs *know*, *love*, *stand*, *remain*, *have*, *hold*, *concern*, *understand*, *lie*, *sit* and *stay* in the English translation.[19]

In section 2.1.2.6, the occurrences of the progressive grams in these segment sets are presented. Progressive grams may, however, also have less frequent and more subtle peripheral uses that cannot be captured using these automatic methods.

Although the Bible corpus is a written source, it contains many direct speech quotations. These passages allow for a comparison of written language and representations of natural speech. Passages are treated as direct speech in the Bible corpus if they are enclosed in quotes but not preceded by the verb 'write', and are not longer than two verses. The latter two criteria intend to exclude quotes from the Old Testament (Hebrew Bible), sermons and similar passages.

2.1.1.4 Methodological issues

Like all methods, the method employed in this section is not without its problems. I will address some issues starting with those regarding the corpus. There are several problems with using a corpus based on Bible translations. The language of many Bible translations may be perceived of as archaic and written sources with written conventions may have restrictions, such that a progressive gram is not allowed or so that a progressive gram has a different distribution. Note however, that both these points mainly apply to standard languages with many speakers and longer written traditions. There may also be translational biases towards the source language. For more reading regarding translational issues regarding the Bible corpus, the reader is referred to de Vries (2009).

The automatic extraction of grams is also limited in what types of structures it can capture. Irregular or suppletive forms cannot be captured automatically since whole paradigms cannot be captured. If we know of these forms, however, (for example, through grammatical description, or if we discover that only part of a paradigm has been captured in the search), we can adjust the search string to include all forms. Periphrastic forms are also not automatically captured as such. If a periphrastic pattern is frequent enough, it is possible that one section of the pattern is captured. For example, for Spanish and Portuguese, in one of the earlier searches, the captured patterns were *est-* and *-ando*, respectively. These were then corrected so that the form that was searched for includes both the auxiliary, in all its conjugations, and the gerund form of the verb. Similar corrections have been done for several other grams. The automatic method then clearly advantages grams realized as a unique string of letters.

[19] These first eight are the most common stative verbs in the corpus.

The method employed here can in principle provide several grams per language, but the search algorithms have been calibrated to avoid too many false positives. Thus, in all cases, only one gram per translation was identified.

In a number of cases, grams were extracted from languages that do not have grammatical descriptions. This is an advantage of the method since it enables the inclusion of languages that are usually unavailable for typological studies, but it is also problematic since there is no external source that can verify the identification of the gram. This is also true for grams that belong to languages with grammatical descriptions but where the captured gram is not mentioned in those descriptions. We can then expect to encounter errors in the extracted forms of the progressive grams, especially for those that belong to languages for which grammatical descriptions are unavailable. But, as will become evident, this will not crucially affect the conclusions drawn since we are mainly concerned with the segments that are shared among these grams, meaning that minor errors in search strings will not affect the outcome in a drastic way.

The identified search strings are viewed as adequate for capturing the distribution of the progressive pattern in the Bible corpus. At times, the search string is simply a morpheme or a freestanding marker that is a simplification of the form of the progressive gram, as a progressive gram also (at least) includes a verb. Attempts at specifying the search string to include, e.g., a verb or a specific verb form were not always more successful than the original search string, most probably due to errors in the automatic annotation of that language.

Due to the nature of the Bible corpus, we can expect to be dealing with highly grammaticalized progressive patterns that are not restricted to the spoken language. Also, using this method will only give us progressive grams with a high frequency and not capture those that for various reasons do not occur often in the New Testament. For example, the Persian *dāštan* progressive is not included in the gram set since it is typically not used in written texts. Also, the more formal Persian *dar hāl=e* 'in state of' pattern, which occurs in the New Testament, is also not included since it only occurs a few times. Similarly, the Tajiki progressive formed with past participle + *istodan* 'to stand, be standing' is also not included since it has almost no occurrences in the Bible corpus. The Iranian languages discussed in Chapter 5 do not have Bible translations included in the Bible corpus. In fact, there are no Iranian languages among the 89 members of the gram set.

2.1.2 Data and analysis

We will now look at the segments in which the 89 progressive grams of the gram set most often occurred. All English examples are from the Lexham English Bible (LEB).[20] The overwhelming majority of the top segments are in the English past

[20] Copyright 2012 Logos Bible Software. Lexham is a registered trademark of Logos Bible Software.

progressive. In the top 50 segments in which the progressive grams occurred, only 11 segments were in the English present tense and only 1 segment in the English future tense. In example 2:3, the top five segments are shown in italics in the verses in which they occur. The segment codes are followed by the percentage of the gram's occurrences in this verse. Thus, 84% of the grams occurred in the segment of the top verse, 76% in the second top segment, and so on. Notably, these five examples all involve events of praying or speaking.

2:3 Top five segments

a) 44009011$39 0.84
And the Lord said to him, "Get up, go to the street called 'Straight' and in the house of Judas look for a man named Saul from Tarsus. *For behold*, *he is praying*,

b) 42003021$21 0.76
Now it happened that when all the people were baptized, Jesus also was baptized, and *while he was praying*, heaven was opened,

c) 40017005$5 0.73
While he was still speaking, behold, a bright cloud overshadowed them, and behold, a voice from the cloud said, "This is my beloved Son, with whom I am well pleased. Listen to him!"

d) 44010030$18 0.72
And Cornelius said, "Four days ago at this hour, *the ninth*, *I was praying in my house*. And behold, a man in shining clothing stood before me

e) 41014043$8 0.71
And immediately, *while he was still speaking*, Judas one of the twelve arrived, and with him a crowd with swords and clubs, from the chief priests and the scribes and the elders.

Example 2:3a) involves an event that is ongoing at one FOC point in time, namely the speech moment. The verse is in the present tense in the English translation and is part of a quote intended to represent spoken language. The rest of the top five segments are different. They are backgrounding clauses in the past tense in the English translation, in the sense presented in Chapter 1. The present and past segments will be further discussed in the upcoming sections 2.1.2.1 and 2.1.2.2, respectively.

The top 50 list seems to suggest that uses of progressives with past time reference are much more common in the Bible corpus than uses with present time reference. However, looking at all occurrences of the members of the gram set,

we see that this is not the case. For this purpose, the occurrences of our 89 grams were checked against Esperanto present and past tense, i.e. the sum of all occurrences that matched Esperanto present tense were calculated, on the one hand, and the occurrences that matched Esperanto past tense on the other. For each gram, there were also a number of occurrences that did not match present or past. The results show that the ratio between present tense and past tense is 58% to 42% in favor of the present tense. This is higher than the general ratio of present vs. past tense in the Esperanto Bible translation where the ratio of Esperanto present versus past tense is 43% to 57% in favor of the past tense (or 49% to 51% if present and future are put together against past, which may be a better calculation for those languages that do not have future tense). This is then an indication that the 89 grams occur more often in the present than in the past. Thus, even in a corpus consisting of written texts, including many narratives, the progressive is more often used in the present. In fact, it turns out that it is only in the top 100 list that the past segments outnumber the present segments, after the top 100, there are more present segments than past ones. In direct speech segments (presented in section 2.1.1.3), the present always outnumbers the past. This is illustrated in Table 4, which shows the Esperanto present and past tense in the top segments, for the whole corpus as well as direct speech segments.

Table 4. Esperanto present and past tense in top segments of the gram set

TOP LIST SEGMENTS	WHOLE BIBLE CORPUS		DIRECT SPEECH	
	PRS	PST	PRS	PST
1-100	34	59	16	0
101-200	50	37	20	0
201-300	50	35	9	1
301-400	64	21	15	0
401-500	62	22	22	1
501-600	50	30	11	2
601-700	57	23	10	4
701-800	52	24	12	1
801-900	50	26	8	2
901-1000	62	22	21	1

In what follows, data is presented starting with a presentation of the present progressive and past progressive segments in 2.1.2.1 and 2.1.2.2, followed by the presentation of the three cluster groups in 2.1.2.3. In 2.1.2.4, the temporal restrictions and preferences of grams are given, and in section 2.1.2.5 the notion of ‘while’ in connection to a number of grams in one of the clusters is discussed. In 2.1.2.6, peripheral uses of a number of grams are presented. Section 2.1.3 offers a section summary.

2.1.2.1 Top segments in the present

The segments in the top list of the gram set that are translated to the present tense in English have FOC contexts. They are often aimed at turning the attention of the listener towards, or at making the speaker aware of an acute and imminent situation. The FOC point may be explicitly given by an element such as *now* indicating that the event is happening right now or by *behold* (modern English equivalent 'look' or similar), which aims to turn the attention of the addressee towards or make them aware of an ongoing event[21], or it may be implicitly given. We already saw such uses in 2:3a). An example is given in 2:4, where no explicit FOC point is given. A little more than half of the progressive grams occurred in this segment.

2:4 42018037$11 0.573

And they told him, "*Jesus the Nazarene is passing by*."

In order to understand this verse, we can look at the two preceding verses: *As he drew near to Jericho, a blind man was sitting by the roadside begging. And hearing a crowd going by, he inquired what this meant*. Thus, the utterance containing the progressive is an explanation of what is happening to a blind man who hears the noise that the crowd is making. It is then evident that the event of Jesus passing by is FOC in the sense that it is explained as happening at the moment of speech. Some other examples of segments that are translated to the English present tense are given in 2:5. Examples 2:5a), b) and c) are quotes meaning that they illustrate spoken language.

2:5 Present progressive segments

a) 43009037$20 0.685
Jesus said to him, "You have both seen him, and he is the one *who is speaking with you*."

b) 41001037$13 0.64
And they found him and said to him, "*Everyone is looking for you!*"

c) 41010033$6 0.618
"*Behold*, *we are going up to Jerusalem*, *and* the Son of Man will be handed over to the chief priests *and the scribes*, *and* they will condemn him to death and will hand him over to the Gentiles.

[21] Interestingly, in discussing the English progressive, Croft (2013:152–154) mentions several subtypes, the first one, referred to as "undirected activity", is exemplified by *Look! It's flying*. However, Croft does not comment on the use of *Look!* here.

d) 52005011$16 0.584
Therefore encourage one another and build up each other, *just as indeed you are doing*.

2.1.2.2 Top segments in the past

Some examples of segments translated to the past progressive in English from the top list were given in 2:3. As already noted, the majority of segments in which the progressive grams occur in the past are backgrounding contexts. These segments often start with 'while' or 'as' in the English translation. These types of sentences have the following structure: an ongoing event provides the background to another event which occurs at the same time or interrupts the event to which the progressive applies. The foregrounded event is typically punctual or perfective and provides the FOC reference time for the progressive. In a typical present progressive, such as *For behold, he is praying*, the FOC point coincides with the speech moment. In a typical backgrounding past ongoing event such as *While he was praying* [*heaven was opened*], the FOC reference point is provided by a telic event, here *heaven was opened*, from which the ongoing event is viewed.

Quite often, these sentences are accompanied by *still* in English so as to emphasize that the event expressed with the progressive had started earlier and was ongoing at the time of the second event. As was seen in examples 2:3c) and e), the event expressed by the progressive gram, i.e. the speaking, is ongoing as another event interrupts it, i.e. the overshadowing of the cloud and the arriving of Judas, respectively.

Backgrounding contexts are typical in narrations, of which the New Testament is full. Although the top segments in the past are often backgrounding in the sense of setting the scene for a telic event, there are also some uses of progressives in the Bible corpus that do not fit that characterization, such as in 2:6. Here the FOC reference point seems to coincide with the event expressed by the progressive.

2:6 41009031$4 0.562

> And from there they went out and passed through Galilee. And he did not want anyone to know, *for he was teaching his disciples and* was telling them, "The Son of Man is being betrayed into the hands of men, and they will kill him. And when he is killed, after three days he will rise."

As will be seen, in the top lists of the TED corpus, past ongoing events that are not backgrounding are also found.

2.1.2.3 The red, green and yellow cluster groups

The areal distribution of the three clusters is shown on the map in Figure 3. As can be seen, while the green and yellow cluster groups are spread out, the red cluster group is focused in Southeast Asia, mainly in Indonesia and its surroundings.

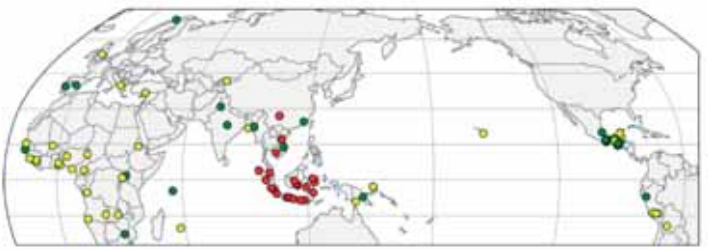


Figure 3. Areal distribution of the red, yellow and green clusters

In Table 5, statistical data regarding the clusters and the whole gram set is provided. As can be seen, the more frequent grams in the gram set are clustered in the yellow group, while the red group includes grams with the fewest occurrences. In the table, the mean recall and precision value of each cluster is given, as well as the mean recall and precision value of each cluster when run against the whole gram set.

Table 5. Statistical data, at the cluster level and for the whole gram set

	YELLOW	GREEN	RED	GRAM SET
NO. OF GRAMS	36	31	22	89
MEAN GRAM SIZE	1937	870	234	1145
MEDIAN GRAM SIZE	1939	794	186	954
MEAN RECALL, GRAM SET	0.59	0.82	0.70	0.69
MEAN PRECISION, GRAM SET	0.80	0.86	0.95	0.86
MEAN RECALL, CLUSTER	0.76	0.85	0.83	-
MEAN PRECISION, CLUSTER	0.75	0.78	0.82	-

Notably, the yellow cluster has a lower mean and median value for recall and precision than the other groups, both as a cluster and for the whole gram set. At cluster level, this means that it is a weaker cluster than the other two clusters, i.e. the overlap in distribution between grams is smaller than the overlap in distribution between the grams in the other groups. As will become evident, the yellow group includes several grams that can be viewed as progressives with peripheral uses or, at times, as grams that cannot be seen as indisputably belonging to the progressive gram type only. These will be mentioned in section 2.1.2.6. The

low recall value for the whole gram set of the yellow cluster reflects the fact that there are many grams in this cluster with a preference for or restriction to present time reference. In fact, the top 20 list of the yellow group only includes segments that are in the English present tense. This does not mean that these grams are all present progressive grams, rather, it means that it is the present ongoing use that they have in common.

The top 20 list of the green cluster almost exclusively includes segments that are translated into the past tense in English, with only one segment in the present. This is similar to the top 20 list of the red cluster, which only includes past progressive segments with the exception of two present progressive segments and one future segment. The red cluster mainly includes grams from the Austronesian language family, with a few Austro-Asiatic and one Hmong-Mien language. As will be discussed in section 2.1.2.5, some of the grams in the red cluster seem to have a meaning of 'while V-ing'.

2.1.2.4 Temporal restrictions and preferences

It was previously mentioned that the grams in the gram set occur more often with the Esperanto present tense than with the past. In this section, the temporal restrictions and preferences of individual grams will be presented. For each gram, the ratio of what corresponds to Esperanto present and past is calculated. To give an example, of the 317 occurrences of the Achinese (Austronesian) progressive gram, 59 correspond to Esperanto present tense and 168 correspond to Esperanto past tense, giving the ratio of present over past as 26%, i.e. 26% present and 74% past, thus indicating a clear preference for the past. Table 6 shows a classification of the grams according to their present-past correlation.

Table 6. Ratio of present over past

RATIO OF PRS PROG OVER (PRS PROG + PST PROG)	TEMPORAL PREFERENCE	NO. OF GRAMS	
		WHOLE CORPUS	DIRECT SPEECH
$0 \leq x > 20$	high past tense preference	5	1
$20 \leq x > 40$	past tense preference	16	1
$40 \leq x > 60$	no temporal preference	39	0
$60 \leq x > 80$	present tense preference	22	21
$80 \leq x \geq 100$	high present tense preference	7	66

The data in the last two columns of the table are also presented in Figures 4 and 5, where Figure 4 includes the whole corpus and Figure 5 the preference for present over past in direct speech. We then see a clear preference for present over past in that progressive grams generally have a majority of uses in segments with present time reference in our gram set. In direct speech, all grams apart from the grams in Bawn Chin and Jola-Fonyi, have a majority of present uses.

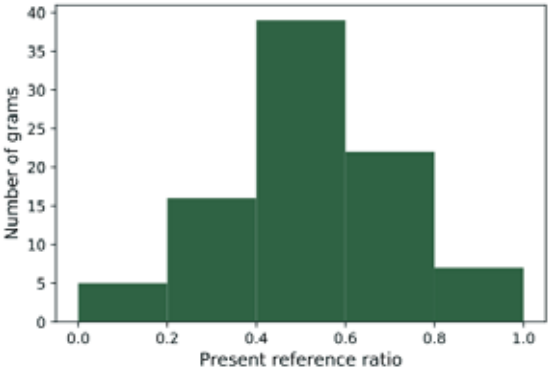


Figure 4. Ratio of present time reference for progressive grams

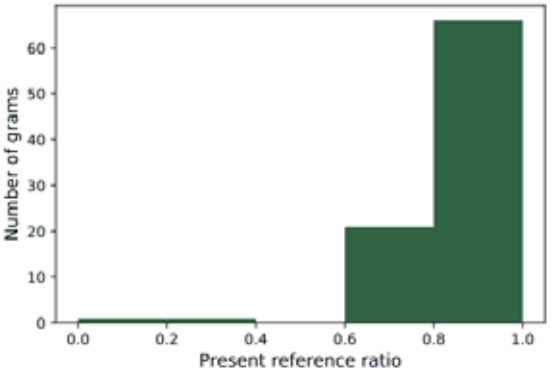


Figure 5. Ratio of present time reference for progressive grams in direct speech

The present-past preference from Table 6 can be illustrated in a map, as in Figure 6, where a darker red indicates a higher ratio of present over past in the Bible corpus.

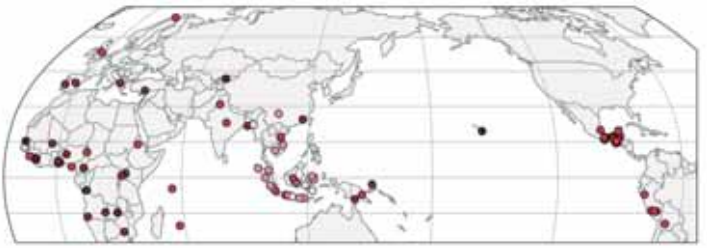


Figure 6. Temporal preference with darker red indicated higher ratio of present over past

As can be seen on the map in Figure 6, the higher correspondence with Esperanto past tense is mainly found in Southeast Asia. These grams are grams of the red cluster group from the Austronesian language family, apart from three non-Austronesian languages, namely Central Khmer (Austro-Asiatic), Vietnamese (Austro-Asiatic)[22] and Hmong Daw (Hmong-Mien). Interestingly, most of these languages are traditionally analyzed as lacking grammatical tense distinctions. It may be that having a majority of past uses is an areal feature. Grams that have a majority of present uses in the New Testament are more spread out, suggesting that cross-linguistically it is less common to have past preference even in written sources. In addition, in direct speech, almost no grams have a majority of past uses.

It was previously noted that most of the grams with a majority of present uses in the Bible corpus belong to the yellow cluster. Grams with no temporal restrictions and preferences are found in all clusters, in fact, in the green cluster, most grams are such grams. The data indicates that even if we might suspect that almost all progressives have a majority of present uses in spoken language, there are grams that have a stronger present preference than others. In what follows, we will look at some temporal restrictions and preferences for the whole Bible corpus.

2.1.2.4.1 No temporal restrictions or preferences

Many grams in the gram set occur in segments of the Bible corpus corresponding to the Esperanto present and past tense without any noted temporal preference. Some of these also have no noted peripheral uses, meaning that they have no occurrences, or close to none, in the future, performative-like, habitual or stative segment sets. Such grams are, for example, *taatmwɨ* in Angaataha (Trans-New Guinea), *rah-* + COP in Hindi (Indo-European) and *rah-* + COP in Panjabi (Indo-European). As illustrated in example 2:7, the Panjabi *rah-* is inflected for number and gender while the copula shows tense. Thus, the gram is not restricted to the present or past, and our data suggests that it occurs more or less equally often in both tenses in the Bible corpus. Notably, the Hindi progressive occurs much less often (210 occurrences) in the Bible corpus than the Panjabi progressive does (1327 occurrences).

2:7 Panjabi, Indo-European (Bhatia 1993:254)

mãi	*so*	*ríaa*	*ãã/sãã.*
I	sleep	PROG.M.SG	am/was

'I am/was sleeping.'

[22] Both *đươ̛ng* and *đang*.

Almost all Mayan grams turn up approximately as often in the present as in the past. Examples from Popti' and Uspanteco are shown here. In Popti', the progressive marker *lañan* seems to have the original meaning 'long and flexible' (Day 1973:32).

2:8 Popti', Mayan (Craig 1977:63)

Lañan *ha-wayi*.
PROG ERG2-sleep
'You are sleeping.'

In Uspanteco the progressive marked with *tijin* occurs in two constructions, where one, 2:9b), includes the preposition *chi*.

2:9 Uspanteco, Mayan (Pixabaj 2006:210–211)

a) Ø *tijin* *in-atín-ik*.
ABS.3SG PROG ABS.1SG-bathe-CS
'I am bathing.'[23]

b) *In* *tijin* *chi* *wiik*.
ABS.1SG PROG PREP eat
'I am eating.'[24]

No Quechua languages and almost no Zapotecan languages show temporal restriction or preference.

2.1.2.4.2 Present progressive grams

There are some grams that almost exclusively correspond to Esperanto present tense in the Bible corpus. There are reasons to believe that some of these are present progressive grams. As suggested in Dahl (1985:94), grams restricted to present time reference can alternatively be viewed as constituting a separate type, i.e. a separate gram type. In, e.g., Hawaiian (Austronesian), a gram *ke* + *nei* marks the progressive.[25] This pattern was previously noted in Dahl (1985:94) as a present progressive gram. An example is given in 2:10.

[23] Orig. 'me estoy bañando'.
[24] Orig. 'estoy comiendo'.
[25] The *nei* marker also occurs in other patterns as in *aku nei*, which marks the distant past, or in *a'e nei*/*aho nei*, which marks the recent past (Elbert & Pukui 1979:60, 92). However, a search for only *nei* in our corpus did not yield any past progressive occurrences, meaning that there is no past progressive pattern with *nei* in the Bible corpus for the language. Attempts were also made to expand the search to *ke*/*ka* + *nei*/*la*, but these combinations did not have a high enough recall and precision value.

2:10 Hawaiian, Austronesian (Elbert & Pukui 1979:60)

Ke kali nei au.
PRS wait now I
'I'm waiting.'

Also the Turkic languages Kirghiz and Uzbek have grams that seem to be present progressive grams. In Kirghiz, *žata-* 'lie' occurs almost exclusively in the present tense, Abylkasymova & Jumabaeva (1997:306) refer to this gram as the 'complex present tense'.[26] The situation in Uzbek is more complex. The Uzbek progressive gram includes two forms, *-jap-* and *-yotgan edi-*, the former given as the 'focal present' and the latter as the 'focal past', which is why they are put together as constituting one gram in this investigation.[27] However, *-yotgan edi-* only occurs 47 times in the Bible corpus, whereas *-jap-* occurs 473 times. This means that the present form is used much more often than the past form, thus the distribution of these two patterns has a clear majority of present uses.

Several Niger-Congo languages exhibit grams having a high correlation with the Esperanto present tense. For example, in Lenje a present progressive gram *-too-* was identified, referred to as "present progressive and progressive where the action continuous from present to immediate future" by Kagaya (1987:24–25). According to Kagaya, however, there are also other patterns covering the ongoing function, such as *-limu ku-* for the present progressive and -(*a*)*li*(*nga*) *ku-*[28] for the present and past progressive. However, there are no occurrences of *-limu ku-* in the Bible corpus, and -(*a*)*li*(*nga*) *ku-* only occurs 17 times and has low recall and precision values. The data then suggests that in the Bible corpus, there is one gram *-too-* that has the present ongoing function, with essentially no corresponding past progressive. Other Niger-Congo languages with grams that occur most often in the present tense are Koongo with *-eti*, Wolof with *ngi* and Southern Kisi with *chō*. The captured Wolof *ngi*[29] marker is referred to as a 'present continuous/presentative' (Dem 1995:4). In the Bible corpus, this marker mainly occurs in present ongoing functions. Interestingly, the pattern also

[26] There is also mention of *otur-* 'sit' (Abylkasymova & Jumabaeva 1997:306) as well as other potential progressive markers, but none of these have high enough recall and precision values to be included in the sample.
[27] The manuscript 'The Uzbek tense/aspect/modality system' in which this information is found has been found online without an author. It is listed in the bibliography, attributed to Anonymous.
[28] Where -(*a*)*li*(*nga*) is not preceded by *ak-*.
[29] It is also noted that the grammatical description gives *ng-* + demonstrative, therefore *nga* and *ngay* were also searched for, but these rendered recall and precision values too low to be included.

occurs in the proximative function. The present progressive, presentative and proximative uses of *ngi* are illustrated in 2:11, in that order.

2:11 Wolof, Niger-Congo (Robert 1989:166, 172 mg; Diouf 2009:149 mg)

a) *Waj bi mu ng-i lekk yàpp.*
dog the it PSNT-PRX eat meat[30]
'The dog (nearby) is eating meat.'

b) *Mu ng-ii.*
it PSNT-PRX
'Here it is.'[31]

c) *Mu ng-i daanu!*
It PSNT-PRX fall[32]
'It is about to fall!'

2.1.2.4.3 Grams with a majority of present uses

Many of the grams of the gram set are not restricted temporally but have a majority of present uses, as shown by their higher occurrence in the segments corresponding to the Esperanto present tense rather than past tense, some of which will be mentioned here. For example, three Mayan grams, namely *woli-* in Chol, *-tzan-* in Aguacateco and *yac-* in Tzeltal, have more occurrences in the present than in the past. Also, the Spanish and Portuguese *estar* + GER[33] progressive grams occur more often in the present segments than in the past ones.

2:12 Tzeltal, Mayan (Polian 2013:169)

Yak(*al*)*-on ta way-el.*
PROG-1ABS PREP sleep-INF[34]
'I am sleeping.'

[30] Orig. Chien le il prés…prox. manger viande 'le chien (à proximité) est en train de manger de la viande'
[31] Orig. 'le voici'.
[32] Orig. Il prés…prox. tomber! 'il va tomber!'
[33] The Portuguese *estar* + *a* + INF, which is included in the TED corpus, is not included in the gram set. It does, however, occur in the *A Bíblia para todos* were it has a low recall value but a high precision value. It has no occurrences in the *Nova Tradução na Linguagem de Hoje* (2000) Bible translation which is the one used for the gram set.
[34] Orig. PROG-B1 P dormir-NF 'Estoy durmiendo.'

Also, several grams in Niger-Congo languages have a majority of present uses, such grams are *-tɛn* in Busa, *le* in Gen, (-)*lɛ* + V in Sekpele[35], *khou* in Venda and the verbal suffix *-ẁwɔ̀*, *-gùwɔ̀* or *-ŋ̀wɔ̀*, extracted as V-*wɔ-* in this work, in Toro So Dogon. A majority of present uses has previously been noted for Gen: in Jondoh (1980:37), the verb *le* 'be' is noted as marking the 'continuous aspect' with or without *wɔ̀* mainly in the present tense.[36]

In Toro So Dogon the marker is given as a present progressive, while the general past imperfective *=bè* also covers the past ongoing and past habitual functions (Heath 2014:186, 201). In this investigation, however, *-wɔ-* correlates with Esperanto past tense as well, but not to a great extent, suggesting that it is a progressive gram with majority present uses.

Similarly, the gram *-tye* + V[37] in Kumam (Nilo-Saharan) occurs both in the present and the past, but more often in the present.

2:13 Kuman, Nilo-Saharan (Hieda 2011:39)

A=tye	*kók.*
1SG=IPFV:be	cry.INF
'I am crying.'	

The Turkic grams in Kirghiz and in Uzbek were discussed above where it was noted that they occur (almost) exclusively in the present tense (the Uzbek past form has very few occurrences). The Turkish *-Iyor-*[38], however, is not restricted to present nor past. It has, however, more occurrences in the present than in the past. This preference is strengthened in our data by the fact that 'while' in Turkish, marked with *-ken*, more often combines with the aorist and "more rarely" with *-Iyor-* (Kornfilt 1997:361). Many grams with a majority of present uses, including Turkish *-Iyor-*, also have peripheral uses and will be further discussed in section 2.1.2.6.

2.1.2.4.4 Grams with a majority of past uses

Most of the grams with a majority of past uses in the Bible texts belong to the Austronesian language family. In fact, among the Austronesian languages, only grams with no temporal preference or a majority of past uses were found. This is

[35] Also called Likpe, e.g. in Ameka (2006).

[36] The exclusion of *wɔ̀* results in a better recall and precision value. Jondoh (1980:37) also notes that *nɔ̀* is used in all other tenses, 'in all aspectual constructions and with all modals'. This marker does not have good recall and precision values in this investigation, also, the examples that are given are not mainly progressive.

[37] The verb form is probably always in the infinitive, but the search is for any kind of verb since the infinitive is not easy to capture.

[38] The *I* in *-Iyor-* represents an archiphoneme "whose missing feature value [is] predictable by general phonological rules (Kornfilt 1997:xxiv). The search string used in this chapter is *-yor-*.

probably, at least partly, due to preference for subordination rather than temporal reference, an issue that will be discussed in 2.1.2.5. In addition to the Austronesian languages, other languages spoken in Southeast Asia also have a majority of past uses: *kampoung-* in Central Khmer (Austro-Asiatic), *đương* in Vietnamese (Austro-Asiatic) and *tabtom* Hmong Daw (Hmong-Mien). These grams all belong to the red cluster group.

The only gram with a majority of past uses not belonging to this group is found in Ozolotepec Zapotec, where a prefix *nge-* is identified, a marker not found in Heise (2003), who instead provides the prefixes *ch-*/*x-* as progressive markers. The latter prefixes do not have high recall and precision values in our data and are therefore not included in the gram set.

There are some grams that co-occur with the Esperanto past tense but do not have high correspondence with the present tense. These are *tangasano* in Muna[39] (Austronesian), *liau*(*ah*) in Bawm Chin (Sino-Tibetan), *-lako-* in Jola-Fonyi (Niger-Congo), *mahassa-* in Bambam (Austronesian) and *teppana* in Madurese (Austronesian). The latter two have some, but few, occurrences in segments corresponding to Esperanto present tense.

In Jola-Fonyi, *-lako-*, which seems to originate from 'stay, sit' (Sapir 1965:104), is given as a past progressive in Hopkins (1995:149). Our data also points in this direction. However, the grammar also mentions other progressive patterns, such as complex patterns formed with nominal infinitives, with the pronominal presentative *-oo-* (with or without *di*). These patterns were either not possible to capture or did not attain high recall and precision values. It is then possible that the language has progressive patterns also covering the present tense that are not included in this investigation.

The Bawm Chin *liau*(*ah*) shows a rather high correlation with segments corresponding to Esperanto past tense in the Bible text, although the language does not mark present and past tense grammatically (Reichle 1981:61). In Reichle (1981) progressive examples with present time reference are also given. Thus, this gram is not restricted to the past, but occurs predominantly in these contexts in the Bible corpus. Examples of *liau*(*ah*) in contexts with present and past time reference are given in 2:14.

2:14 Bawm Chin, Sino-Tibetan (Reichle 1981:62–63)

a) *A* *hawng kal* *liau*.
He come PROG
'He is coming.'

[39] Van den Berg (1989:160) also provides a marker *naando/naandoo* preceding a verb as marker of the progressive, but this pattern does not have good recall and precision values.

b) *Buh ka ei liau asi ruangah ka kal kho lo.*
rice I eat PROG ASS because I go can not
'Because I was eating rice I could not go.'

The Muna *tangasano* occurs almost exclusively in segments that correspond to the Esperanto past tense, although this language does not mark tense grammatically either. Since many of these segments start with 'while' or 'as' in the corpus, they are also subordinated. In fact, René van den Berg (p.c.) comments that this marker is probably an "adverb marking continuative or progressive aspect irrespective of tense, typically used in subordinate clauses". Two examples from the Bible corpus are given in 2:15 where a) is translated into the English present progressive (the top segments of the general top list) and b) is translated into the future-progressive tense.

2:15 Present and future time reference with *tangasano*

a) 44009011$39
Nokowambamo dua Ompu, Kalamo we kaangka'a konea'ano Sala Melaa. Ondofi we lambuno Yudas seemie mai'aono we Tarsus, neano Saulus. *Ampa aitu tangasano nosambahea.*
'And the Lord said to him, "Get up, go to the street called 'Straight' and in the house of Judas look for a man named Saul from Tarsus. *For behold*, *he is praying*,'

b) 42017035$3
Ane dahodua hobhine tangasano megilino kahitela, seemie dawowo'oe, seemieno dahumunsae.
'*There will be two women* grinding at the same place; one will be taken and the other will be left.'

I will discuss *tangasano* alongside other progressive markers in Austronesian languages in section 2.1.2.5. It then seems as if only Jola-Fonyi *-lako-* may be a clear case of a past progressive gram in our gram set.

2.1.2.5 'While' and the progressive

In this section, some grams in the red cluster are discussed, as well as some additional patterns that may be of interest to this discussion. Interestingly, none of the grams in the red group occurred to any extent in the future, performative-like, habitual or stative segment sets. Some of these languages are under-described, but looking at those for which there are descriptions available we see that in several languages the progressive is marked with a word meaning 'middle' or 'half' or similar. For example, *pintanga'* in Balantak (Austronesian) is glossed

as ‘in.the.middle’ in the grammatical description. As illustrated in example 2:16, it is translated as ‘while’ followed by a progressive pattern in English, although there is no word for ‘while’ nor any other progressive marker in the Balantak sentence.

2:16 Balantak, Autronesian (van den Berg & Busenitz 2012:32)

Pintanga'	*bisara*	*koi-ya'a,*	*uar-kon-on-mo*
in.the.middle	speak	like-DEM3	go.out-CAUS-PV.1-PRF

a	*panganon*.
ART	proposal.gift

‘While they are speaking thus, the marriage proposal gifts are taken out.’

According to the grammatical description, *tanga* has the meaning ‘middle’. There are a number of grams in the red group that seem to have a similar meaning: in the closely related Muna (Austronesian), the gram is *tangasano*, which, as already mentioned, typically marks ongoing uses in subordinate clauses. Additionally, *se-tanga* has the meaning ‘a half’ (van den Berg 1989:113). René van den Berg (p.c.) notes that *tanga-* is probably related to the Indonesian/Malay *tengah* ‘middle’, in addition, he notes that this word may be a loan from the more prestigious language Wolio (Austronesian), which is not included in our gram set.

Similarly, for Achinese (Austronesian), *teungoh* is given in Asyik (1987:167–168, 115), both as the marker of progressive and with the meaning ‘middle/in the middle of, during’. It seems reasonable to assume that the Balantak, Muna and Achinese grams have arisen from words meaning ‘middle’ or similar.

2:17 Achinese, Austronesian (Asyik 1987:167–168)

a) *Ayah*	*teungöh*	*geu-peugah*	*haba ngön jamèe*.
father	PROG	3-tell	story with guest

‘Father is/was talking with a guest.’

b) *Beuklam*	*jaga-lōn*	*teungöh*	*malam.*
last-night	awake-1	middle	night

‘Last night I awoke in the middle of the night’

Interestingly, Sneddon (2006:89) and Sneddon et al. (2010:205) report that in Indonesian (Austronesian) a marker *tengah* ‘middle’ as well as *lagi* ‘again, still, more, other’ can be used for the ongoing function in addition to *sedang*, although they are not as frequent as the latter. *Sedang* is also found in the Austronesian languages Lampung Api, Standard Malay and Minangkabau, included in the gram

set.[40] These words may also be related to Vietnamese (Austro-Asiatic) and Bengkulu Malay (Austronesian) *đang/dang*. The Indonesian *sedang* has an additional meaning of 'medium, moderate, average' (Intan Fuji, p.c., Google translate). It is then possible that *sedang* and *đang*/*dang* as progressive markers are also related to a meaning 'middle' or similar.

In addition to the meaning of 'middle', there also seem to be a connection between the meaning 'while' and some other grams in the red cluster. *Kenyeke* in Sasak (Austronesian) is reported to mark 'in the process of' (Austin 2012:241–242), but it is also translated as having such meanings as 'while' or 'as' (Eades 1998:122). Similarly, in Standard Malay, *sedang* also has the additional meaning of 'while, as'[41], and Vietnamese *đương* is also translated as 'in the act of, during, while' (Thompson 1987:270).

There is then a connection between the meaning 'middle', 'while' and the progressive grams that have a majority of past/subordinate uses in the Bible corpus. The impression is that some of these grams have a meaning of 'while V-ing'. Speculatively, such patterns could have grammaticalized from the meaning of 'while in the middle of V-ing' or 'while engaged in V-ing'. This would then explain why some of these grams also mean 'while' and have preference for subordinate past ongoing uses in the Bible corpus.

In Chapter 3, the Persian *mašġul* 'busy' + INF is noted to have partial overlap with the *dāštan* progressive. It has a semantics very close to the English 'busy' when used with a gerund of the verb, i.e. 'being busy V-ing'. Interestingly, in Central Khmer (Austro-Asiatic), we find *kampoung-*, which is described as both a progressive and as having the meaning 'be engaged in/busy in' (Haiman 2011:267). What is more, in both Indonesian and Persian, the words for 'while' are derived or have a similar form to the respective progressive patterns. In Indonesian, we find *sedang-kan* 'while'[42] as well as *lagi pas* 'while'. *Lagi* is a marker, which in addition to the ongoing function, also has other meanings, one of them being 'while' with or without *pas* 'when'. Sneddon (2006:93) comments that *pas* "is often followed by *lagi* 'action in progress' […] to specify that the event occurred when something else was in progress", which could be a description of a backgrounding context. As can be seen, *lagi* is glossed as 'still', which is one of its meanings.

[40] Despite the resemblance, René van den Berg (p.c.) does not think that the Muna [mnb] (Austronesian) *se-tanga* 'a half' is related to the Indonesian *sedang*.
[41] Found on Wiktionary.
[42] This investigation has captured a freestanding marker *sedang*, thus excluding *sedangkan* 'while'.

2:18 Indonesian, Austronesian (Sneddon 2006:93)

Pas	*lagi*	*cerita*	*trus*	*gua*	*tiba-tiba*	*ketawa.*
when	still	tell.story	then	I	suddenly	laugh

'While I was telling the story I suddenly laughed.'

The Persian *dar hāl=i ke* 'while', where *ke* is the subordinator 'that', is similar in its form to the *dar hāl=e* progressive. An example from the Bible corpus is given in 2:19, where 2:19a) shows the progressive pattern and b) the 'while' pattern. The semantic closeness of the FOC ongoing meaning and 'while' may be assumed to give rise to these similarities in form.

2:19 Persian, Indo-European

a) 62002008$33

[…]	*nur=e*	*haġiġi*	*dar*	*hāl=e*	*deraxšidan*	*ast.*
	light=EZ	true	in	state=EZ	shine.INF	COP.3SG

'[…] the true light is already shining.'

b) 44001009$12

[…]	*dar*	*hāl=i*	*ke*	*hame*	*negāh*	*mi-kard-and* […]
	in	state=INDF	that	all	look	IPFV-do.PST-2PL

'[…] as they were looking […]'

2.1.2.6 Grams with peripheral uses

In section 2.1.1.3, several comparative segment sets were presented, containing segments with future, performative-like and habitual uses as well as segments containing stative predicates.[43] The grams of the gram set were run against these segment sets with the aim of capturing peripheral uses of the grams. In order to exclude false positives, the segments were checked manually for all grams with at least 15% occurrence in these segment sets. In this section, I will discuss cases that seem fairly clear as well as mention cases where these uses are less clear.

[43] Attempts to check for proximative uses discussed in Chapters 3, 4 and 5, failed as it was difficult to establish whether the use of a gram has an intended reading as proximative or ongoing. For example, looking at different English translations, it was noted that there were some segments where the 'about to' pattern was used in one translation and the progressive pattern was used in another translation. For what it's worth, the grams in the Niger-Congo languages Kuwaa, Susu and Gokana occurred more often than other grams in segments that are typically translated with 'about to' in English. Whether or not this reflects proximative use needs to be confirmed.

2.1.2.6.1 Future segment set

In the future segment set, the grams *de* + V-*nù*[44] in Kuwaa, *chō* in Southern Kisi (Niger-Congo), *yac-* in Tzeltal (Mayan) and *-sha-* in Eastern Apurímac Quechua (Quechuan) occurred often. Two examples of segments with future uses in which some of these grams occur are given in 2:20. In 2:20a) the Southern Kisi, the Tzeltal and the Kuwaa grams occur, in 2:20b) the Tzeltal and the Eastern Apurímac Quechua grams occur. As illustrated, these are typical contexts for future grams. Although the method cannot ensure the exclusion of cases where grams are combined with future markers, this is not the type of segment included in the English translation of the segments in the future segment set.

2:20 Future segments

a) 43016020$20
Truly, truly I say to you, *that you will weep and lament, but the world will rejoice*; you will become sorrowful, *but* your sorrow will change to joy.

b) 40010021$5
"And brother will hand over brother to death, and a father his children, and children will rise up against parents *and* have them put to death,

Interestingly, the gram *chō*, or *co* as it is given in Childs (1995), in Southern Kisi is described as a marker that is used both as the present progressive and for marking the future tense, an example of which is given in 2:21. The corpus data presented here confirms this description.

2:21 Southern Kisi, Niger-Congo (Childs 1995:117)

a) *ŋ̀ cò cììikíàŋ lɔ́ɔ́ ŋ̀ cò hùnɔ̀ɔ-ó.*
we AUX meet time you AUX come-SUF
'We will see you when you come.'

b) *sàà cò ndú tàmbá lòòlùlló.*
Saa AUX her Tamba beat
'Saa is beating Tamba for her.'

[44] Marchese described the progressive as consisting of *de* 'be at' and the nominalizer *nu*. The automatic search only captured V-*nù* but the segments in the lists were checked so that *de* 'be at' precedes the nominalizer.

2.1.2.6.2 Segment sets with performative-like uses

Some of the members of the gram set frequently occur in segments with performative-like uses. These are *n(i)-…-a*[45] in Nyoror (Niger-Congo), *ngi* in Wolof (Niger-Congo) and V-*eni* in Bine (Western Fly). In 2:22a) all three patterns occurred, in 2:22b) the Bine and Nyoro patterns occur.

2:22 Performative-like segments

a) 46001014$2
I give thanks that I baptized none of you except Crispus and Gaius,

b) 44016018$27
And she was doing this for many days. But Paul, becoming greatly annoyed and turning around, said to the spirit, *"I command you in the name of Jesus Christ* to come out of her!" And it came out immediately.

There are also grams that occur in the segments with performative-like uses but less frequently than the grams just mentioned. The performative uses of these grams are more difficult to establish. These are grams belonging to Toro So Dogon (Niger-Congo), Gen (Niger-Congo), Koongo (Niger-Congo), Hawaiian (Oceanic), Merey (Afro-Asiatic), Turkish (Turkic) and Coatecas Altas Zapotec (Oto-Manguean).

It is probably the case that there are many segments in our corpus with performative function. Performative uses are perfective in the sense that the event becomes completed as it is uttered. Therefore, progressive patterns that also occur in these contexts can be viewed as progressives that are expanding and taking over uses that are not typically covered by progressives. For those languages that typically use their simple present in performative contexts, the use of a progressive in these contexts could be seen as part of the expansion of the progressive towards present tense. Notably, most grams mentioned here have a majority of present uses, among which are also what has been assumed to be the present progressive grams in Hawaiian and Wolof.

2.1.2.6.3 Habitual segment set

The occurrences of the grams in the habitual segment set are difficult to analyze, since many of these segments also allow for a FOC interpretation. This means that

[45] The progressive is marked by the prefix *ni-*, whose vowel may be elided before a subject marker beginning with a vowel (Rubongoya 1999:219). The search string also includes the subject marker and the final vowel *-a,* which may be seen as an indicative marker. In Rubongoya (1999:242), a form *-ruku-* is presented for negated and relative clauses. This marker, however, does not seem to have the same distribution as the progressive grams in the Bible corpus.

when these contexts contain one of the grams of the gram set, it is difficult to establish whether the use has a generic reading or FOC one. Therefore, only those grams will be mentioned here that also occur in segments that have other elements indicating habituality, such as 'always', 'often' or 'every year'. These are *faya* in Merey (Afro-Asiatic), *-Iyor-* in Turkish (Turkic) and COP + *naku-* in Mbunda (Niger-Congo). In 2:23a) the Merey and Turkish grams occur, in 2:23b) the Mbunda gram occurs.

2:23 Habitual segments

a) 40009014$13
Then the disciples of John approached him, saying, "Why do *we and the Pharisees fast often, but* your disciples do not fast?"

b) 42002041$4
And his parents went every year to Jerusalem for the feast of the Passover.

For *de* + V-*nù* in Kuwaa (Niger-Congo), Marchese (1986:65, 66) notes that although not much data is available, it appears as if the progressive has extended its uses to also express habitual actions. This gram occurs in 26% of the segments in the habitual segment set.

2.1.2.6.4 Stative segment set

In the Bible corpus, combinations of the members of the gram set with stative verbs were mainly limited to posture verbs. Thus, all Quechua languages of the gram set, *-Iyor-* in Turkish (Turkic), *taxa* in Kara (Austronesian) and the English progressive occurred with segments containing the posture verbs *stand*, *lie*, *sit* and *stay*. Two such segments are given in 2:24.

2:24 Segments with posture verbs

a) 40026069$4
Now Peter was sitting outside in the courtyard, and a female slave came up to him and said, "You also were with Jesus the Galilean."

b) 66007009$31
After these things I looked, and behold, a great crowd that no one was able to number, from every nation and tribe and people and language, *standing before the throne* and before the Lamb, dressed in white robes and with palm branches in their hands.

Dahl (1985:94) has previously noted stative uses with *-sha-* in Quechuan. Similarly, such uses are known for Turkish and English. Notably, in our data, the Cajamarca Quechua (Quechuan) gram *-yka-* also has such uses. This gram differs from the other Quechuan grams, not only in its phonological form, but also in that it has a lower frequency than the other ones. For Kara, examples of *taxa* with the verb *stand* are given in Dryer (2013:28, 242). Dryer (2013:201), who also (mainly) uses the Bible as his source, refers to *taxa* as a "continuative" and notes that it "indicates a continuing act or state […] often occurring where the English translation uses the progressive". In 2:25, examples of *taxa* as indicating ongoingness, with the verb *stand* and indicating "continuing state", are given.

2:25 Uses of *taxa* in Kara, Austronesian (Dryer 2013:201, 242)

a) *Mi taxa seng nase'*?
2PL CONTIN look.for who
'Who are you looking for?'

b) *E tamo ta mataa i taxa tigina xulu-na lifu* […]
and if NONS man 3SG.INC CONTIN stand on-3SG house
'And if a man stands on the roof of the house […]'

c) *Rutul taxa waan xulu-na xaati* […]
3TRI CONTIN be.at on-3SG boat
'They were in a boat […]'

There are also a few grams that occur with posture verbs where the data is harder to interpret. These are the grams in Kuwaa (Niger-Congon), Kumam (Niger-Congo), Morisyen (Creole) and Zacatlán-Ahuacatlán-Tepetzintla Nahuatl (Uto-Aztecan).

Turkish is an exception to the generalization that the members of the gram set are not found with non-posture verbs, as Turkish *-Iyor-* occurs with 'know' and 'understand' as well. This is to be expected since previous research has noted such uses. In fact, Johanson (2000:39, 90) refers to *-Iyor-* as a "low focal item", whereas progressives are viewed as "high focal". Low focal items are, for example, simple presents and past imperfectives. This means that *-Iyor-* has expanded towards the imperfective, or it has gone from high focal to low focal, in Johanson's (2017:101) terminology. Although it occurs in several of the segments sets, its general distribution in the Bible corpus is similar to other progressive grams as its general precision value is 76%. As will be shown in section 2.2.2.3, however, its distribution in the TED corpus is quite different from other progressives. I will assume that this similarity is an outcome of a more restrictive use of *-Iyor-* in the Bible. In example 2:26, the uses of *-Iyor-* in the present, past and with a stative verb are illustrated.

2:26 Turkish, Turkic (Kornfilt 1997:34, 341, 218)

a) […] *tiyatro-ya* *gid-iyor-um*[46].
theater-DAT go-PROG-1SG
'[…] I am going to the theater.'

b) *Dün* *saat* *beş-te* *Hasan* *kahve* *iç-iyor-du.*
yesterday o'clock five-LOC[47] Hasan coffee drink-PROG-PST.3
'Yesterday at five o'clock Hasan was drinking coffee.'

c) *Ben-im* *hasta ol-duğ-um-u* *bil-iyor-mu-sun*[48]*?*
I-GEN sick be-FNOM-1SG-ACC know-PROG-Q-2SG
'Do you know that I am sick?'

2.1.2.6.5 Conclusion on peripheral uses

Almost all the grams with peripheral uses belong to the yellow cluster. This is somewhat expected since the yellow cluster includes the grams with the highest number of occurrences in the gram set. The grams *faya* in Merey (Afro-Asiatic), *-sha-* in Eastern Apurímac Quechua (Quechuan), *de* + V-*nù* in Kuwaa (Niger-Congo) and *-Iyor-* in Turkish (Turkic) occurred in several segment sets. Notably, their occurrences in the Bible corpus are relatively high: 3648, 3016, 2422 and 2100, respectively. They can therefore be assumed to be borderline cases that have expanded towards the imperfective, which would be in line with previous research.

As previously illustrated, the comparison with the segment sets representing peripheral uses mainly capture general and more frequent tendencies. Grams with less frequent peripheral uses, or uses with more subtle interpretations, cannot be discussed here. For example, we know that the English progressive has many peripheral uses that are not reflected here. I will discuss some uses of this kind in Chapter 4.

2.1.3 Summarizing section 2.1

In this section, it was shown that progressives typically have a majority of present uses, even in a written source such as the Bible. Some grams have a stronger present preference than others, and there are grams that are restricted to present time reference. A number of grams mainly spoken in Southeast Asia that occurred more often in the past were also noted. This preference was shown to be at least partly linked to their preference for subordination. Some of them were shown to

[46] Orig. go-PRS.PROG-1SG.
[47] Orig. five-ABL.
[48] Orig. know-PRS.PROG-Q-2SG.

have originated from the meaning 'middle', therefore, it was suggested that they may have grammaticalized into meaning 'while V-ing'.

The most favorable contexts for progressives in the present and past were also presented. It was noted that the most favorable contexts with present time reference included FOC ongoing events and utterances with *behold*, which demands the attention of the listener to be turned towards an ongoing event. The most favorable contexts in the past included backgrounding uses. Finally, a number of grams with future, performative-like, habitual and stative uses were also discussed. Not surprisingly, these grams have high frequencies in the corpus. This is compatible with the assumption that progressive grams grammaticalizing towards the imperfective expand their functional domains toward such uses in the process.

2.2 TED corpus

In this section, the distribution of 14 progressive grams in 12 languages in the TED corpus is investigated. The section has as its main aim to complement the observations regarding uses in the present and past made in the Bible corpus. The distribution of some additional grams in comparison to the 14 progressive grams will also be presented. In section 2.2.1, the TED corpus, method and methodological issues, as well as the inclusion of grams in the gram set(s), are presented. Section 2.2.2 presents the results, and section 2.2.3 summarizes the findings.

2.2.1 The corpus, method and methodology

2.2.1.1 The corpus

The TED corpus is compiled by Östen Dahl and consists of non-annotated subtitle texts from approximately 1900 TED talks and TEDx talks. A TED talk is a recorded presentation held, most often in English, in front of a live audience.[49] Originally, the topics of the talks were focused on technology and design, but today all types of topics are included. The talks are typically 20 minutes long and freely accessible.[50] The TED corpus then consists of subtitles in about 100 languages for a selection of the talks. The number of translations for each video varies, the maximum number being 60, which means that the corpus is partially parallel. Consequently, the number of words per language differs. For example, while 3 639 000 words are available in English, 2 579 000 words in Persian and

[49] Presenters almost never use teleprompters: https://www.quora.com/Have-some-presenters-at-TED-used-teleprompters

[50] https://www.ted.com/talks

1 759 000 in Indonesian, there are also languages where only a few thousand words are available. In the corpus, the original line divisions are kept. A line is a stretch of text that can be shown simultaneously on the screen and corresponds to 2-3 seconds of speech. Languages are then linked by these lines. Most often, the languages that are included are so called 'standard' languages with many speakers, and there is an overrepresentation of Indo-European and European languages. In what follows, I will use the term *line* and not *segment*, since the smallest linked unit in the TED corpus is a line.

2.2.1.2 Method, gram set and methodological issues

In section 2.2.2, top lists of occurrences of a number of grams will be analyzed. In order to better understand the data, the parts of the video presentations where the top progressive lines occur were examined. Thus, the analysis is based on the video clips rather than on the written texts in the corpus. Searches for relevant patterns in the corpus were based on information about progressive grams in grammatical descriptions for those languages that have a relatively high number of words in the corpus. The language with the smallest number of lines included in the gram set is Finnish with 67 537 lines.

The results of the searches identified one group of grams with high mutual similarity and another group of grams that were less similar to the first group and also to each other. I refer to the former group as the *primary gram set* and to the latter group as the *secondary gram set*.

Fourteen grams are included in the *primary gram set*. These are given in Table 7. The main part of section 2.2.2 concerns the analysis of the top lists of these grams. Since the grams all belong to well-known languages, I will not provide information on language family when referring to these languages. The members of the *primary gram set* all have a precision value above 60% (I will present and explain how precision is calculated in section 2.2.2). As can be seen, Persian is represented by both the *dāštan* and the *dar hāl=e* progressives. Since the texts in the TED corpus are aimed at representing spoken language in writing, the translators may differ in their preferences of using the colloquial *dāštan* or the more formal *dar hāl=e* progressive for Persian. Also, Portuguese is represented by two patterns, the *estar* + *a* + INF and *estar* + GER progressives. These will be shown to have very different frequencies in the corpus.

Table 7. The progressive grams of the primary gram set

LANGUAGE	LANGUAGE FAMILY	LINES IN CORPUS	PROGRESSIVE GRAM
English	Indo-European	519936	*be* + GER
Finnish	Uralic	67537	*olla* + INF3
French	Indo-European	489130	*être* + *en train de* + INF
Indonesian	Austronesian	251424	*sedang*
Italian	Indo-European	488278	*stare* + GER
Mandarin Chinese	Sino-Tibetan	482266	*zhèngzài*
Persian	Indo-European	368444	*dāštan* + IPFV
			dar hāl=e + INF
Portuguese	Indo-European	286015	*estar* + *a* + INF
			estar + GER
Spanish	Indo-European	470471	*estar* + GER
Swedish	Indo-European	171726	*hålla* + *på* + *att/och*
Thai	Tai-Kadai	244258	*kamlaŋ*
Vietnamese	Austroasiatic	412134	*đang*

The *secondary gram set* consists of grams that are referred to as progressives or similar in grammars, but have a precision value below 50% (see Table 11, section 2.2.2.3). This means that in the TED corpus, they do not share their majority of uses with the grams of the primary gram set. For example, the Indonesian *lagi* marks the progressive but also means 'still, again, more, other' (Sneddon 2006), and the Mandarin Chinese *zài* is a marker of the progressive as well as having locative and other functions (Po-Ching & Rimmington 2004:105). As can be noted, the Turkish *-Iyor-* is included in the secondary gram set, for reasons I will discuss in the aforementioned section. The two German grams are included in the secondary grams set since their frequencies in the corpus are very low, which makes it more difficult to analyze their distributions. The Swedish posture verb constructions formed with *sitta* 'sit', *ligga* 'lie' and *stå* 'stand' are treated as one pattern, since it has been shown that the main difference between these patterns is the position of the body (Kvist Darnell 2008:247). I have also chosen to treat the Spanish motion verb constructions together.

The grams of the secondary gram set are given in Table 8. They will be compared to the grams in the primary gram set and are briefly discussed in section 2.2.2.3.[51]

[51] Some grams that were searched for but not included in either gram set: the Italian *stare* + *a* + INF occurs merely 14 times in the corpus, the Italian *andare* + GER occurs 12 times, the Italian *venire* + GER and the Portuguese *vir*/*ir* + GER did not occur in the corpus, the German *sein* + *beim* + INF only occurred a few times and the German *sein* + *daran* has possibly one occurrence.

Table 8. The grams of the secondary gram set

LANGUAGE	LANGUAGE FAMILY	LINES IN CORPUS	GRAM
German	Indo-European	405807	*sein* + *am* + INF
			sein + *dabei zu*
Indonesian	Austronesian	251424	*lagi*
Mandarin Chinese	Sino-Tibetan	482266	*zài*
Persian	Indo-European	368444	*mašġul* + INF
Spanish	Indo-European	470471	*ir/venir/andar* + GER
Swedish	Indo-European	171726	*sitta/ligga/stå* + *och* + V
Turkish	Turkic	426430	*-mAktA-*[52]
			-Iyor-
Vietnamese	Austroasiatic	412134	*đương*

Similar to what was done for the Bible corpus, progressive patterns in the TED corpus are captured through certain search strings. For example, the search string for the Swedish *hålla* + *på* + *att/och* gram finds all relevant forms of *hålla* followed by the preposition *på* and either *att* or *och*.[53] The search string for the Spanish *estar* + GER gram finds all relevant forms of *estar* followed by a word ending in *-ando* or *-endo*, which in turn can optionally be followed by pronominal clitics, and so on. The results of the search strings of the two grams included for German were manually checked, and false positives were excluded, this was possible due to the low number of occurrences of these grams in the corpus. In some cases, however, it was not possible to entirely exclude false positives. As previously mentioned, in Persian, the verb *dāštan* 'have' in combination with the main verb taking the imperfective *mi-* prefix is a progressive gram. The *dāštan* verb is also a main verb meaning 'have'. The search for *dāštan* + *mi-* sometimes including two verb phrases, one where *dāštan* is used as a main verb and one where a verb taking the *mi-* prefix is the main verb. Attempts were made to avoid such false positives. Since such cases are rare, and since we are concerned with top lists in this investigation, i.e. the lines in which most of the languages use their progressive grams, these cases are not problematic.

In this section, unlike in the Bible corpus, we will only be dealing with one general top list since there is no clustering of grams. The calculation of percentages for the TED corpus is more problematic than for the Bible corpus since we are dealing with a partially parallel corpus. When presenting lines from

[52] *A* in *-mAktA-* represents an archiphoneme.

[53] The occurrences of the pattern with *och* are much less common than the occurrences of the pattern with *att*. As both *att* and *och* are pronounced /ɔ/ in spoken language, the difference between the two patterns is only shown on the following verb form, which is infinitive after *att* and finite after *och*. Although Blensenius (2015) shows that the two patterns can have different uses, he also notes a great deal of overlap in their uses.

the top list, the name of the presenter is followed by the time of the utterance, as given in the corpus. In 2:27 then, 1312 refers to 13 minutes and 12 seconds. This is followed by the number of languages that use their progressive in this line, the number of languages that have a line for this utterance and the ratio of these two values. In this example then, eight languages had this line in the corpus and all used their progressive. As was done with the segments in the Bible corpus, the line containing the progressive, and to which 1312 refers, is given in italics, with some additional lines provided for context.

2:27 Luis_von_Ahn1312 (8, 8, 1.0)

> Now, the crazy thing about this method is that it actually really works. First of all, people are really, really learning a language. *We're mostly done building it, and now we're testing it*. People really can learn a language with it.

In both our corpora we have an issue with translational biases. In the TED corpus, there is a bias towards English since most of the TED talks are given in English. In fact, all the top sentences that were investigated included talks that were held in English. Also, even if the subtitles are based on spoken language, there are still written language conventions that may have affected the translations. In addition, when discussing tense, I will take the English present or past tense to represent present or past time reference. This is, of course, an overgeneralization, but I will assume that the general conclusions drawn still hold.

2.2.2 Data and analysis

As already mentioned, the inclusion of a gram in the primary gram set is dependent on its precision value being higher than 60%. In Table 9, recall and precision values for the grams in the primary gram set are given. The recall value refers to the occurrences of a gram in the top 20 and top 200 lines, depending on whether or not the language has that line in the corpus. For example, for French there are transcripts of 17 lines of the top 20 list, and in all these the progressive gram is used. In the top 200 list, French has 177 lines available, and the progressive is used in 99 of these.

Table 9. Recall and precision for the primary gram set

GRAM	TOP 20			TOP 200			PRECISION		
	OCC.	AVAL. IN	%	OCC.	AVAL. IN	%	OCC.	TOT. OCC.	%
English *be* + GER	20	20	100	199	200	99	14984	24100	62
French *être* + *en train de* + INF	17	17	100	99	177	56	820	898	91
Finnish *olla* + INF3	0	0	-	3	4	74	230	378	61
Indonesian *sedang*	6	6	100	35	42	83	1120	1402	80
Italian *stare* + GER	20	20	100	188	193	97	8207	9380	87
Mandarin Chinese *zhèngzài*	18	18	100	144	190	76	1870	2399	78
Persian *dar hāl=e* + INF	6	9	67	45	90	50	1457	2208	66
Persian *dāštan* + IPFV	3	9	33	32	90	36	1835	2890	63
Persian both patterns	9	9	100	77	90	86	-	-	-
Portuguese *estar* + *a* + INF	7	7	100	86	89	97	6302	7768	81
Portuguese *estar* + GER	0	7	0	2	89	2	234	256	91
Portuguese both patterns	7	7	100	88	89	99	-	-	-
Spanish *estar* + GER	16	16	100	185	185	100	8846	10008	88
Swedish *hålla* + *på* + *att/och*	2	2	100	8	13	62	141	156	90
Thai *kamlang*	6	6	100	57	61	93	3187	4232	75
Vietnamese *đang*	11	11	100	153	155	99	8935	13770	65

The last three columns in Table 9 calculate precision in the following way: a list of all the lines where at least five languages have a line and at least two progressive grams occur were created, and the occurrences of each gram are checked against this list. French, for example, has 898 lines in this list, of which 820 contain the progressive gram. This gives the French gram a precision value of 91%. Since we are dealing with very few grams, and since the corpus is only partially parallel, the recall and precision values have to be taken with a grain of salt, since smaller changes in the gram set (such as the addition of a gram to the set) could change the figures more drastically than it would have in the Bible corpus. Therefore, I will not discuss the difference in precision value between the grams in the primary gram set, rather, I have treated these figures roughly for the inclusion and exclusion of grams in the gram set.

For Persian, an additional line has been added in Table 9 where the recall values of both *dāštan* + IPFV and *dar hāl=e* + INF patterns are added together. Once we put them together, the recall values rise. This means that in typical progressive contexts either one or the other will be used. The same has been done in Portuguese, but here the *estar* + *a* + INF pattern is much more frequent than the *estar* + GER pattern. As can be seen, in both top lists the *estar* + *a* + INF pattern is used almost exclusively. However, *estar* + GER has a high precision value, which means that when it is used, it is used in contexts where other progressives are also used. This shows that the percentage is a bit tricky for two reasons: first, it depends highly on the language having a translation for that line, and secondly, it depends on other patterns that may also be suited in those lines, thus occupying the space as it were

In Table 10, the languages and progressive grams of the primary gram set are given in the first two columns, followed by the total number of occurrences of the gram in the corpus. The last column calculates the ratio between the number of occurrences of the progressive gram and the total number of words in the corpus. As can be seen, the English progressive is the most frequently used gram, with 6.6 occurrences per 1000 words, whereas the lowest frequency is 0.1 occurrences per 1000 words.

For each gram in the primary gram set, it was calculated how many lines in the corpus it shared with the English progressive, and what percentage among these were in the present in English. The average present percentage for the grams in the sample was 80. No gram occurred as many times or more often in the past. This is somewhat lower than the present tense preference in direct speech in the Bible corpus.

Table 10. Frequencies of grams in the primary gram set in the corpus

LANGUAGE	PROGRESSIVE GRAM	TOTAL OCC. OF GRAM	FREQUENCY OF GRAM PER 1000 WORDS
English	*be* + GER	24100	6.6
Finnish	*olla* + NMLZ	378	0.8
French	*être* + *en train de* + INF	898	0.3
Indonesian	*sedang*	1402	0.8
Italian	*stare* + GER	9380	2.7
Mandarin Chinese	*zhèngzài*	2399	0.7
Persian	*dar hāl=e* + INF	2208	0.9
	dāštan + IPFV	2890	1.1
	Total	5098	2.0
Portuguese	*estar* + *a* + INF	7768	3.9
	estar + GER	256	0.1
	Total	8024	4.0
Spanish	*estar* + GER	10008	3.0
Swedish	*håller* + *på* + *att/och*	156	0.1
Thai	*kamlang*	4232	2.5
Vietnamese	*đang*	13770	4.8

We will now turn to the analysis of the present and past tense lines of the top segments of the members of the primary gram set.

2.2.2.1 Present time reference

Most of the top lines in the English present tense have a FOC interpretation and involve utterances that have to do with change in which the speaker presents a new project or new work or a change in the world that is currently taking place, i.e. telic events. In fact, the verbs *change* and *happen* very frequently occur in the top lines. This is not surprising since the TED talks often involve presentations of new ideas and projects that are promoted as important, given current developments. As was noted, the top lines in the Bible corpus often involved *praying* and *speaking*, i.e. atelic events.

2:28 FOC present contexts

a) Paul_Root_Wolpe0413 (9, 9, 1.0)
But something much, much more powerful *is happening now*. These are normal mammalian cells genetically engineered with a bioluminescent gene taken out of deep-sea jellyfish.

b) Jan_Chipchase1400 (7, 7, 1.0)
And I know TED is about big ideas, *but actually, the benchmark for a big idea is changing*. If you want a big idea, you need to embrace everyone on the planet, that's the first thing.

c) Frederic_Kaplan0328 (8, 8, 1.0)
We are setting up a 10-year digitization program which has the objective of transforming this immense archive into a giant information system.

d) Don_Tapscott0015 (8, 8, 1.0)
Openness. It's a word that denotes opportunity and possibilities. Open-ended, open hearth, open source, open door policy, open bar. (Laughter) *And everywhere the world is opening up*, and it's a good thing.

e) Eleni_Gabre-Madhin1443 (6, 6, 1.0)
So, the ECX is an Ethiopian exchange for Ethiopia. *We're creating a system that serves all market actors*, that creates integrity, trust, efficiency, transparency and enables small farmers to manage the risks that I have described.

f) Anne-Marie_Slaughter1638 (5, 5, 1.0)
The revolution for human equality can happen. *It is happening*. It will happen.

The segments translated to the English present tense in the Bible corpus often contained *behold*, which was used as a requirement that the attention of the listener be turned towards an ongoing event. Such uses are not common in the TED corpus, only one such line was found: in example 2:29, the speaker uses the present progressive in a relative clause when referring to wasps metamorphosing inside cocoons. In doing so, he briefly turns and points towards the screen next to him where the imagery is changed in that instance to show a picture of these cocoons. The progressive is given together with the turning and pointing, thus as part of a strategy to turn the attention of the audience towards the depicted metamorphosis on the prompter.

2:29 Ed_Young0426 (8, 8, 1.0)

Some of the wasps seemed to stay behind and controlled it into defending their siblings *which are metamorphosing* into adults within those cocoons.

We can assume that attention-requiring uses are generally rare in the TED corpus since in a presentation, the presenter typically already has the attention of the audience, and there are no sudden or unexpected events that are taking place

around him/her in which an attention-requiring phrase could be used. The showing of events on the prompter is controlled by the speaker, and consequently, when there is a wish for the audience to look at those events, other strategies are used, e.g. *And the thing I want you to look at here*[54] and *Let's look at the squares*.[55]

In Chapters 3 and 4, subjective readings and temporary uses are discussed. Such uses were not found in the Bible corpus and do not seem to be very frequent in the TED corpus, either. It is noted, however, that the word *actually* recurs in the top list. Such utterances can perhaps be argued to mark surprise, as in 2:28b). Progressive utterances carrying the notion of surprise have previously been discussed for English, French and German progressive grams (De Wit & Brisard 2014; De Wit et al. 2013; Anthonissen et al. 2016). Two more examples are given in 2:30.

2:30 Lines with 'actually'

a) Janna_Levin1319 (9, 9, 1.0)
In this Hubble image, we see two galaxies. They look like they're frozen in some embrace. And each one probably harbors a super-massive black hole at its core. But they're not frozen; *they're actually merging*.

b) Eleni_Gabre-Madhin 1328 (6, 6, 1.0)
Now, over the last century, we tend to think of commodity exchanges as the purview of Western industrialized countries, and that the reference prices for cotton, coffee, cocoa -- products produced mainly in the south -- are actually a reference price, or a price discovered in these organized commodity exchanges in the northern countries. *But that is actually changing*.

The utterance in 2:31 is a narrative, but the tense used in English is present tense. Interestingly, the progressive is given as we switch tense. Here, the event marked by the progressive refers to several utterances given after one another. This example can be viewed as an iterative ongoing use.

2:31 Talithia_Williams1306 (5, 6, 0.83)

But okay, what was Donald doing just before? *So Donald goes into this slew of medications he was taking*. He lists, "I took this decongestant and then I took this nasal spray,"

[54] Anil_Ananthaswamy1255.
[55] Arthur_Benjamin0204.

2.2.2.2 Past time reference

Similar to the past uses in the Bible corpus, the top lines corresponding to the English past tense in the TED corpus often refer to backgrounding events where another event interrupts or occurs at the same time as the event to which the progressive applies. Some examples are given in 2:32.

2:32 Backgrounding past contexts

a) Carl_Safina1540 (5, 5, 1.0)
At an aquarium in South Africa was a little baby bottle-nosed dolphin named Dolly. *She was nursing, and one day a keeper took a cigarette break* and he was looking into the window into their pool, smoking.

b) Sean_Gourley0034 (8, 9 0.89)
So as a naive New Zealander I thought, well I'll go to the Pentagon. Can you get me some information? (Laughter) No. So I had to think a little harder. *And I was watching the news one night in Oxford.* And I looked down at the chattering heads on my channel of choice. And I saw that there was information there.

c) Jennifer_Granholm0727 (7 8 0.88)
…and I was standing in the back of the room during one of the demonstrations and standing next to one of the Chinese officials, *and we were watching, and he says,* "So, Gov, when do you think the U.S. is going to get national energy policy?"

d) Michael_Pollan0055 (6, 7, 0.86)
Like a lot of my ideas, like a lot of the tools I use, I found it in the garden; I'm a very devoted gardener. *And there was a day about seven years ago: I was planting potatoes*, it was the first week of May – this is New England, when the apple trees are just vibrating with bloom; they're just white clouds above. […] And the question I asked myself that afternoon in the garden

e) Jehane_Noujaim0636 (6, 7, 0.86)
because before the war started, there was kind of this media war that was going on. *And I was watching television in New York*, and there seemed to be just one point of view that was coming across,

f) Maz_Jobrani0556 (8, 10, 0.80)
They have a mall there, the Dubai Mall. It is so big, they have taxis in the mall. *I was walking. I heard "Beep, beep."* I'm like, "What are you doing here?" He goes, "I'm going to the Zara store. It's three miles away.

In many backgrounding events, the FOC point is provided by another telic event, as, for example, in 2:32b) by "I looked down", in c) by "he says", in d) by "the question I asked myself", and in f) by "I heard 'Beep, beep'". In examples 2:32a) and e), it is unclear if "one day" and "[one point of view] coming across" are examples of FOC or DUR reference points.

There are also examples where the backgrounding context occurs with what could be viewed as a DUR reference time. In 2:33a), the presenter is talking about a robot that is measuring the prosody of the speaker's voice during the time that the speaker is working on the robot. In 2:33b), the progressive applies to a process that took place over a period of time when several other events were happening. In this example, however, 'meanwhile' could perhaps be analyzed as adding a FOC reference time too. Interestingly, the stretched out duration of the reference time is explicitly given in these examples.

2:33 DUR backgrounding contexts in the past?

a) Rodney_Brooks1151 (5, 5, 1.0)
And when I was dealing with my robot over here, *Chris, the robot, was measuring the prosody in my voice,* and so we have the robot measure prosody for four basic messages

b) Peter_Hirshberg1838 (5, 6, 0.83)
By this time, computing had kind of leapt into media territory, and in short order much of what we're doing today was imagined in Cambridge and Silicon Valley. Here's the Architecture Machine Group, the predecessor of the Media Lab, in 1981. *Meanwhile, in California, we were trying to commercialize a lot of this stuff.*

FOC past contexts that are non-backgrounding and more equivalent to the typical FOC present ongoing uses are also found in the data, some examples are given in 2:34. In these examples, the reference time can be seen as referring to the same moment as the event to which the progressive applies.

2:34 Non-backgrounding past contexts

a) Chris_Domas0513 (5, 6, 0.83)
In this case, I was looking for a very advanced, very high-tech piece of code that I knew I could hack, but it was somewhere buried inside of a billion ones and zeroes. Unfortunately for me, I didn't know *quite what I was looking for*. I didn't know quite what it would look like, which makes finding it really, really hard.

b) Ian_Dunbar0847 (5, 6, 0.83)
The dog broke his down-stay 22 times in four and a half hours, while she cooked dinner, because we had a lot of aggression related towards food. The breaks got fewer and fewer. *You see, the punishment was working*. The behavior problem was going away. She never raised her voice. If she did, she would have got bitten.

c) Carl_Safina1313 (4, 5, 0.80)
Sisters kicked out other sisters. That one on the left tried for days to rejoin her family. They wouldn't let her because they were jealous of her. *She was getting too much attention from two new males*, and she was the precocious one. That was too much for them.

d) Eve_Ensler0639 (8, 10, 0.80)
Cancer exploded the wall of my disconnection. I suddenly understood that the crisis in my body was the crisis in the world, and it wasn't happening later, *it was happening now*.

Similar to FOC ongoing uses in the present, the examples refer to new and topical events. The example in 2:34d) has an urgent and dramatic sense. These uses were not very common in the data. The overall impression is that backgrounding uses are the most typical uses in the past, and that the sense of focality as well as the emotive effect is generally less prominent when progressives apply to past time reference than to present.

Another interesting example is given in 2:35. The utterance is given with a dramatic gesture after the ironic *I'm bringing in the end of the world. You know?* The line contains two clauses, both of which include the progressive.[56] The first is a past progressive in the English translation and is given as a background to a second present progressive that has a more urgent sense and is uttered ironically.

2:35 Paola_Antonelli0351 (5, 6, 0.83)

I'm bringing in the end of the world. You know? *We were talking about the rapture? It's coming*. And Jonathan Jones is making it happen.

2.2.2.3 Secondary gram set

Table 11 presents the frequencies of grams of the secondary gram set as well as their recall and precision values. For recall and precision, the grams of the secondary gram set are run against the **primary gram set** one by one. This is done in order to see whether these grams have a similar distribution as the progressive

[56] This means that we do not actually know in which clause the grams occur, and if they occur in both, as in the English translation.

grams in the primary gram set. The members of the secondary gram set are not tested against each other as was done for the grams of the primary gram. Attempts at creating a top list of the secondary gram set failed mainly because it was not possible to find a set of lines with translations into a sufficient number of languages and a sufficient number of occurrences of the members of the secondary gram set.

For example, the Mandarin Chinese *zài* is checked against the top 20, the top 200 and the precision list of the primary gram set. We then see that *zài* has quite low occurrences in the top 20 and 200. It also has a low precision value. As expected then, it has other uses than ongoingness which gives it a different distribution than the members of the primary gram set.

The two German grams are included here, since it is interesting to note their relatively low frequencies in the corpus. Due to this, however, I will not discuss them further. Regarding the other grams, a comparison between Table 9 and Table 11 shows that the grams in the secondary gram set have much lower recall and precision values than the grams in the primary gram set.

The data shows that most of the uses of these grams are not shared with the grams in the primary gram set. But the number of shared lines with the primary gram set differs between the members of the secondary gram set. The Persian *mašġul* + INF pattern has some overlap with the grams in the primary gram set. In Chapter 3, it will be confirmed that it partly overlaps with the progressive *dāštan* pattern.

The Spanish motion verb patterns "only carry a durative meaning" (Bertinetto 2000:577). This could explain their low precision value.[57] Similarly, the Swedish posture verbs are also suggested in Bertinetto et al. (2000a:530) to have a natural durative meaning, however, I am not certain in this case that this is the correct analysis. Although they may easily combine with durative adverbials, an utterance such as *Jag sitter och läser* 'I am sitting and reading' has in my mind a FOC reading, given, for example, as an answer to *What are you doing?* The difference in distribution of both the motion verb and the posture verb patterns in comparison with the patterns in the primary gram set suggests that they are perhaps better described as something other than progressives. It may also be that they are quite marginal cross-linguistically.

[57] Bertinetto (2000:580) notes that the *andar* + GER progressive is more interchangeable with the *estar* + GER pattern. However, even run independently, none of the motion verb constructions received a high precision value.

Table 11. Frequency, recall and precision for the members of the secondary gram set run against the primary gram set

GRAM	FREQUENCY OF GRAM PER 1000 WORDS	TOP 20			TOP 200			PRECISION		
		OCC.	AVAL. IN	%	OCC.	AVAL. IN	%	OCC.	TOT. OCC.	%
German *sein* + *am* + INF	0	0	14	0	1	149	1	12	27	44
German *sein* + *dabei zu*	0	0	14	0	0	149	0	19	27	70
Indonesian *lagi*	1.5	0	6	0	0	42	0	85	2589	3
Mandarin Chinese *zài*	17.2	1	18	6	30	190	16	4641	58087	8
Persian *mašġul* + INF	0.2	0	9	0	1	90	1	153	453	34
Spanish *ir/venir/andar* + GER	0.1	0	16	0	0	185	0	49	335	15
Swedish *sitta/ligga/stå* + *och* + V	0.1	0	2	0	0	13	0	33	135	24
Turkish *-mAktA-*	1.2	0	15	0	3	156	2	362	3488	10
Turkish *-Iyor-*	21.1	9	15	60	91	156	58	6821	62897	11
Vietnamese *đương*	0.1	0	11	0	0	155	0	6	260	2

The Turkish *-mAktA-* consists of the infinitive marker *-mAK* followed by the locative case *-DA*. Both *-mAktA-* and *-Iyor-* are used for ongoing events but may also occur in habitual contexts (Göksel & Kerslake 2005:332–333). *-mAktA-* has a formal connotation, as opposed to the marker *-Iyor-* which is more neutral. Kornfilt (1997:358) discusses both as having ongoing uses but notes that neither *-Iyor-* nor *-mAktA-* are limited to such uses. Both the Turkish *-mAktA-* and *-Iyor-* markers have a low precision value in Table 11. However, *-Iyor-* has a higher recall value. This means that, as expected, it occurs to an extent in contexts that are most favorable for progressive grams, but that it also has other uses. Due to the observations mentioned in previous literature and the results from this chapter, both *-Iyor-* and *-mAktA-* will be regarded as patterns with uses towards the imperfective in this thesis.

Notably, *-Iyor-* has a much better precision value in the Bible corpus than in the TED corpus. Comparing the incidence of the marker to the English progressive in both corpora, we see that it occurs much more frequently in the TED corpus. In the TED corpus *-Iyor-* has a frequency of 21,1 per 1000 words while the English gram has 6,6. In the Bible corpus *-Iyor-* occurs 2100 times and the English gram 1252 times. I will then assume that the uses of *-Iyor-* in the Bible corpus are much more restricted than in the TED corpus. The difference of *-Iyor-* in our two corpora is probably, at least partly, explainable by the difference between a conservative and written source and a source closer to spoken language.

We may also note that the Vietnamese *đương* has a much lower precision value in the TED corpus than in the Bible corpus. Phan (2013:25, 44, 65) refers to *đang/đương* as durative and "related to viewpoint aspect", but also shows that *đương* has other uses, such as occurring with statives and meaning 'road'. While the distribution of *đương* indicates that it is a progressive marker in the Bible corpus, it cannot be regarded as such in the TED corpus. This probably has to do with the preferences in the Bible translation in which *đương* is found. The Bible translation in the Bible corpus is from 1934 and therefore may perhaps contain a different use than the more modern TED transcriptions.

Finally, I wish to stress that it is possible that some of the grams of the secondary gram set would have had better recall and precision values against a different, preferably larger, and areally and genealogically more diverse, gram set. What this section has shown is the extent to which these grams share their distribution with the grams of the primary gram set.

2.2.3 Summarizing section 2.2

The data from the TED corpus confirms that the grams included in the primary gram set occur more often with present time reference than with past. This is expected since we are dealing with translations of spoken language. Also, the most favorable contexts for progressives with present time reference are FOC

ongoing uses. Due to the nature of data, not many attention-requiring contexts were found. Instead, progressives typically applied to telic events referring to changes in the world. In the past, similar to the findings in the Bible corpus, backgrounding uses were typical, but FOC ongoing uses were also found. In the past, both FOC and DUR reference points were noted. The impression is that focality is less prominent in the past.

A number of grams that have been referred to as progressive or similar in the literature were also presented in the secondary gram set. These were shown not to share their distribution with the grams of the primary gram set to a large extent. I will leave for future research to investigate their uses in more detail.

3 The *dāštan* progressive in Persian

3.1 Introduction

This chapter provides a descriptive account of the most favorable as well as peripheral uses of the periphrastic progressive construction formed with the verb *dāštan* 'have' in colloquial Persian. The grammatical features of the Persian *dāštan* construction are commonly noted in the literature (e.g. Dehghan 1972:200–201; Lazard 1992:160; Mahootian 1997:241–244; Mace 2003:102–103): it occurs only in the indicative and imperfective; it cannot be negated; it is incompatible with the verb *budan* 'to be', the verb *dāštan* 'to have' and stative verbs such as *dānestan* 'to know' as the main verb; and it may occur in passive constructions (formed with the verb *šodan* 'become'). *Dāštan* always precedes the main verb in a clause and items such as objects and prepositional phrases may intervene between the two. In addition, the *dāštan* progressive is often described as mainly restricted to colloquial speech (Windfuhr & Perry 2009:461), in written Persian, the more formal *dar hāl=e* 'in state of' progressive is available.

The *dāštan* construction is not only used in Persian, it has also been copied into many neighboring languages and dialects, mainly within Iran. Some examples are Gazi (Stilo 2007a), the Jewish dialects of Isfahān (Stilo 2007b) and Hamadān (Stilo 2003), Luri (data from PROGQ), Sistani (Ahangar 2010), Sarhaddi Balochi of Granchin (Ahangar 2007), Yazdi (Dehghan 1972), Kermāni (Dehghan 1972) and Āvarzamāni (Dehghan 1972). The situation constitutes a typical example of borrowing from a dominant language into dominated varieties.

The data used in this chapter has been collected via a questionnaire filled out by 26 native speakers of Persian and via the collection of utterances containing the pattern in five modern Iranian movies. Although grammatical features of the pattern are mentioned in previous literature, little has been done regarding its functional scope. The chapter focuses on the uses of the Persian *dāštan* progressive, adding some points to the discussion of its grammatical features. Apart from FOC ongoing meaning, which will be shown to be the most common use of the pattern, uses such as proximative, futurate, iterative are also presented, and intensifying and subjective readings are discussed. This chapter will also present the *dar hāl=e* and *mašġul* + INF patterns, and to lesser extent the *gir* construction, given in the responses to the questionnaire in section 3.2.2.2.

The questionnaire used here is an adjusted and translated version of the progressive questionnaire used in Bertinetto et al. (2000b). The adjusted

questionnaire will henceforth be referred to as PROGQ. It aims to investigate the different uses of the *dāštan* construction and also enable a comparison with those studies in the EUROTYP project concerning the progressive in a number of European languages (i.e. Bertinetto et al. 2000a; Bertinetto 2000; Ebert 2000; Tommola 2000; de Groot 2000). The information collected from the Iranian movies enables a comparison to more natural-like speech with pragmatic and contextual data.

In this chapter, if no other information is given, the examples are standard Persian, as spoken in Iran. Unless another reference is given, all Persian examples are glossed by me. All interpretations based on intuition are checked with native speakers for acceptability. Throughout this chapter, I will refer to these consulting informants when native speaker acceptability is required. All examples lacking a reference to the PROGQ or to an author were constructed for the purpose of illustration.

This section continues with an overview of the morphological features of the *dāštan* construction in 3.1.1 and the relation of this pattern to other auxiliary patterns in Persian in 3.1.2. This is followed by the presentation of data from the PROGQ in section 3.2 and continues by presenting the occurrences of the *dāštan* construction in the Iranian movies in section 3.3. In section 3.4 the results of both the PROGQ and the movies are discussed with focus on the different uses. Section 3.5 concludes this chapter.

3.1.1 Morphology and syntax of the *dāštan* progressive

The *dāštan* construction is formed with the verb *dāštan* 'have', which is semantically completely bleached, followed by the main verb, which takes the imperfective *mi-* prefix. 'Main verb' refers to the verb that contains the semantic content of the verb phrase. In the *dāštan* construction, both verbs are inflected for tense, person and number, as shown in examples 3:1a) and b). The paradigm of this verb is composed of the present stem *dār-,* historically 'hold, keep, dwell' and the past stem *dāšt-*, which has been imported from the verb **darz* 'to attach, fasten; to load; to sew' (Cheung 2007:57, 59, 62).

As shown, only the main verb takes the *mi-* prefix which is expected in view of the fact that *mi-* generally does not combine with *dāštan*. Thus, tense and person-number marking is redundantly shown on both the *dāštan* verb and the main verb. When used as a main verb, *dāštan* retains its meaning 'have', as seen in 3:2.

3:1 The *dāštan* progressive

a) *dār-am* *dars* *mi-xun-am.*
have.PRS-1SG lesson IPFV-read.PRS-1SG
'I am studying.'

b) *dāšt-am* *dars* *mi-xund-am.*
have.PST-1SG lesson IPFV-read.PST-1SG
'I was studying.'

3:2 *Dāštan* as main verb

a) *se=tā* *so'āl* *dār-am.*
three=CL question have.PRS-1SG
'I have three questions.'

b) *se=tā* *so'āl* *dāšt-am.*
three=CL question have.PST-1SG
'I had three questions.'

The present progressive and past progressive patterns are built on the present indicative tense and past imperfective, respectively. This means that the removal of the *dāštan* verb from the examples in 3:1 leaves us with the present indicative tense and past imperfective, respectively.

Apart from the present and past forms of the *dāštan* construction, other forms are also mentioned in the literature. Dehghan (1972:200) mentions the progressive pluperfect *dāšte bud mirafte bud* 'he had been [in the process of] going' but notes that this form is extremely rare. Lazard (1992:160) and Windfuhr (1979:102) note a pattern referred to as 'completed past' and 'reported', respectively, as in, e.g., *dāšte bāzi mi-karde* 'was playing', where *dāšte* is a perfect form of *dāštan*. Lazard (1989:273) also notes the 'inferential progressive' *dāšte mi-rafte-ast*. None of these forms occur in the data and will therefore not be discussed.

The marking of aspect in Persian is a complex matter intertwined with both tense and mood, and will not be addressed in detail here (for an overview, the reader is referred to e.g. Windfuhr 1979:83–113). Table 12 illustrates the modal and aspectual asymmetry of the imperfective *mi-* prefix relevant for the forming of the *dāštan* construction. The present and past tenses are formed with present and past stems, respectively. The present tense is often referred to as the present/non-past, since it is also used with future time reference, in what follows, for the sake of simplicity, I will refer to the present/non-past tense as the present tense. When combined with present stems, the *mi-* marker can be viewed as a marker of the indicative opposed to the subjunctive, whereas when combined with past stems, the *mi-* prefix marks the imperfective and is opposed to the perfective

form. For present stems in the indicative, except for the verbs *dāštan* 'have' and *budan* 'to be', no corresponding prefixless verb forms exist, and for past stems no corresponding forms with *be-* exist. Since present tense can typically be said to be imperfective, the present indicative pattern is also a present imperfective indicative. In the literature, the *mi-* prefix is often referred to as the imperfective marker (e.g. Windfuhr & Perry 2009:451), consequently it will be glossed as such in this thesis.

Table 12. Temporal, aspectual and modal oppositions of the *mi-* prefix

		INDICATIVE	SUBJUNCTIVE
PRESENT STEMS		*mi-*V.PRS-PN	*be-*V.PRS-PN
	PERFECTIVE	**IMPERFECTIVE**	
PAST STEMS	V.PST-PN	*mi-*V.PST-PN	

Historically, the present tense form was prefixless. In modern Persian, the imperfective prefix *mi-*, originating from the Middle Persian adverb *hamē* 'always' (Nyberg 1974:91) or 'forever' (Skjærvø 2009:239), is obligatory in present tense verb forms in the indicative. The verb *dāštan* 'have', however, has preserved the old present tense form and does not combine with the *mi-* prefix (Lazard 1963, cited in Haspelmath 1998:43-44). In certain compound constructions, however, *dāštan* combines with the *mi-* prefix.[58] The verb *budan* 'to be' is also not combinable with the *mi-* prefix unless we consider the stem variant *bāš-*, which is only used in formal language (Mace 2003:98).

It can be added that listeners are often able to distinguish between *dāštan* as part of the *dāštan* progressive or as the main verb early in the processing of such utterances, since typically the *dāštan* element occurs after the subject noun phrase (if there is one) when occurring in the *dāštan* construction but after the object when occurring as main verb. The *dāštan* verb is also typically less stressed when occurring in the *dāštan* construction as when occurring as main verb. This means that in 3:1, independent of any contextual information, once the *dāštan* verb is uttered, the listener knows that it is most likely part of a *dāštan* construction.

In order to investigate the meaning that progressive grams add to the events to which they apply, studies often contrast utterances containing the progressive to utterances containing simple forms (see, e.g., Johanson 2000; De Wit et al. 2013; De Wit & Brisard 2014; Anthonissen et al. 2016). This will also be done in this chapter where the progressive is contrasted to the simple present and past imperfective. However, such comparisons are not a trivial matter. In answering the PROGQ, informants did not always use the simple present or past imperfective as an alternative to the *dāštan* construction, at times other forms such as, e.g., the

[58] For example, *bar mi-dār-am* 'I am removing/taking'.

simple past or the perfect were also used. However, the use of the simple past or perfect changes the interpretation of the sentence in a more radical way than the use of the simple present or past imperfective. The inclusion of the ongoing use with present and past imperfective grams then legitimizes such a comparison.

3.1.2 The *dāštan* progressive and other periphrastic patterns

The Persian *dāštan* progressive is often described as including an auxiliary and a main verb (see, e.g., Lambton 1957; Dehghan 1972; Jeremiás 1993; Mace 2003; Windfuhr & Perry 2009; Davari & Naghzguy-Kohan 2017). The periphrastic patterns discussed here all include elements that have been referred to as auxiliaries in the literature. Here, I will refer to these elements as TAM elements. However, to my knowledge, there is no satisfactory definition of the notion of auxiliary in Persian. In fact, Windfuhr (1979:100), in discussing some of the patterns in Table 13, points out that even if there may be certain similarities between constructions, "each of the aspectual and modal verbs has its peculiar semantic-syntactic rules which criss-cross the traditional and modern classificatory list". In this section, the structures of different periphrastic elements in Persian are presented, showing that there is no general constructional schema for periphrastic patterns in Persian.

Table 13 shows the main periphrastic patterns that are noted in the literature in a simplified way. It does not include all possible tense-aspect forms for the constructions mentioned, nor does it include all periphrastic patterns or possible two-verb constructions in Persian, which is complicated in and of itself given the complex nature of light verb constructions in Persian. Also, there are many more features than tense, person/number marking and negation to be discussed for these verbs, here, only the forms and constructions relevant for the *dāštan* construction are included. In the table, the TAM elements of the constructional schemas are given in bold.

Several observations can be made here. First, there are three patterns that combine a finite TAM element with a non-finite main verb, namely the future formed with *xāstan* 'want', the impersonal construction formed with *šodan* 'become' and the passive formed with *šodan* 'become'. In these patterns, the TAM elements are semantically bleached. These could then be seen as auxiliary patterns due to these features. However, it is noted that the passive differs radically from the other two in that it follows the main verb and in that the main verb has a participle form. In fact, all other patterns in the table precede the main verb. The impersonal construction with *bāyad/bāyest* 'is/was necessary' is similar in structure to the impersonal pattern with *šodan* 'become', but it does not take person number.

Table 13. Periphrastic patterns in present and past tense, and marking of negation

TAM ELEMENT	EX. OF PERIPHRASTIC PATTERN	PRESENT TENSE CONSTRUCTIONAL SCHEMA		PAST TENSE CONSTRUCTIONAL SCHEMA		NEG	FUNCTION
dāštan 'have'	*dār-am mi-xun-am.* 'I am singing.'	**V.PRS-PN**	*mi*-V-PRS-PN	**V.PST-PN**	*mi*-V.PST-PN	-	PROG
bāyad/bāyest 'is/was necessary'	*bāyad be-xun-am* 'I must sing.'	**not conjugated for PN**	*be*-V.PRS-PN	**not conjugated for PN**	*mi*-V.PST-PN	on TAM element	*modal*
šāyad 'is possible'	*šāyad be-xun-am.* 'I might sing.'	**not conjugated**	*be*-V.PRS-PN	**not conjugated**	V.PST-PN or *mi*-V.PST-PN	on main verb	*modal*
tavānestan 'be able to'	*mi-tun-am be-xun-am.* 'I can sing.'	***mi*-V.PRS-PN**	*be*-V-PRS-PN	***mi*-V.PST-PN**	*be*-V.PRS-PN	on TAM element	*modal*
xāstan 'want'	*mi-xā-m be-xun-am.* 'I want to sing.'	***mi*-V.PRS-PN**	*be*-V.PRS-PN	***mi*-V.PST-PN**	*be*-V.PRS-PN	on TAM element	*modal*
xāstan 'want'	*xāh-am xund.* 'I will sing.'	**V.PRS-PN**	V.PST/SHORT.INF[59]	-	-	on TAM element	FUT
bāyad/bāyest 'is/was necessary'	*bāyad hamiše rāst goft.* 'One must always tell the truth.'	**not conjugated for PN**	V.PST/SHORT.INF	**not conjugated for PN**	V.PST/SHORT.INF	on TAM element	IMPERS
šodan 'become'	*mi-š-e intour goft?* 'Can you say this (thus)?'	***mi*-V.PRS-PN**	V.PST/SHORT.INF	***mi*-V.PST-PN**	V.PST/SHORT.INF	on TAM element	IMPERS
šodan 'become'	*xund-e mi-šav-ad.* 'It is read.'	V.PTCP	***mi*-V.PRS-PN**	V.PTCP	**V.PST-PN**	on TAM element	PASS
gereftan 'take'	*mi-gir-e mi-xāb-e* '(S)he falls asleep.'	***mi*-V.PRS-PN**	*mi*-V.PRS-PN	**V.PST-PN**	V.PST-PN	on TAM element	TAM

[59] The past stem standing alone may also be referred to as the short infinitive.

Second, looking at person-number marking we see that all elements except *bāyad/bāyest* 'is/was necessary' and *šāyad* 'is possible' show person-number marking in redundancy with the main verb, *šāyad* is also not conjugated for tense. The past tense form *bāyest* is not frequent in spoken language. This raises the question of whether one should regard *bāyad/bāyest* and *šāyad* as verbs at all since they do not show verb-like properties. Historically *bāyestan* and *šāyestan* were proper verbs, and *bāyad* still has some remaining verbal morphology in constructions like *(mi)-bāyest-(i)* 'you should have done', whereas *šāyad* does not. The negation pattern shows that *bāyad/bāyest* are able to take the negation marker *na-*, while *šāyad* is not.

Third, we can note that only *tavānestan* 'be able to' and *xāstan* 'want' (as a modal), show the same constructional schema in this table. In these cases, it seems as if information regarding tense is retained from the potential auxiliary while the main verb is marked for mood: while the TAM element shows the clausal tense, the main verb shows what can be analyzed as a default tense, namely the present; in reverse, while the main verb shows subjunctive mood, the TAM element shows what can be said to be the default mood, namely the indicative. Both tense and mood are obligatorily marked in these constructions, that is, these verbs must have either a present or a past tense stem and fill the prefixal slot for mood. Interestingly, the *xāstan* pattern can also have proximative use, as noted by Lambton (1957:54) who gives the example *mixāst bemire* '(S)he was about to die'. It is also discussed in Jahani (2017:264), who provides the example *mixād bārun biyād* 'It is going to rain' referring to an "imminent event".

Fourth, the pattern with *gereftan* 'take', e.g. *gereft-and xābid-and* 'they took to sleeping/fell asleep' is a colloquial construction that can be used pejoratively or ironically (Windfuhr 1979:104). Windfuhr provides examples in the imperative and past tense, but according to a consulting native speaker this pattern also exists in the present. It does not seem to have grammaticalized by analogy to the other patterns mentioned here. Similar to the *dāštan* progressive however, in the present tense, it also combines with the indicative form of the verb, i.e. *mi*-taking form.

Fifth, and most importantly for the purposes of this investigation, the *dāštan* construction shows a unique constructional schema not shared by any other construction. In both the present and the past, the construction is only compatible with the indicative/imperfective *mi-* marker, which gives it a unique representation. The reason why the *dāštan* element itself is not combinable with the *mi*-marker has to do with properties of the *dāštan* verb itself, as *dāštan* as a main verb cannot combine with the *mi-* marker. However, it is not clear why the main verb shows past tense in the past: it could have easily followed the pattern of the *tavānestan* 'be able to' and *xāstan* 'want' (modal) constructions, where the main verb is in the present tense. Additionally, similar to the future, passive and impersonal with *šodan*, but not the other patterns, the *dāštan* verb in the *dāštan* progressive construction is semantically bleached.

This means that the creation of the Persian *dāštan* pattern cannot have happened in analogy to other periphrastic patterns. As Table 13 shows, Persian does not have a typical auxiliary schema as the other periphrastic patterns are likewise more or less unique patterns.

3.2 The Progressive Questionnaire (PROGQ)

3.2.1 Method and methodological issues

The modified version of the PROGQ used here consists of 69 Persian sentences. The original questionnaire was created by Bertinetto et al. (2000b), is in English, and was designed to define and capture the main characteristics of a potential progressive construction in a language. Similar to the questionnaire used in Dahl (1985), it contains sentences, sometimes accompanied by additional context given in square brackets, where the predicate is presented in capital letters in the infinitive, leaving the informant to fill out the appropriate form. In order to enable an examination of the full range of the progressive, in addition to typical contexts where a progressive gram is expected to occur, both the original questionnaire and PROGQ also include contexts that are not FOC, sentences with stative verbs, sentences with achievement verbs, negated sentences, habitual contexts and so on. In the original version of the questionnaire, a division of questionnaire sentences into various headings such as 'motion verbs', 'phasal verbs', 'durative adverbials' and so on was made. The sentences in the PROGQ are in a randomized order, which is why these headings were not kept.

The original questionnaire was translated from English into Persian, was shortened and was again expanded by a few additional sentences, randomized and adjusted to better fit the cultural and linguistic features of the region. Appendix B presents the PROGQ and Appendix C, an English translation of the PROGQ (note that this translation does not exactly correspond to the original English questionnaire). An example of questionnaire sentence 55 is given in 3:3. In 3:3a), the sentence as it appears in the PROGQ is given, in 3:3b), an answer is given. Examples from the PROGQ include Q followed by the number of the questionnaire sentence. When information regarding the results of the PROGQ is provided, it will follow the questionnaire number, as in Q55:22, which means that 22 informants used the *dāštan* construction for questionnaire sentence 55. Each informant has a unique label, when questionnaire responses are presented, the informant label is given in square brackets.

3:3 Q55

a) Q55:22
[Somebody's calling and asking for Maryam. I answer:
- Maryam is near me,] she SING a song.

b) Q55:22 [FS]
Somebody's calling and asking for Maryam. I answer:
- Maryam is near me,] *dār-e* *āvāz* *mi-xun-e*.
have.PRS-3SG song IPFV-read.PRS-3SG
'She is singing.'

The 26 informants were recruited through a posting on the LINGUIST List as well as via personal contact. The informants were all born in Iran although some reside abroad. The ages of the individuals range from 25-60, the median age being 33,5 and the group consists of 16 women and 8 men.[60] Out of 26 informants, 17 come from northern Iran.[61]

Informants were instructed to fill out the PROGQ with answers that correspond to what they would say rather than write. Since the conventions of the written language typically do not allow for the *dāštan* construction, and since only the main verb is given in the examples without the potential *dāštan* periphrasis, an example sentence using the *dāštan* construction was given to exemplify the instructions. Filling out a questionnaire is by no means a natural communication situation, and the instructions to write the way you speak do not provide an ideal situation. Thus, the degree of informality between the informants varies so that some may be very colloquial-like while others are less so.

3.2.2 Data and analysis

Table 14 summarizes the results of the questionnaire in a matrix. Vertically, the sentences in which the *dāštan* construction was used are ordered from the sentence where least informants used the pattern to the sentence where most informants used it, as seen in the last column. Horizontally, informants are ordered from the informant who used the *dāštan* construction most restrictively to the informant who used it the most, as seen in the last row. The first column gives the numbers of the sentences in the PROGQ, and the first row gives the informant labels. For example, the second to last row in the table shows that in Q01, 24 out

[60] Four informants did not provide information regarding age. Two informants did not provide information regarding gender.
[61] The informants originate from the following cities: Tehran (7 informants), Behshahr (5), Mashhad (3), Esfahan (3), Shiraz (2), Neka (1), Qazvin (1), Qom (1), Gachsari (1), Bushehr (1), Kerman (1).

of 26 informants used the *dāštan* construction which is more than any other questionnaire sentence, and the second last column shows the informant with the label [MT] who used the *dāštan* construction the most, namely in 30 of the 69 sentences of the PROGQ.

The table includes all 482 occurrences of the *dāštan* construction when occurring with the predicate provided in the PROGQ. All answers in which a predicate other than the assigned one given in capital letters was used were disregarded. For instance, [FF] used the *dāštan* construction but with the predicate *pust gereftan* 'peel' (lit. 'take peel') instead of the given *pust kandan* 'peel' (lit. 'take off peel') in Q49. This and other similar answers are not included in Table 14. However, some other changes were accepted. In Q03, for example, FF omitted the *bāyad* 'is necessary' element and added *ehtemālan* 'probably' preceding the *dāštan* construction, these cases are included. In a few cases, informants have misunderstood the sentences or not paid attention to the given context, creating contradictory sentences such as, e.g., *I lay in the sun for so long that I got burned, [but fortunately I didn't]* in answering Q58. These answers were also disregarded. Furthermore, the table does not include occurrences of the *dar hāl=e* 'in state of', *mašġul* 'busy' and *gir* 'seize' patterns, which will be discussed in section 3.2.2.2. In the table, only sentences where at least one informant used the *dāštan* construction are shown.[62]

Table 14 shows that there are certain sentences where most informants use the *dāštan* construction, sentences where informants vary in their use of the *dāštan* construction, as well as sentences where few informants use the construction. Also, at a first glance, it looks as if there is no sentence for which all informants provided the *dāštan* construction. However, this is misleading since the *dāštan* construction is used in Q57 by all informants apart from those who misunderstood the sentence. This sentence is discussed in section 3.2.2.1. In addition, the table also shows that the informants who used the *dāštan* construction the least generally provide the construction in those sentences that have the highest probability of containing the construction, while the informants that used the *dāštan* construction the most also used the construction in the least probable cases. This clearly indicates that although there are great individual variations and preferences, there is overall agreement as to which of the PROGQ sentences constitute the most favorable uses of the *dāštan* construction.

[62] Sentences where the *dāštan* construction is unexpected (i.e. negated sentences, stative/be/*dāštan* verbs as main verbs and imperatives (Dehghan 1972:200–201; Lazard 1992:160; Mahootian 1997:241–244; Mace 2003:102–103) and where no informant provided the pattern are: 7, 13, 18, 28, 30, 43b, 50, 59, 62, 68. Sentences where the *dāštan* construction could have occurred but did not are: 4, 8, 10, 12, 20, 26, 31, 33, 35, 36, 38, 43a, 54, 63.

Table 14. Results of the PROGQ

Q↓ INF→	SS	NM	MH	PK	RS	SA	AA	ME	FS	MN	AB	SE	MD	FN	MG	FF	SZ	AC	PV	AD	BT	VS	AF	BB	KA	MT	SUM
5															•												
41																										•	
47																•											1
69						•																					
6																				•			•				
24											•															•	2
40											•														•		
51																						•			•		
67																		•					•				
32															•					•				•			
39		•																		•					•		3
42																		•	•							•	
45																					•	•				•	
25															•	•								•		•	4
17									•										•	•	•						
19														•		•		•	•					•			5
29																	•	•		•				•		•	
53															•		•				•		•		•	•	6
61a																	•				•	•	•	•		•	
9									•			•				•		•	•	•							
48														•			•	•				•	•	•	•		
58													•				•			•	•	•	•	•		•	7
61b													•				•		•		•	•	•		•		
21								•			•			•				•	•		•			•			8
60											•		•				•			•		•	•	•	•	•	9
3								•		•					•	•	•	•	•			•			•	•	10
37			•							•		•		•					•	•	•	•			•	•	
11	•		•	•						•		•		•		•			•	•	•		•	•			
44							•	•			•	•	•				•	•		•		•	•	•	•	•	13
46									•			•			•		•	•	•	•	•	•	•	•	•	•	
34										•	•	•	•	•	•	•		•	•		•	•	•	•	•	•	15
49							•		•	•			•		•		•	•	•	•	•	•	•	•	•	•	
15							•		•	•	•		•	•	•	•	•	•	•			•	•	•	•	•	
22							•	•	•		•		•	•	•	•	•	•		•		•	•	•	•	•	16
27									•	•			•	•	•	•	•	•	•	•	•	•	•	•	•	•	
23		•			•	•	•				•	•	•	•		•			•	•		•	•	•	•	•	
52					•	•		•	•	•	•	•		•	•			•	•		•	•	•	•	•	•	17
65				•				•				•	•	•	•		•	•		•		•	•	•	•	•	
56					•	•		•		•		•	•	•		•	•	•	•	•	•	•	•	•	•	•	18
2							•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	19
16						•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	20
14				•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
55					•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	22
64			•	•	•	•	•	•	•		•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	
66		•		•		•	•	•	•	•		•	•	•		•	•	•	•	•	•	•	•	•	•	•	
57	•	•	•	•	•	•			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	23
1	•		•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	24
	3	4	5	6	8	11	12	14	16	16	17	18	20	20	21	21	24	24	26	26	26	27	28	29	29	30	

The results confirm the aforementioned grammatical restrictions of the *dāštan* construction (Dehghan 1972:200–201; Lazard 1992:160; Mahootian 1997:241–244; Mace 2003:102–103; Windfuhr & Perry 2009:461). Thus, it does not occur in sentences containing negation (Q18:0, Q30:0, Q50:0), imperative constructions, (Q28:0), constructions with the verb *budan* 'to be' as main verb (Q07:0, Q13:0, Q43b:0, Q68:0), and with stative verbs such as *dānestan* 'to know' (Q62:0) and *dust dāštan* 'to like' (Q59:0). The *dāštan* progressive is also not expected to occur with the modal verb *bāyad* 'is necessary', which requires the subjunctive form of the verb. In those sentences in the PROGQ where *bāyad* was given, all those informants who used the *dāštan* construction omitted the *bāyad* element (Q03:10, Q41:3). The questionnaire sentences where the *dāštan* construction did not occur despite being possible, are several habitual uses (Q04:0, Q20:0, Q31:0, Q36:0), a sentence with *gereftan* 'to grip/hold' as main verb (Q63:0), a passive construction (Q54:0), a sentence with *šuru kardan* 'begin' as a main verb (Q12:0) and sentences containing the verbs *avizān kardan* 'to hang' (Q08:0), *nešastan* 'to sit' (Q33:0) and *istādan* 'to stand' (Q35:0). In the latter three, the use of the *dāštan* construction would have rendered the proximative reading, I will discuss such uses in section 3.4.2.2. Also, the sentences with future time reference Q10:0 and Q43:0 yielded no occurrences as well as the sentence containing the gradual adverb *lahze be lahze* 'moment by moment'. As will become evident, however, there are other sentences with future time reference and gradual adverbs in the PROGQ where the *dāštan* construction occurs.

Table 15 divides the results of the PROGQ into FOC and DUR contexts, as well as into several sub-uses. The division follows to some extent the original questionnaire by Bertinetto et al. (2000b), where the questionnaire sentences are grouped according to various features. The division is based on more or less explicit information in the questionnaire sentences, or an at least highly probable interpretation of the sentence used with the *dāštan* construction. Only sentences where the *dāštan* construction is expected or possible were included, meaning that sentences containing negation, stative verbs, the verb *budan* 'to be', the verb *bāyad* 'is necessary' and imperatives were all omitted from the table. Also, habitual contexts and sentences with posture verbs were omitted. Groupings similar to those in the table have also been done for the selected data from the five Iranian movies presented in section 3.3.

The FOC contexts were divided into ongoing, proximative, and futurate/proximative uses. It is not always easy to determine such uses as sentences may have several interpretations, I will discuss such issues in section 3.4. The ongoing uses also include absentive uses, since these uses are also ongoing as discussed in section 3.4.2.5. 'Futurate/proximative' refers to sentences that are vague in regard to these uses. The DUR contexts are all ongoing. They are further divided where 'delimited duration' refers to contexts where there is an indication of a delimited period of time during which an event was taking place.

The median value given in the right-most column refers to the number of informants providing the *dāštan* construction. We then see that although the *dāštan* construction can be used in both types of context, there are more informants who used the *dāštan* construction in FOC contexts than in DUR ones. Notably, there are several proximative and futurate/proximative uses where no or very few informants used the *dāštan* construction. In fact, the ongoing uses alone have a median value of 15,5 informants. All the questionnaire sentences are admittedly not ideal for comparing the FOC uses to the DUR uses, since there may be other reasons for informants using or not using the *dāštan* construction in these sentences, but I believe that the indication towards a majority of FOC uses still holds.

Table 15. Uses of the *dāštan* construction in the PROGQ

CONTEXT	TYPE OF USE	RESULTS PROGQ	MEDIAN VALUE
FOC	Ongoing	Q01:24, Q02:19, Q03:10, Q09:7, Q11:13, Q14:22, Q15:16, Q21:8, Q22:16, Q24:2, Q27:16, Q32:3, Q37:10, Q40:2, Q44:13, Q46:13, Q49:15, Q54:1, Q55:22, (?)Q61b:7, Q64:22, Q65:17	
	Absentive	Q16:20, Q34:15	
	Proximative	Q12:0, Q23:17, Q38:0, Q39:3, Q41:1, Q45:3, Q52:17, Q57:23, Q58:7, Q67:2, Q69:1	
	Futurate/Proximative	Q10:0, Q29:6, Q43a:0,	**10**
DUR	Ongoing		
	Delimited duration	Q06:2, Q19:5, Q47:1, Q51:2, Q56:18, Q60:9, Q61a:6 , Q66:22	
	Gradual process	Q17:4, Q26:0, Q48:7	
	Narrative	Q25:4	
	'*Continually*'	Q53:6	**5**

The results of the questionnaire confirm previous analyses of the *dāštan* construction as a progressive gram, as it most often has ongoing use, and does not occur in habitual contexts to a great extent or with stative predicates.

In the upcoming section, we will look at the sentences where the *dāštan* construction was given the most, that is, the most favorable contexts for the pattern.

3.2.2.1 Most favorable contexts for the *dāštan* construction

The sentences where the greatest number of respondents chose the *dāštan* construction are given in 3:4. As illustrated, they all have a FOC reference point, and all but one are in the present tense.

3:4 Top 6 sentences where most informants used the *dāštan* construction

a) Q01:24 [AA]
[Father says to his child:
- Don't disturb me,] *man* *dār-am* *nāme mi-nevis-am.*
I have.PRS-1SG letter IPFV-write.PRS-1SG
'I am writing a letter.'

b) Q57:23 [FF]
pesar=e javān *dāšt* *mi-mord.*
boy=EZ young have.PST.3SG IPFV-die.PST.3SG
'The young boy was about to die [but finally they found the right medicine and cured him].'

c) Q66:22 [BT]
[I am very tired,]
az *vaġti* *ke* *bidār* *šod-am*
from time that awake become.PST-1SG

dār-am *nān* *mi-paz-am.*
have.PRS-1SG bread IPFV-bake.PRS-1SG
'I have been baking ever since I woke up.'

d) Q64:22 [AA]
[Somebody's calling and asking for Ali. I answer:
- Ali is near me,] *u* *dāre* *āmāde* *mi-š-e.*
(s)he have.PRS-3SG ready IPFV-get.PRS-3SG
'He is getting ready.'

e) Q55:22 [SA]
[Somebody's calling and asking for Maryam. I answer:
- Maryam is near me,] *u* *dār-e* *āvāz* *mi-xun-e.*
(s)he have.PRS-3SG voice IPFV-read.PRS-3SG
'She is singing.'

f) Q14:22 [ME]
[Look at Aryan on the sofa]

u	*dār-e*	*xāb*	*mi-bin-e.*
(s)he	have.PRS-3SG	sleep	IPFV-see.PRS-3SG

'He is dreaming.'

Q01:24 given in 3:4a) is the questionnaire sentence where the *dāštan* construction was used the most. This sentence is taken from Dahl (1985:92), where it was shown to have the highest incidence of responses with progressive constructions across languages. It is thus not surprising that the Persian *dāštan* progressive occurs frequently in this context.[63] In this example, the speaker is referring to his engagement in the event, specifically his being busy with working, which is ongoing at the moment of speech. In addition, this example can be imagined to be uttered in an irritated tone.

In Q57:23, given in 3:4b), the informants who did not use the *dāštan* construction clearly misunderstood the sentence.[64] In this sentence, the *dāštan* construction in applying to an achievement event yields the proximative reading rather than a progressive one. We will return to the proximative uses in section 3.4.2.2.

3:4c) is similar to a) in that it refers to the subject's engagement in the activity. In addition, the sentence can be imagined to be uttered as a complaint. I will discuss this example and others similar to it in section 3.4.2.1.

3:4d) and e) are two out of many phone-call dialogues in the PROGQ. They are clear examples of FOC contexts where no explicit reference point is given. If someone calls on the phone and asks about Maryam, the answer *She is singing* is referring to the time of the speech event. In both these examples, the use of the *dāštan* construction implies that the subject is occupied with an activity at the time of utterance.

Q14:22 in 3:4f) contains the predicate *xāb didan* 'to dream', which in Persian consists of the components 'sleep/dream' and 'see', similar to other languages such as, e.g., Russian, Turkish and Hindi/Urdu. This context is very similar to the typical present progressive contexts including *behold* (meaning *look*) that are found in Chapter 2, which aims at making the listener aware of a dynamic event occurring at the moment of speech. Q14:22 implies that there is something to be seen, that there is something in Aryan's behavior on the sofa which makes us believe that he is dreaming, such as some movement of the body or something similar. 'To dream' is then perceived as a clearly dynamic situation here.

[63] In retrospect, it would have been better not to have this sentence as the first sentence of the questionnaire in order to avoid impact from the example sentence given on the instruction sheet.
[64] Three informants AA, AF and ME, provided the past 'died' or negated past 'did not die'. Looking at Q57, it becomes clear that these answers are misunderstandings of the given context: "The young boy died/did not die [but finally they found the right medicine and cured him]."

We can then see that all the examples above, apart from 3:4b), which is a proximative use, have to do with the *engagement* and/or *busyness* of the agentive subject. I understand *engagement* and *busyness* to be somewhat stronger than simply 'carrying out the event' (which is expected of an agent) and mean that the subject has a higher degree of activity and involvement in the event. Linked to engagement and busyness is what has been referred to as an *emotive effect*, which has been noted to arise when progressives are used (Comrie 1976:37). As noted, both 3:4a) and c) can be analyzed as having such.

Most of the sentences where the *dāštan* construction was not given, or given to a lesser extent, have DUR contexts of different sorts, which have already been shown to be less favorable for the *dāštan* construction. These examples favor the use of non-progressive, imperfective grams. Some examples of context where informants did not provide the *dāštan* construction are given in 3:5. The questionnaire sentence in a) is a habitual sentence, b) a narrative sentence and c) a sentence expressing a gradual event, none of which are FOC. Example e) is interesting since the context is FOC and explicitly requires the attention of the listener. Here, however, the subject of the clause is inanimate without volition, what is more, the shining of the sun is not a very dynamic event. In comparison, Q37:10, 'The water BOIL. [Shall I make tea?]', is similar in that it refers to a FOC context without an agentive subject. In this example, however, more informants provided the *dāštan* construction. This is probably due to the boiling of water being more dynamic than the shining of the sun.

3:5 Less favorable contexts for the *dāštan* construction

a) Q04:0
Last year, Aryan VISIT us three times.

b) Q25:4
[It was a bright summer day.] The bees HUM, the birds SING and the cows GRAZE in the pasture. [Suddenly, the earth opened and the devil came out.]

c) Q26:0
[Moment by moment] the policeman TAKE NOTES of what (s)he said.

d) Q46:2
[Look out the window!] The sun SHINE.

Other less favorable contexts were ones where patterns other than the *dāštan* construction were more suitable and/or sentences where informants made an interpretation other than the interpretation available with that construction. In, for example, Q67:2, 'The climber REACH the top of the mountain [when suddenly

he had a heart attack and died.]', most informants used the past tense form rather than the progressive, which would have rendered the proximative reading.

I conclude that the most favorable contexts for the *dāštan* construction are contexts that are FOC, also, these uses often have to do with the engagement or busyness of an agent in the event.

3.2.2.2 Other patterns with ongoing uses

When answering the PROGQ, in addition to the *dāštan* construction, informants provided three other constructions that also have ongoing uses: these are the *dar hāl=e* progressive[65], the *mašġul* construction and the *gir* construction. Already Dehghan (1972:205) mentioned the *dar hāl=e* and *mašġul* constructions as patterns of 'progressive tense' used in literary Persian, although he also noted their not being very common. These patterns were discussed in Chapter 2, section 2.2.2.3, where it was shown that the *dar hāl=e* construction has a distribution similar to other progressives, while the *mašġul* construction only overlaps with that of the progressives to a limited extent.

The total number of occurrences of the *mašġul* construction is 25, the *dar hāl=e* progressive 19 and the *gir* construction 3. These can be compared to the total number of the occurrences of the *dāštan* construction, which is 482. It is thus clear that when instructed to use colloquial speech, speakers mainly use the *dāštan* construction for marking progressivity. The *gir* construction is only provided by one informant, [FN], from the southern Iranian city Bushehr and could thus be a dialectal or idiolectal pattern.

The *dar hāl=e* progressive consists of a prepositional phrase with *dar* 'in' + *hāl* 'state' + *ezafe*[66] + main verb in the infinitive + copula or the existential verb *hast*, as illustrated in example 3:6. The infinitival construction in this example is a combination of the light verb construction *dorost kardan* 'to fix, to prepare' and *ġazā* 'food'. The sentence thus has a literal meaning 'I am in the state of food-making'.

3:6 Q09 [ME]

[If you come at 8 o'clock]
dar *hāl=e* *ġazā* *dorost* *kardan* *hast-am.*
in state=EZ food right do.INF be.PRS-3SG
'I will be (lit 'am') cooking. [Come a little later, please.]'

[65] I referred to this pattern as the 'be-in-state-of' pattern in Chapter 1.
[66] *Ezafe* is an enclitic particle that attaches to the head in a phrase and connects it to a dependent element (see, e.g., Lazard 1992).

The *mašġul* construction contains the adjective *mašġul* 'busy' in an *ezafe*-construction with a verb in the infinitive followed by a copula, similar to the *dar hāl=e* progressive. Thus, 3:7 has the literal meaning '(S)he was busy doing work'. [67] In both the *dar hāl=e* and *mašġul* patterns, the copula may be realized in its long or short form. Also, in 3:7, the infinitival verb could be omitted, i.e. *mašġul=e kār bud* '(s)he was busy working', although such an example was not provided by any informant.

3:7 Q65 [SA]

[Last night at 8 o'clock, when Aryan arrived, Maryam]
mašġul=e *kār* *kardan* *bud.*
busy=EZ work do.INF be.PST.3SG
'She was busy working.'

The example in 3:7 is very similar in meaning to the English translation in that the 'busy' elements in both languages retain their literal meaning.

The *gir* 'seize, trap' construction is very similar to the *mašġul* and *dar hāl=e* constructional schema. Since it only occurred three times in the data, it will not be discussed further.

3:8 Q03 [FN]

[- What is Nima doing?
- I don't know but I think that:]
bāyad *al'ān* *gir=e* *ġazā* *xordan* *bāš-e.*
is.necessary now trap=EZ food eat.INF be.SBJV.PRS-3SG
'(S)he must be busy/trapped eating now.'

We will now turn to the occurrences of the *mašġul* and *dar hāl=e* constructions in the PROGQ, which are given in Table 16, with the questionnaire sentences given in column 4. The table includes all occurrences of the two patterns in the questionnaire responses. Similar to what was done for the *dāštan* construction, the questionnaire sentences have been divided into FOC and DUR contexts. The imperative and negated sentences are given separately. It is important to keep in mind the relative low frequency of occurrence for these two constructions in comparison to the *dāštan* construction. Thus, the functional span of these

[67] Consulting informants suggest that the Kurdish of Kermanshah, Saqqez and Marivan, as well as Hawrami as spoken in Paveh, may have a progressive construction formed with *xarik* 'busy'. The present progressive in Sorani Kurdish includes *xerík* plus the present subjunctive or indicative, which is suggested to have originated from the meaning "it is imminent that it happens" (Jügel et al. 2017:164–165).

constructions cannot be examined fully, but rather certain tendencies are noted and discussed here.

Similar to the *dāštan* construction, we note that both the *mašġul* and *dar hāl=e* constructions occur more often in FOC than in DUR contexts. Especially the *dar hāl=e* construction occurs in FOC contexts almost exclusively, whereas the *mašġul* construction appeared twice as often in FOC contexts as in DUR ones (similar to the *dāštan* construction). This, however, does not mean that the *dar hāl=e* constructions cannot occur in DUR contexts.

The highest frequencies for both constructions are found in two questionnaire sentences in which the *dāštan* construction cannot occur, that is, in Q03 (6 *mašġul* and 7 *dar hāl=e* constructions) and Q42 (3 *mašġul* and 5 *dar hāl=e* constructions), where the modal element *bāyad* 'is necessary' occurs. As already mentioned in section 3.1.1, the *dāštan* construction only occurs in the indicative mood and cannot combine with *bāyad* 'is necessary', which requires the subjunctive form of the main verb. Both these sentences involve FOC ongoing events, i.e. typical sentences where progressives are expected. In fact, many informants simply omitted *bāyad*, at the price of losing the modal meaning. In, e.g., Q03, instead of providing the intended sentence *He must be eating right now*, they provided a sentence *He is eating right now* including the *dāštan* construction.

Further, we may note that *mašġul* occurs in a negated and an imperative sentence, where the *dāštan* construction is normally excluded. The *dar hāl=e* pattern was not used in these sentences, although it would be possible in principle.

The *mašġul* construction differs from the *dāštan* construction in several respects, the main difference being that it is less grammaticalized as it retains its meaning of 'busy'. Another feature of the *mašġul* construction is that it requires an agentive subject, that is, the subject of the predicate is typically animate and acting volitionally. This is perhaps most evident in Q46 '[Look out the window!] The sun SHINE', in which the use of *mašġul* was not accepted by a consulting native speaker. Both the *dāštan* and *dar hāl=e* constructions were used in this sentence in the responses.

Like the *dāštan* construction, the *mašġul* construction does not seem to go well with stative events. On the other hand, it does not seem to have the proximative reading. Example 3:9, which is an additional example not included in the PROGQ, was not accepted by the consulting native speaker in a proximative reading.

3:9 *mašġul=e* *oftādan* *bud.*
busy=EZ fall.INF be.PST.3SG
'(S)he/it was (in the midst of) falling'

Table 16. Occurrences of the *mašġul* and *dar hāl=e* constructions in PROGQ

CONTEXT/ USE	*maš.*	*hāl*	QUESTIONNAIRE SENTENCE	TOTAL	
				maš.	*hāl*
FOC ONGOING	-	1	Q01: [Father says to his child: - Don't disturb me,] I WRITE LETTER.	14	17
	6	7	Q03: [- What is Nima doing? - I don't know but I think that:] he must EAT right now.		
	2	3	Q09: If you come at 8 o'clock, I still COOK. [Come a little later, please.]		
	1	-	Q11: [Somebody's calling and asking for Maryam. I answer: - Maryam is near me,] she WORK right now.		
	3	5	Q42: [- What is Yasaman doing? - I don't know but I think that:] she must TEACH right now.		
	-	1	Q46: [Look out the window!] The sun SHINE.		
	1	-	Q55: [Somebody's calling and asking for Maryam. I answer: - Maryam is near me,] she SING a song.		
	1	-	Q65: [Last night at 8 o'clock,] when Aryan arrived, Maryam still WORK.		
FOC ONGOING ABSN	2	-	Q34: [On the phone: - Is Nima home right now? - No, he PLAY CARDS [as usual].	2	2
	-	1	Q16: [On the phone: - Is Nima with you right now?] - No, he PLAY CARDS [in the next room].		
FOC PROX?	-	1	Q57: The young man DIE [but finally they found the right medicine and cured him].	-	1
DUR	1	-	Q19: [During the whole time of the class] Maryam TALK to her neighbor [in fact, she carried on even afterwards].	6	1
	2	-	Q25: [It was a bright summer day.] The bees HUM, the birds SING and the cows GRAZE in the pasture. [Suddenly, the earth opened and the devil came out.]		
	1	-	Q60: [Think! As we are talking] the earth TURN around the sun.		
	-	1	Q61a: Yesterday, while Maryam STUDY in her room, Aryan PLAY in the courtyard.		
	1		Q61b: Yesterday, while Maryam STUDY in her room, Aryan PLAY in the courtyard.		
	1	-	Q66: [I am so tired] I BAKE BREAD all day since I got up this morning.		
NEG	1	-	Q50: The boss was angry, because when he came in Nima NOT WORK.	1	-
IMP	2	-	Q28: [For goodness sake] WORK when the boss comes back!	2	-

The *mašġul* construction can be shown to focalize the event. A sentence like 3:10 which is ambiguous for focality in the simple present, becomes unambiguous when the *mašġul* construction applies to it.

3:10 *mašġul=e* *sigār* *kešidan=am.*
busy=EZ cigarette pull.INF=COP.PRS.1SG
'I am busy smoking.'

Q25 consists of three predicates: [It was a bright summer day.] The bees HUM, the birds SING and the cows GRAZE in the pasture. [Suddenly, the earth opened and the devil came out.]. In this sentence, two informants used the *mašġul* construction but only for one of the three predicates in the sentence. They then seem to have used it for stylistic variation in order to not repeat the same construction three times.

Unlike the *mašġul* construction, the *dar hāl=e* construction does not have agency as a requirement for the subject, which is evident in example 3:11.

3:11 Q46 [AA]

[Look out the window!]
xoršid *dar* *hāl=e* *tābidan=e*.
sun in state=EZ shine.INF=COP.PRS.3SG
'The sun is shining.'

It is not clear whether the *dar hāl=e* pattern has a proximative use or not. Looking at example 3:12, where one informant used the *dar hāl=e* pattern, it is unclear if we have a proximative reading or a progressive one, that is, if the event is viewed as being about to happen or if it refers to a longer period of time during which the young boy was dying. A consulting native speaker noted that perhaps both readings are available.

3:12 Q57 [SA]

pesar=e *javān* *dar* *hāl=e* *mordan* *bud*
boy=EZ young in state=EZ die.INF be.PST.3SG
'The young boy was about to die/was (in the process of) dying [but finally they found the right medicine and cured him]'

Although the *dar hāl=e* pattern occurs almost exclusively in FOC ongoing contexts, it my impression that it does not necessarily have a FOC sense in the same way as the *dāštan* progressive. Rather, my impression is that it marks ongoingness. This impression, as well as the issue of proximative reading, needs to be further examined, preferably with non-elicited, naturally occurring data.

This chapter will continue by discussing the uses of the *dāštan* construction.

3.3 The *dāštan* construction in five Iranian movies

3.3.1 Method

This section presents the *dāštan* construction as it appears in five Iranian movies produced in Iran between 2002 and 2012. The movies were selected for having dialogue similar to natural speech. The total length of the films is approximately eight and a half hours. Table 17 lists the movies.

In what follows, examples from the movies will be accompanied by the title of the movie in English and the time of the utterance containing the *dāštan* construction. All occurrences of the *dāštan* construction are given in Appendix D where they are marked with one or several tags referring to their uses. These uses are discussed together with the results from the PROGQ in section 3.4.

Table 17. The five Iranian movies used in the studies

MOVIE TITLE	DIRECTOR	PROD. YEAR	LENGTH
زندان زنان 'Women's Prison'	Manijeh Hekmat	2002	1h 45min
سنتوری 'The Music Man'	Dariush Mehrjui	2007	1h 47min
جدایی نادر از سیمین 'A Separation'	Asghar Farhadi	2011	2h 3min
سعادت آباد 'Felicity Land'	Maziar Miri	2011	1h 24min
من مادر هستم 'I am a Mother'	Fereydoun Jeyrani	2012	1h 37min

3.3.2 Data and analysis

Table 18 shows the uses of the 84 utterances found in the movies. As shown, the *dāštan* construction has many more FOC uses than DUR ones, more so than in the result from the PROGQ. In addition, we also find proximative and iterative uses of the *dāštan* construction. Also, it is at times difficult to distinguish between, for instance, progressive and proximative uses or between futurate and proximative uses, as certain contexts are ambiguous or vague in this respect. These issues will be discussed in section 3.4.

Table 18. Functional tags of the results from five Iranian movies

CONTEXT	TYPE OF USE	NUMBER OF OCCURRENCES
FOC	Ongoing	48
	Proximative	15
	Futurate/Proximative	9
	Ongoing/Proximative	3
DUR	Ongoing	2
	Iterated ongoing	2
Other	FOC ongoing/DUR ongoing	2
	Gradual process/proximative	1
	Untagged	2
TOTAL		**84**

The present and past tense of the 84 *dāštan* constructions found in the movies are presented in Table 19. The results show that the *dāštan* construction is much more often used in the present than in the past. This ratio is similar to the ratio of present over past found in Chapter 2 in the TED corpus.

Table 19. Present or past tense in the five Iranian movies

MOVIE TITLE	NO. OF *DĀŠTAN* CONSTRUCTIONS	PRESENT TENSE	PAST TENSE
'Women's Prison'	8	7	1
'The Music Man'	11	11	0
'A Separation'	31	24	7
'Felicity Land'	18	17	1
'I am a Mother'	16	15	1
TOTAL	**84**	**74**	**10**
PERCENTAGE	**100%**	**88%**	**12%**

Two examples are given in 3:13, both of which are utterances in FOC contexts. 3:13a) constitutes what in the literature has been referred to as historical or narrative present and is also noted for, e.g., the English and French (*être en train de* INF) progressive patterns (see, e.g., Comrie 1976:73; De Wit et al. 2013:850). In both a) and b), the *dāštan* construction in the present tense seems to highlight focality. Notably, both examples are given with a non-neutral tone expressing irritation and irony, respectively, i.e. expressing *emotive effect*. In fact, in a), the use of the *dāštan* construction seems to have a FOC and emotive effect only and does not seem to mark that the event of saying is ongoing.

3:13 FOC ongoing contexts in present tense

a) ‘A Separation’, 1:10:06 and 01:10:08

diruz	*dār-e*	*mi-g-e*	*man*	*aslan*
yesterday	have.PRS-3SG	IPFV-say-3SG	I	at.all

ne-mi-dunest-am	*bārdār*	*bud-e.*	*emruz*
NEG-IPFV-know.PST-1SG	pregnant	be.PST-3SG	today

dār-e	*mi-g-e*	*man*	*aslan*	*hol=eš*
have.PRS-3SG	IPFV-say.PRS-3SG	I	at.all	push=3PC.SG

na-dād-am.
NEG-give.PST-1SG
‘Yesterday he was (lit. ‘is’) saying: I didn’t know at all she was pregnant. Today he is saying I didn’t even push her.’

b) ‘A Separation’, 1:03:45
[The mother says to her daughter in an ironic way.]

če	*ġad*	*ham*	*bābā=t*	*negarān*	*bud*
how	much	also	father=2PC.SG	worried	be.PST.3SG

man	*dār-am*	*mi-r-am.*
I	have.PRS-1SG	IPFV-go.PRS-1SG

‘And your father, he was so upset that I am leaving.’

Surprisingly, one example of a negated *dāštan* construction is found in the movie data. This is a rather special case where what is negated is what has just been stated by another speaker. I will discuss this example and the restriction on negation in Chapter 4, section 4.4.

3.4 Discussion

In this section, we will look at and examine the data presented in this chapter. The data showed that the main use of the *dāštan* construction is to mark that an event is ongoing at the reference time, most often in FOC context. The construction is also far more often used in the present tense than in the past, as evident by the data collected from the movies. In what follows, I will further discuss the different uses of the construction, starting with ongoing uses in the present and past, which are the most frequent ones, and continuing with the peripheral uses, which are also less common.

3.4.1 The *dāštan* construction in present and past tense

The data collected from the five movies shows that the *dāštan* construction is far more often used in the present tense than in the past tense. Assuming that present time reference is more common than past time reference in spoken language, this is not surprising. As will become evident, there are certain contexts in the data, namely backgrounding contexts, that are only found with the *dāštan* construction in the past tense. At times, differentiating between FOC and DUR contexts is not easy to do, below I will give examples from the clearest cases.

Typically, the *dāštan* construction in the present tense occurs in FOC contexts denoting an ongoing event. Often, the FOC reference time is not expressed explicitly in the clause, by, e.g., a phrase such as 'right now' or 'look!', but is obtained from the context. A few examples from the movies are given in 3:14. In order to better understand these sentences, in what follows, the contexts of these sentences are reproduced.

3:14 FOC contexts in the present tense

a) 'Women's Prison', 09:55
[A female prisoner rushes to the other prisoners, alarmed at what is happening and informs the others that the guards are taking their things from their cells to the main office.]
tamām=e *asās-ā=mun=o* *dār-an*
all=EZ things-PL=1PC.PL=DIR have.PRS-3PL

mi-bar-an *daftar*
IPFV-take.PRS-3PL office
'They are taking all our stuff to the office.'

b) 'The Music Man', 01:26:41
[The subject, an addict, breaks down after realizing that he will not receive any drugs.]
in *me'de=ye* *man* *dār-e* *mi-suz-e.*
this stomach=EZ I have.PRS-3SG IPFV-burn.PRS-3SG
'This stomach of mine is burning.'

c) ‘Felicity Land’, 30:30
[The speaker, worried about the consequences of a phone call, is informing the others about the fact that the subject is making a phone call, as if urging them to prevent it.]
dār-e *beh=eš* *zang* *mi-zan-e.*
have.PRS-3SG to=3POSS call IPFV-hit.PRS-3SG
‘He is calling her.’

d) ‘I am a Mother’, 01:24
[A woman is sitting in a therapist’s office, remembering a dream and telling it as she envisions the event, i.e. as if it occurs before her eyes in that moment.]
dār-e *mi-r-e* *be* *samt=e* *daryā.*
have.PRS-3SG IPFV-go.PRS-3SG to direction=EZ sea
‘She is walking towards the sea.’

All these utterances are given in FOC contexts. A consulting native speaker confirmed that the FOC reading of examples a) and d) would be weakened if the *dāštan* element were omitted, i.e. if the simple present tense was used instead. Interestingly, when given the movie contexts in which b) and c) occurred, a consulting native speaker noted that they have a sense of being incomplete without the *dāštan* verb, i.e. with the simple present. In both cases, we are dealing with alarming events. It may then be that the *dāštan* construction is on its way to being conventionalized, i.e. less optional, in such contexts.

Interestingly, all these examples are uttered in order to turn the attention of the listener towards or make the listener aware of an ongoing event. As previously mentioned, these types of contexts are typical present ongoing uses for progressives.

Contexts with DUR reference times are also found, although they are not as common as FOC ones, an example is given in 3:15. I believe that, in this example, the use of the *dāštan* construction intensifies the statement. In section 3.4.2.1, the DUR uses will be further discussed.

3:15 ‘Felicity Land’, 53:31, DUR contexts in the present tense

dah *sāl=e* *dār-i* *hamin* *mozaxraf=o*
ten year=COP.3SG have.PRS-2SG this rubbish=DIR

mi-g-i.
IPFV-say.PRS-2SG
‘For ten years (now), you have been (lit. ‘are’) saying the same rubbish.’

In the past tense, we also find FOC uses of the *dāštan* construction, two examples are given in 3:16. Both examples are narratives and include backgrounding contexts. In a), the progressive is given in an utterance following a backgrounding event, in b) the progressive is part of a backgrounding event.

3:16 FOC contexts in the past tense

a) 'Women's prison', 01:31:22
vaġti *az* *birun* *mi-umad-am* *did-am*
when from outside IPFV-come.PST-1SG see.PST-1SG

dāštan *mi-bord-an=eš* *enferādi=ye pāin.*
have.PST-3PL IPFV-bring-3PL=3PC.SG solitary=EZ down-stairs
'When I was coming in from outside, I saw that they were taking her to the solitary confinement down-stairs'

b) 'A separation', 01:33:17
hamun *mouġe* *ke* *dāšt* *bā* *mo'alem=et*
that moment that have.PST.3SG with teacher=2PC.SG

tu hāl *harf* *mi-zad,*
in hall talk IPFV-hit.PST.3SG

'At that moment when she was speaking to your teacher in the hall, [I heard everything they were saying from the kitchen.]'

In example 3:17, two utterances in DUR contexts in the past are given. In 3:17a), which has a clear sense of reproach, the explicit duration of 'half an hour' is given in the clause containing the *dāštan* construction. Similar to the example with the DUR context in 3:15, I believe that the *dāštan* construction intensifies the utterance, and possibly also the sense of reproach. 3:17b) is better analyzed as a DUR backgrounding use of the *dāštan* constructions where Q61a:6, referring to the first clause containing the *dāštan* construction, provides the backgrounding context, and Q61b:7, referring to the second clause containing the *dāštan* construction, provides the event taking place during that time.[68]

[68] Note that *dār hāl=i ke* 'while' is similar in its form to the *dār hāl=e* progressive.

3:17 DUR contexts in the past tense

a) 'Felicity Land', 01:16:26
[- By the way, what were you and Bahram talking about?
- Nothing.]
- pas nim-sā'at tu bālkon dāštin či migoftin?
so half-hour in balcony have.PST-2PL what IPFV-say.PST-2PL
'- So what were you talking about then for half an hour on the balcony?'

b) Q61a:6 and Q61b:7 [SZ]
diruz dar hāl=i ke maryam dāšt tu
yesterday in state=INDF that Maryam have.PST.3SG in

otāġ dars mi-xund, āriyan dāšt tu
room study IPFV-read.PST.3SG Aryan have.PST.3SG in

hayāt bāzi mi-kard.
garden play IPFV-do.PST.3SG
'Yesterday, while Maryam was studying in her room, Aryan was playing in the garden.'

We have then seen that most uses of the pattern are in the present tense and most typically involve ongoing events in FOC context. In the present, the progressive is often used to make the listener aware of an ongoing event. In the past, the *dāštan* construction is found in FOC contexts as well as in DUR contexts, some of which are backgrounding contexts.

3.4.2 Peripheral uses of the *dāštan* construction

In this section, the peripheral uses that were found in the data are discussed in turn, starting with uses in DUR contexts and moving on to proximative, iterative, futurate and absentive uses. As will become evident, the uses that are discussed are quite different from one another: some uses are readings that arise in certain contexts, others are meanings that arise when the *dāštan* construction applies to different types of events.

3.4.2.1 Uses in DUR contexts

Looking at the data from the PROGQ, two questionnaire sentences with DUR reference times, Q56:18 and Q66:22, stand out due to the high proportion of the occurrence of the *dāštan* progressive. Q66:22 was previously given in 3:4c),

in 3:18 questionnaire sentence Q56:18 is given. As can be seen, the sentences are very similar to one another.

3:18 Q56:18 [BT]

[I am very tired,]
az *vaġti* *ke* *pā* *šod-am*
from time that foot become.PST-1SG

hamintour *dār-am* *ġazā* *dorost* *mi-kon-am.*
repeatedly have.PRS-1SG food make IPFV-do.PRS-1SG
'I have been (lit. 'am') cooking ever since I got up.'

In both Q66:22 and Q56:18, the event expressed by the verb stretches from the morning of the day of the speech act to the speech act itself with no FOC reference time. The utterance also carries a sense of complaining. Given the contextual information 'I am very tired', we can assume that the baking has been quite intensive. The question arises as to why the progressive is common in these DUR examples but not in others. I will suggest that in Q56:18 and Q66:22 the use of the *dāštan* construction is triggered by the need to increase the emotive effect and mark the engagement of the subject in the event, i.e. to mark that the subject was in the midst of, or busy with, cooking and is therefore tired. Looking at the DUR contexts where no or few informants provided the *dāštan* construction in 3:19, we see that they do not involve complaint, irritation or similar, nor an especially high level of engagement or 'busyness' of the subject in the event.

3:19 DUR contexts

a) Q06:2
[Yesterday, during my sleep] Yasaman PLAY for 2 hours all by herself.

b) Q26:0
[Moment by moment] the policeman TAKE NOTES of what (s)he said.

c) Q47:1
When the secretary is ill, the boss TYPE his own letters.

Q51:2, given in 3:20, which resembles Q56:18 and Q66:22, is also given with a DUR reference time in a context where the subject is explicitly said to be tired. In Q51:2, however, only two informants provided the *dāštan* construction. Notably, in 3:4c), 3:18 and 3:20, the English translation includes a perfect progressive.

3:20 Q51:2 [VS]

diruz	*vaġti*	*āriyan*	*umad*	*xune*	*xeili*
yesterday	when	Aryan	came.PST.3SG	home	very

xaste	*bud*	*čon*	*tamām=e*	*hafte=ro*	*dāšt*
tired	be.PST.3SG	because	all=EZ	week=DIR	have-PST.3SG

saxt	*kār*	*mi-kard.*
hard	work	IPFV-do.PST.3SG

'Yesterday when Aryan got home, he was very tired because he had been working (lit. 'was working') hard all week.'

I suggest that the lower occurrence of the *dāštan* construction in 3:20 has to do with the lesser need for emotive effect in this sentence. In 3:20, the sense of consequentiality is not as strong as in 3:18, since the sentence is in the past tense and is conceived of as a narrative. Thus, while 3:18 can be uttered as a complaint that may result in some favorable action or sympathy for the subject, 3:20, having happened in the past as well as being a narration about a third person, cannot have a similar effect to the same extent. It is perhaps not surprising that the utterances with progressives given in the present tense have a more urgent and alarming character than utterances in the past, which often appear in backgrounding contexts in narratives. The reluctance to use the *dāštan* construction in 3:20 may also be affected by the length of the sentence, which is longer than those presented in Q66:22 and Q56:18, the adding of yet another verb may be conceived of as making the sentence too heavy.

Emotive and intensifying uses are not restricted to events with DUR reference times. Johanson (2000:86) notes that "[a] higher focality degree may also be chosen to express a higher internal dynamicity, intensity or actional density of what is going on at [the orientation point]". Mahootian (1997:114) notes that the *dāštan* construction may "emphasize a sentence" in examples such as 3:21a). This sentence involves a sense of irritation, which is presumably why the clause in the translation has been reduplicated. Interestingly, again, the simple present tense sounds odd in this context, as confirmed by a native speaker. Similar emotive and intensified uses are also found in many of the utterances from the five Iranian movies, one of which is given in 3:21b), where the *dāštan* construction is used contrastively.

3:21 Intensive/emphatic FOC uses

a) Mahootian (1997:114)
- *bo-ro* *xākrube=ro* *be-ndāz* *dur*.
IPFV-go.PRS garbage=DIR SBJV-throw out

- *dār-am* *mi-r-am*.
have.PRS-1SG IPFV-go.PRS-1SG
'-Go throw out the garbage.
- I am going, I am going.'

b) 'A Separation', 57:33
- *touhin* *na-kon.*
insult NEG-do.PRS.2SG

- *man touhin* *dār-am* *mi-kon-am* *hāj-āġā yā in*?!
I insult have.PRS-1SG IPFV-do-PRS-1SG hadji-Sir or this.one
'- Don't be insulting.
- Am I the one who is insulting, Sir, or this person?!'

Notably, in all examples with DUR time span in this chapter, the DUR contexts are explicitly marked. Also, when asked about the use of the *dāštan* construction in DUR contexts, consulting informants often comment that the use requires a specific situation and sounds odd in isolation. This is probably a reflection of their being less frequent than FOC uses. The DUR uses then seem to be pragmatically marked in a way that FOC contexts are not.

3.4.2.2 Proximative use

This section looks at utterances with proximative use. This meaning arises when the *dāštan* construction applies to events viewed as achievements. Such uses have previously been noted in the literature (e.g. Dehghan 1972; Jahani 2017; Rafiei 2017). As mentioned in Chapter 1, some languages use different patterns for the proximative 'being on the verge of V-ing' and the avertive 'was on the verge of V-ing but did not V' (Kuteva 2001:77, 92, 94). In Persian, the *dāštan* construction can be used for both these meanings. Johanson (2000:153–154; 2017:31) notes that progressive grams are often ambiguous in marking progressive and proximative uses. This section will show that the proximative uses also have typical features distinguishing them from ongoing uses.

In the PROGQ, many sentences with events viewed as achievements have none or very few occurrences of the *dāštan* construction, i.e. Q12:0, Q38:0, Q39:3, Q41:0, Q45:3, Q67:2. As was already noted in section 3.2.2, the questionnaire sentence where all informants (who did not misunderstand the

context) used the *dāštan* construction is Q57:23, 'The young man DIE [but finally they found the right medicine and cured him]', where its most prominent reading is the proximative, or more specifically the avertive, i.e. 'was on the verge of happening but did not'. I believe that the lack of other constructional options for this sentence has to do with the level of ease with which we can construct avertive contexts in a questionnaire, in the case of Q57:23 by adding the contextual information '[but finally they found the right medicine and cured him]'. A few more examples of proximative uses are given in 3:22.

3:22 Proximative use

a) Q23:17 [PV]
[Hurry up!]
ġatār dār-e mi-r-e.
train have-PRS-3SG IPFV-leave.PRS-3SG
'The train is about to leave.'

b) Q52:17 [PV]
[We have to eat these apples.]
unā dār-an mi-pus-an.
they have-PRS-3PL IPFV-rot.PRS-3PL
'They are about to rot.'

c) 'A separation', 40:47
[The speaker, in referring to finding his father alone in the apartment, fallen from his bed.]
dāšt mi-mord man resid-am.
have.PST.3SG IPFV-die.PST.3SG I arrive.PST-1SG
'He was about to die when I arrived.'

d) 'Woman's prison', 26:17
[In the women's prison, shouted by one of the prisoners to make the guards aware of what is happening.]
golandām dār-e bačč=aš mi-a-d!
Golandam have.PRS-3SG child=3PC.SG IPFV-come.PRS-3SG
'Golandam's baby is about to come/is coming!'

e) Q58:7 [MD]
unġadr *tu āftāb* *deraz* *kešid-am* *ke* *dāšt-am*
that.much in sun long drag.PST-1SG that have.PST-1SG
siāh *mi-šod-am.*
black IPFV-become.PST-1SG
'I lay in the sun for so long that I was about to get burned [but fortunately I didn't].'

Several things are observed here. To begin, these uses typically refer to events that are very close to happening. In 3:22a), for example, there may be signs of the train starting to leave (such as speaker announcements or people hurrying to get on or similar). In 3:22b), the apples may have started to show signs of going bad, perhaps some are even rotten, while others are not, as suggested by one informant. The situation in 3:22c) is also urgent and alarming, with the subject being in very bad shape at the time of the arrival. In 3:22d) the woman being referred to is already in labor. Similarly, in 3:22e) we have to assume that there are signs showing that the subject almost got burned. What seems typical of proximative uses is that the event referred to can, or even preferably should, be prevented (Östen Dahl, p.c.), although this is not always the case, as in 3:22d).

Removing the *dāštan* element, i.e. using the simple present or past imperfective, creates sentences that are ill-formed given these particular contexts. Instead, other tense-aspect patterns can be used, such as the past perfective in 3:22a), b) and d), giving rise to the meanings *The train has left!*, *The apples have rotted* and *Golandam's baby came/has come* (preferably with the emphasizer *dige* 'other') which could be used hyperbolically to refer to the same situation. This means that the proximative uses cannot be compared to present tense and past imperfective in the same way as the FOC ongoing use. This has to do with the meanings available for present and past imperfective grams: while they may include the ongoing meaning, they do not include the proximative one.

Ongoing and proximative uses are often difficult to tell apart, and the choice is not always important in actual discourse. Some utterances can be seen as *ambiguous*, i.e. the ongoing and the proximative interpretations describe two different situations. A situation in which a cup is on the edge of a table about to fall is unambiguously proximative, since the cup is not in the air, in the process of falling. The situation with a cup in the air, on the other hand, is unambiguously ongoing. In Persian, an utterance containing the *dāštan* construction such as *dāre miofte* 'it is about to fall/falling' can be used to refer to both the proximative and the ongoing situation. In other cases, the difference is not that distinct, rather, one meaning is more prominent than the other. In 3:22c), e.g., the dying of the subject is presented as having been imminent and something that almost happened but was prevented.

In some cases, the same situation can be conceptualized in different ways, one ongoing and one proximative, the event is then *vague* regarding these two

uses. 3:22d) can be understood as an extended process if the arriving of the child is perceived of as a stretched out event, in which case the ongoing reading 'the child is in the process of coming' arises, or seen as a punctual event in which case the proximative 'the child is about to come' reading arises.

Many of the utterances with proximative use in the five movies were used hyperbolically. It is at times difficult to differentiate between a progressive and proximative use when the *dāštan* construction is used hyperbolically. In example 3:23a), the speaker is worried that her husband will find out about her secret. Her having a heart attack is thus clearly an exaggeration. Similarly, the utterance in 3:23b) is an exaggeration and not a factive evaluation of the speaker's mental health. These examples could suggestively be analyzed as someone on the verge of having a heart attack/going crazy expressed as exaggerations, or they could be analyzed as FOC ongoing uses which receive the proximative use as they are used hyperbolically.

3:23 Hyperbolic uses with achievements

a) 'Felicity Land', 38:58
dār-am *sekte* *mi-kon-am.*
have.PRS-1SG heart.attack IPFV-do.PRS-1SG
'I am having a heart attack.' Or 'I am about to have a heart attack.'

b) 'Felicity Land', 44:35
lāle *man* *dār-am* *az* *negarāni* *divune*
Laleh I have.prs-1sg from worry crazy

mi-š-am, *mi-fāhm-i?*
IPFV-become.PRS-1SG IPFV-understand-2SG
'Laleh, I am going (lit. 'about to become') crazy from worrying, do you understand?'

It has already been mentioned that the *dāštan* construction cannot combine with stative predicates. For example, *dāram midānam* or *dāram hastam*, intended to mean 'I am knowing' and 'I am existing', are ill-formed. The *dāštan* construction can, however, occur together with the posture verbs *nešastan* 'sit down, sit', *derāz kešidan* 'lie down, lie' and *istādan* 'stand up, stand', in which case the proximative interpretation arises. But while these verbs are stative in the perfect and pluperfect, they are understood as having dynamic meaning in other forms and are referred to as "change of state verbs" in Windfuhr & Perry (2009:456). In the data, no examples with the *dāštan* construction combining with posture verbs were found, but such combinations were accepted by consulting informants and were noted to have proximative uses.

3:24 *Dāštan* with posture verbs

a) *dār-am* *mi-šin-am.*
have.PRS-1SG IPFV-sit.PRS-1SG
'I am about to sit down.'

b) *dār-am* *derāz* *mi-keš-am.*
have.PRS-1SG long IPFV-drag.PRS-1SG
'I am about to lie down.'

c) *dār-am* *vāi-mi-st-am/mi-ist-am.*
have.PRS-1SG PVB-IPFV-stand.PRS-1SG/IPFV-stand.PRS-1SG
'I am about to stand up.'

In Chapter 2, it was shown that a number of progressive patterns have uses with posture verbs, but these uses had ongoing rather than proximative meaning. This difference is due to the Persian examples all referring to achievement events, while the examples in Chapter 2 referred to durative events.

3.4.2.3 Futurate use

The *dāštan* construction may occur in contexts with future time reference, i.e. futurate uses, as seen in example 3:25. Most often these uses involve a motion verb.

3:25 Q29:6 [BB]

maryam *fardā* *dār-e* *mi-r-e.*
Maryam tomorrow have.PRS-3SG IPFV-leave.PRS-3SG
'Maryam is leaving tomorrow/Maryam is about to leave tomorrow.'

In Q29:6, the progressive gram is used for marking future time reference. As we will see in Chapter 4, not all progressive grams can occur in such contexts. Q09:6 in 3:26a) is different from 3:25 in that the progressive refers to an ongoing event that is happening in the future.

3:26 Q09, futurate uses

a) Q09:6 [SE]

agar	*to*	*sā'at=e*	*8*	*bi-ā-i*	*man*	*hanuz*
if	you	hour=EZ	8	SBJV-come.PRS-2SG	I	still

dāram	*ġazā*	*dorost*	*mi-kon-am.*
have.PRS-1SG	food	make	IPFV-do.PRS-1SG

'If you come at 8 o'clock, I will still be cooking. [Come a little later, please.]'

b) Q09 [MN]

agar	*to*	*sā'at=e*	*8*	*bi-ā-i*	*man*	*hanuz*
if	you	hour=EZ	8	SBJV-come.PRS-2SG	I	still

dār	*hāl=e*	*ġazā*	*dorost*	*kardan*	*xāh-am*	*bud.*
in	state=EZ	food	make	do.INF	want.PRS-1SG	be.PST

'If you come at 8 o'clock, I will still be cooking. [Come a little later, please.]'

The futurate use of the *dāštan* pattern has also been noted by Jahani (2008:169), who provides a similar example with the motion verb *raftan* 'go, leave', as in '*Quchali, how well on time you came.* ***I am leaving*** (i.e. I intend to leave any moment). *Don't leave your sister alone*.' (my emphasis). Recall that Jahani (2017:261) refers to the *dāštan* construction as a progressive/prospective. Dahl notes that the future uses of motion verbs tend to refer to events happening at a point in the near future and involve a preparatory phase: "One can truly say *I'm going to town* when one has started to prepare oneself for the trip (2000b:312)". Most of the examples of the *dāštan* construction with future time reference include motion verbs, most commonly the verb *raftan* 'go, leave'. These examples can be seen as vague or ambiguous with regard to futurate and proximative reading. In example 3:27, the event of leaving can be seen as extended, in which case we are dealing with an ongoing reading, or as punctual, i.e. the point at which the speaker leaves for Canada, in which case the proximative reading arises. This difference, however, has no practical consequence, and in a real-life situation the intention of the person uttering this sentence can be unspecified in this regard.

3:27 ‘The music man’, 01:30:27

dār-am *mi-r-am* *kānādā.*
have.PRS-1SG IPFV-go-1SG Canada
‘I am leaving for Canada.’

Thus, FOC ongoing uses with future time reference and proximative uses can be difficult to distinguish, but we can identify typical features for both: FOC ongoing uses with future time reference often include motion verbs or explicit future references and are typically intentional in that they include a preparatory phase, while proximative uses apply to events viewed as achievements and typically involve an unintentional situation that should be prevented.

3.4.2.4 Iterative ongoing use

The *dāštan* construction normally does not go very well with verbs or predicates that are interpreted as semelfactives. Yet, when semelfactive verbs are used as main verbs together with the *dāštan* construction, an iterative reading may rise if the event occurs several times. If the event occurs once, the proximative reading is available, in which case the event is viewed as an achievement rather than as a semelfactive. The sentence in 3:28 can mean either that the subject is hitting the object repeatedly or, in an alarming situation uttered as a warning perhaps, is about to hit the object. Also, in a slow motion picture, the reading that the subject is in the midst of the hitting is possible. A native speaker notes that the proximative use is better if the emphatic marker *dige* ‘other’ precedes the utterance. Notably, the iterative reading of hitting the object repeatedly is quite difficult to distinguish from the ongoing reading of ‘being in the process of hitting the object’, since both would refer to the same event: the subject is in the midst of hitting the object over and over. What then seems to happen in iterative ongoing uses is that the semelfactive event is interpreted as repeated, as being a durative atelic event, i.e. an activity. In the proximative reading, we can expect the subject to show signs of wanting to hurt the object, leading us to believe that (s)he is about to do so. Both alternatives constitute FOC uses. The uses with the *dāštan* construction differ from the use with the simple present in that the latter does not include the proximative reading but also includes the generic reading, i.e. that this is something that generally happens.

3:28 Semelfactive event

dār-e *mi-zan-e=š.*
have.PRS-3SG IPFV-hit.PRS-3SG=3PC.SG
‘(S)he is hitting her/him.’

In principle, repeated achievements in combination with the *dāštan* construction can also give rise to iterative ongoing readings if the event is perceived of as repeated, as in *dāre gol mikane* '(S)he is picking flowers'.

Only two sentences that are possibly iterative were found in the data. Both of them have additional items adding iterative meaning to the utterance and are different from 3:28 in this regard: 3:29a) contains *yek riz* 'over and over', and 3:29b) contains *haminjur* 'this way/repeatedly', which both multiply the event. In my view, similar to other DUR contexts, the *dāštan* construction has an intensifying effect on these events. Removing the *dāštan* verb would make the utterance less FOC and decrease the sense of engagement in the event. This does not mean, however, that it is only the *dāštan* construction that carries the sense of intensification in these utterances.

3:29 Iterative uses

a) 'I am a Mother', 35:06
yek riz *dār-e* *so'āl* *mi-kon-e.*
over.and.over have.PRS-3SG question IPFV-do.PRS-3SG
'She is asking questions over and over again.'

b) 'A Separation', 01:11:00
išun *az* *hamun* *aval* *dār-an* *haminjur*
they from that first have.PRS-3PL this.way/repeatedly

touhin *mi-kon-an.*
insult IPFV-do.PRS-3PL
'She is being (repeatedly) insulting from the very start.'

3.4.2.5 Absentive contexts

The *dāštan* construction may be used in examples such as 3:30.[69]

3:30 Q34:15 [BT]

[On the phone:
- Is Nima home right now?]
- *na, u* *dār-e* *varaġ* *bāzi* *mi-kon-e.*
- No, (s)he have.PRS-3SG card game IPFV-do.PRS-3SG
'No, he is (off) playing cards [as always].'

[69] An informant comments that it is odd to answer with a negation and then give the progressive. In her opinion, the absentive use works best if the location is given, or if it is clear to both speakers that the subject is always at a certain location when playing cards.

In some languages, there are specific patterns marking these contexts that are referred to as *absentives*. De Groot (2000:695) defines the absentive as "the grammatical expression of absence" involving the information in 3:31.

3:31 The information provided by absentives (de Groot 2000:696)

> i. The subject is not present;
> ii. The subject is involved in an activity indicated by the lexical verb;
> iii. Based on pragmatic knowledge, it is predictable how long the subject will be away, or there is an assumption about the period of time that the subject will be away;
> iv. The subject will return after a period of time.

Thus, in absentive contexts, there is a place at which the subject is expected to be but is not. Some languages, for instance, Swedish or Dutch, have dedicated constructions for the absentive, but in some languages, progressives are used in absentive contexts (de Groot 2000:694). Persian is such a language where the *dāštan* construction may be used also in absentive contexts, but does not have an absentive meaning in itself. What the *dāštan* construction expresses is not the absence of the subject, but rather that the subject is doing something *as we speak*, i.e. ongoingness at the reference time. The absentive reading is then acquired from the context or from other elements in the clause. This is quite different from the other uses discussed, which arise depending on the type of event to which the progressive applies or have to do with the nature of the reference point. This means that the uses in absentive contexts need not be seen as a type of peripheral use for the *dāštan* construction.

In Dutch, there is a dedicated absentive construction that differs from several progressive constructions in the language. While the progressive examples that are provided indicate that the event is ongoing at a FOC reference time, the absentive may also indicate that the subject is on its way to or from the event (de Groot 2000:702–703). This reading is not possible in Persian, i.e. even without the context in brackets, example 3:30 cannot mean that someone is on their way to go play cards.

3.5 Concluding Chapter 3

In this chapter, different uses of the *dāštan* construction were investigated. It was shown that, as expected, the pattern most often has FOC ongoing use. It occurs much more often in the present tense than in the past tense, as shown in the data from the movies. In the present tense, uses often refer to the engagement and/or busyness of the agentive subject in the event. At times, the *dāštan* construction is used in contexts having an emotive component, such as irritation, complaint or

hyperbolic reading. It was also shown that the *dāštan* construction occurs in DUR contexts and has proximative, futurate and iterative ongoing uses and occurs in absentive contexts as well. Although the simple present and past include the ongoing, it was shown that the use of the simple forms when denoting an ongoing event have less emphasis on the engagement of the subject in the ongoing activity in comparison to the progressive.

The uses of the *dāštan* construction that were presented differ from one another in fundamental ways. While the ongoing function of the construction arises with events viewed as activities and accomplishments, the proximative function arises with events viewed as achievements. These readings are then dependent on the event to which the *dāštan* construction applies. Both proximative and futurate uses were also shown to have certain characteristics. It was shown that proximative uses typically refer to events occurring in a near future, are often unintentional and expected to be prevented. At times, they are also hyperbolic expressions. Futurate uses most often occur with motion verbs and are intentional. The iterative ongoing uses were noted to be very similar to the ongoing uses with activities and achievements, the difference being that the event is a repeated instantaneous rather than a dynamic one. The use of the *dāštan* construction in absentive contexts was shown to simply mark ongoingness where the absentive reading is given by other elements. Also intensification as arising in DUR contexts was noted. I will conclude, then, that the *dāštan* construction has three main functions, FOC ongoing (which is the most frequent one), proximative and futurate. Often, the pattern is used in contexts that are pragmatically marked in some way.

In the upcoming chapter, I will suggest some explanations for the uses discussed here in cross-linguistic comparison.

4 Cross-linguistic comparison and explanations

Many scholars, often concerned with the English progressive, have attempted to explain the different functions, uses and readings that arise with progressive grams. I am not in a position to offer a solution to all problems discussed, nor will I discuss all the different suggestions or approaches in the literature. I will however present some thoughts on these issues with respect to some of the grams discussed in Chapters 2 and 3, with special focus on the *dāštan* construction. The discussion is based on the data from these chapters, previous research and additional data collection based on native speaker consulting. This chapter starts by discussing and providing some explanations for the different uses that arise when the *dāštan* construction applies to different types of events in 4.1 which is followed by a typological probe in which the possibility of such uses is investigated in some of the grams discussed in Chapter 2. Additional uses such as temporariness and subjective uses are discussed in section 4.3. Section 4.4 discusses the negation restriction of the *dāštan* construction from a cross-linguistic perspective.

4.1 *Dāštan* construction and event types

It is generally noted that a progressive gram views an event without regard to its endpoint. For example Dowty (1977:57), in discussing the *imperfective paradox* in English, points out that the English progressive entails a reading where the telos, i.e. endpoint, of accomplishments is merely a possible outcome. From a cross-linguistic point of view, Bertinetto & Delfitto (2000:193) note that telic events viewed imperfectively are "detelicized". Regarding the progressive, it has been noted that progressive grams "exclude limit-oriented readings" (Johanson 2000:57), or put differently, progressive grams make the telos potential rather than actual (Östen Dahl, p.c.). In utterances where progressives are used, as the focus is on one point in time at which the event is ongoing, the attention is turned away from any endpoint or goal or result included in the meaning of the predicate. Similar to a snapshot of a dynamic situation (Mittwoch 1988:233–234), the progressive gram views the event at one point without regarding its endpoints. This cross-linguistic explanation for the reading that arises when progressives

apply to events viewed as accomplishments is also applicable to the *dāštan* construction.

Achievements are characterized by being punctual but also typically being associated with an implicit stage leading up to the event. What then seems to happen with the *dāštan* construction in combination with events viewed as achievements is that the progressive gram detelicizes the endpoint, i.e. what would be the event itself if given without the progressive, while taking scope over the preparatory phase. Thus, in a sentence like 3:22a), '[Hurry up!] *ġatār dār-e mi-r-e!*' 'The train is about to leave!', the event to which the progressive refers is the preparation of the train leaving (e.g. the speaker announcement or people rushing to get on), while the actual leaving of the train is potentialized, meaning that its fulfillment may or may not be realized. This is what gives rise to the proximative reading of *being on the verge of happening*. In this sense, the *dāštan* construction can be said to make punctual achievements into events that are somewhat more extended. It seems as if the *dāštan* construction enables an interpretation where the punctual event becomes more stretched out, but not to the extent that it is viewed as an accomplishment. Also, typically semelfactive events can be viewed as achievements. Here, again, the *dāštan* construction refers to the preparatory phase.

In this way, the *dāštan* construction changes the way the verbal meaning is interpreted, which is most evident with telic verbs. This explanation however cannot explain why events can be viewed as, e.g., accomplishments given one context and as achievements given another. Surely, contextual information as well as knowledge about the world affects the interpretation of utterances, e.g. we know that we do not usually eat rotten apples or that in our part of the world, trains do not usually move while we board them.

Thus, the event type with which the *dāštan* construction combines determines the main reading of the utterance. In order to see whether this also applies to other progressives, a typological probe is conducted in the next section.

4.2 Typological probe

This section looks at the uses of a number of progressive grams with achievement and semelfactive events in order to see whether they give rise to similar readings as the *dāštan* construction. The ongoing use of these grams with activities and accomplishments is often uncontroversial, but less is known about the readings that arise when these grams combine with achievements and semelfactive events. Also, the futurate uses with motion verbs is discussed. In this section, if no other reference is given, the examples are provided by informants and glossed by either them or me.

The typological enquiry includes 9 languages and 10 progressive patterns given in Table 20. The inclusion of grams was dependent on the availability of informants. All these grams, as well as others, were also discussed in the study of the TED corpus in Chapter 2. The Turkish *-Iyor-* is also included here, although it is has been shown to be on its way to becoming an imperfective. One informant per language has been consulted, apart from German, where three informants were consulted. Examples capturing mainly achievements and semelfactives were taken from the PROGQ or created, and informants were asked whether the progressive in their language gives rise to the proximative and iterative readings with these events or not. The informants were either consulted via e-mail or in person, at times, additional events were also discussed as a result of the dialogue.

Table 20. Grams included in the typological probe

LANGUAGE	LANGUAGE FAMILY	GRAMS
English	Indo-European	*be* + GER
Finnish	Uralic	*olla* + V.NMLZ[70]
French	Indo-European	*être* + *en train de*+ INF
German	Indo-European	*sein* + *am* + INF
Italian	Indo-European	*stare* + GER
Spanish	Indo-European	*estar* + GER
Swedish	Indo-European	*hålla* + *på* + *att/och* + V
Thai	Tai-Kadai	*kamlaŋ*
Turkish	Turkic	*-Iyor-*
		-mAktA-

Not all grams with ongoing uses can apply to achievement events, for example *zài* in Mandarin Chinese (Sino-Tibetan) has previously been noted to not combine with achievements (Smith 1997:75). Among the grams discussed here, the progressive marker *kamlaŋ* in Thai (Tai-Kadai) and *-mAktA-* in Turkish (Turkic) do not have the proximative reading. The Thai informant only accepted the proximative construction in Thai for this function, the proximative being built on the progressive *kamlaŋ* plus the future marker *cà* plus a main verb (Smyth 2002:69). The Turkish informant only accepts ongoing or habitual uses with *-mAktA-*, i.e. not the proximative reading. The Finnish (Heinämäki 1995:144) as well as Turkish *-mAktA-* grams mainly occur in intentional contexts with agentive subjects. In addition, the Turkish *-mAktA-* is typically used in formal settings and not in spoken language, although Göksel & Kerslake comment that it may be used in informal speech 'where a speaker wishes to emphasize the intensity of the ongoing event' (Göksel & Kerslake 2005:332).

[70] Also referred to as the third infinitive.

Regarding the other grams, however, all informants accepted the proximative use to some degree given the right context. A few examples are given in 4:1. As can be seen, the ongoing reading is also available unless it is not possible given the context.

4:1 Proximative uses of progressive grams

a) Swedish (Indo-European)
[Hurry up!]
Tåg-et *håll-er* *på* *att* *åk-a.*
train-DET hold-PRS on INFM go-INF
'The train is about to leave/is leaving!'

b) Turkish (Turkic)
Öl-üyor-du.
die-PROG[71]-PST.3
'(S)he was dying/about to die.'

c) French (Indo-European)
La *pierre* *est* *en* *train* *de* *tomber.*
The stone be.PRS.3SG in train of fall.INF
'The stone is about to fall/is falling.'

d) Spanish (Indo-European)
Estaba *muri-endo-se* [but they found a medicine that cured him]
be.PST.3SG die-GER-3REFL
'He was about to die.'

e) Finnish (Uralic)
Hän *oli* *voitta-ma-ssa*.
3SG be.PST.3SG win-NMLZ-INE
'(S)he was about to win/was winning.'

Among those who accepted the proximative reading, the English and German informants were the most reluctant. For example, the German informants did not accept the proximative reading with an example such as *Die Kaffeetasse ist am fallen*/*kippen* when intended to mean 'The coffee mug is about to fall/tilt', only the FOC ongoing meaning was accepted, i.e. that the mug is already in the air falling towards the floor. In addition, they commented that the combination of the progressive with these verbs sounds odd. Similarly, the English informant only

[71] *-Iyor-* is glossed as PROG, following Kornfilt (1997), although it is a gram with uses towards the imperfective.

accepted the ongoing reading with *The glass is falling*. In other languages, such as, e.g., Persian and Swedish, both the proximative and the ongoing reading are available in such examples.

Johanson (2017:31) provides *The train is leaving* as an example of a proximative use for the English progressive. This example was accepted by the English informant as having proximative use, with the comment that it works better containing the additional future referent '*in two minutes*', as in '*Hurry up, the train is leaving in two minutes*!'. This example is vague with regard to proximative and near future reading, which was also noted in section 3.4.2.2 to be common for the *dāštan* construction. In all other attempts to combine the English progressive with an event viewed as an achievement, the event was interpreted as an accomplishment and rendered an ongoing reading.

For the German gram, the proximative reading is not completely unattested, both Ebert (2000:615) and Anthonissen et al. (2016:15–16) mention it. It is my impression however, that this use is rather restricted. After some consideration, one German informant accepted the example 4:2a) as constituting a proximative use, which was later also accepted by the second but rejected by a third informant.[72]

4:2 German (Indo-European)

a) [We have to save the bird from drowning in the petrol],
der *ist* *am* *Sterben*.
3SG be.3SG at die.INF
'it is about to die.'

b) Anthonissen et al. (2016:16)
"Sie *waren am Erfrieren und Verhungern*," sagt Kriminalinspektor Gösta Hellberg.
'"They were freezing and starving," says Detective Inspector Gösta Hellberg.'[73]

The examples provided by Anthonissen et al. (2016:15–16) of achievement verbs with the German progressive gram, in which the event containing the progressive is italicized, rendered some disagreement among the German informants: while the 'third informant' only accepted the example in 4:2b) as grammatical (and

[72] The 'third informant' was generally more critical of the proximative use than the others, which may or may not have something to do with his being from the Eifel region in Rhineland-Palatinate (but speaking standard German). The German gram has previously been shown to have special features in the Rhineland dialect, see, e.g., Ebert (2000).

[73] A better translation could be "'They were about to perish from cold and hunger', says Detective Inspector Gösta Hellberg".

rendering a proximative reading), the other two accepted the examples given but did not agree on whether they should be interpreted as progressive or proximative.

I will conclude then that the proximative reading with the English and German grams is restricted, perhaps somewhat more so in English than in German. In these languages then, the progressive grams typically do not apply to achievement events. It is noted that these two grams share this feature although they are on rather different ends of the grammaticalization cline of the progressive towards the imperfective. Their level of maturity is, e.g., reflected in their number of occurrences in the TED corpus presented in Chapter 2, where the English progressive has an overwhelming higher number of occurrences than the German progressive: 24 100 and 27, respectively.

Chapter 3 showed that, for the *dāštan* construction, both ongoing and proximative readings can be used hyperbolically to refer to an event that is about to happen. This is also noted for some of the grams discussed here. For example, the Turkish informant accepted the proximative reading of a stone that is about to fall with *-Iyor-* if it is uttered several times as a warning. The proximative use with *-Iyor-* is also noted by Kornfilt (1997:359). Interestingly, Kornfilt provides an example with a posture verb. Here, the proximative use is linked to the use of the dative, which makes the event dynamic, instead of the use of the locative (Agnes Korn, p.c.). In this sense, it is similar to the uses of the Persian *dāštan* construction with posture verbs discussed in Chapter 3, section 3.4.2.2.

4:3 Turkish (Turkic)

Düş-üyor *düş-üyor* *düş-üyor*!!
fall-PROG fall-PROG fall-PROG.3
'It is falling/about to fall!!'

4:4 Turkish (Turkic) (Kornfilt 1997:359)

Koltuğ-a *otur-uyor-um.*
armchair-DAT sit-PROG-1SG[74]
'I am sitting down in the armchair.' ("I am about to sit down in the armchair.")

In a similar manner, the Italian informant comments that the progressive can be used in an event if the situation is urgent or if something is happening suddenly, perhaps uttered as a warning when referring to a falling stone, or a car that is just about to hit someone. The rotting of apples from Q52, in which 17 Persian-speaking informants used the *dāštan* construction, can also be viewed as a hyperbolic use of a progressive utterance. The use of the progressive in this

[74] Orig. sit-PRS.PROG-1SG.

sentence was accepted by some informants, at times somewhat skeptically. The German and English informants did not accept this use. In 4:5, an example from Italian is given.

4:5 Italian (Indo-European)

[We must eat these apples,]
stanno and-ando a male.
be.PRS.3PL go-GER to bad
'they are going bad/rotting.'

Moving on to the iterative ongoing uses, informants were asked if their progressive gram could apply to sentences with repeated semelfactive events such as, e.g., *He is sneezing*. In Italian, Turkish (for *-Iyor-*) and English the 'over and over' reading was accepted. For English, the iterative use is already noted in Comrie (1976:42). In Spanish, Finnish and German, informants only accepted their grams with repeated semelfactives if limited duration was added. The Spanish informant accepted the progressive gram with 'sneeze', adding *mucho* 'a lot' to mean that these days he has been sneezing a lot. Similarly, the Finnish informant only accepted the iterative use of the progressive gram with the verb 'knock' if *koko ajan* 'all the time' was added to the sentence, and the German informants only accepted the progressive gram with *schlagen* 'hit' when *die ganze Zeit* 'all the time' was added.

4:6 Turkish (Turkic)

Hasan (çok) öksür-üyor.
Hasan a.lot cough-PROG.3
'Hasan is coughing (a lot).'

It then seems as if for some grams the use of the progressive is better if the duration of the event is explicitly given. In some languages, the progressive provides such a stretched out reading. It is, however, not my impression that the iterative uses are common, and they generally work best with semelfactives that naturally occur several times, such as 'knock', or 'hack'.

4:7 Iterative uses

a) Italian (Indo-European)
Sta buss-ando alla porta.
be.PRS.3SG knock-GER on.the door
'(S)he is knocking on the door.'

b) German (Indo-European)
Der *ist* *am* *Hacken*.
3SG be.PRS.3SG at chop.INF
'He is chopping.'

The Thai *kamlaŋ* stands out in that it cannot occur at all in examples such as *He is sneezing* or *Tom is hitting Jerry*, not even when adverbials such as *all day* are added, in fact, the informant notes that the progressive marker is incompatible with this type of adverbial. Similarly, the Turkish informant did not accept the use of *-mAktA-* with 'hit', 'sneeze' is also unacceptable since the gram requires an agentive subject.

In Chapter 3, section 3.4.2.3, the futurate uses of the *dāštan* construction were discussed in contexts such as the questionnaire sentences Q29 and Q09, the English translations are repeated here for clarity. What is of interest here are cases where progressives are used to refer to future time reference without additional future markers.

4:8 Futurate uses

a) Q09
If you come at 8 o'clock, I still COOK. [Come a little later, please.]

b) Q29
Maryam LEAVE tomorrow.

While the progressive in Q09 refers to an event that is ongoing at a future reference time, Q29 has future time reference, but the event is not interpreted as ongoing. The use in Q29 is then a further development than the use in Q09, as it no longer marks ongoingness. We may suspect that the progressive-future grams such as the Kisi (Niger-Congo) *-chō-* discussed in Chapter 2 are developments from uses such as in Q29.

The progressives discussed here differ in their ability to occur in these two sentences. Most of the futurate uses of the progressives discussed here are already discussed in Bertinetto (2000:560, 587–588), Tommola (2000:656, 669) and Ebert (2000:641, 645).[75] Building on their findings and adding informant data on the Turkish and Thai patterns as well as data from Persian, we can note that while Persian, English and Finnish progressives occur in both Q29 and Q09, French and the Turkish *-mAktA-* patterns cannot occur in either. The Thai pattern cannot be used in Q09 since it is restricted to the realis domain. In Q29, it is typically used with the future marker *cà*, but the informant notes that the future marker may perhaps be left out in spoken language if the event of leaving is certain. The

[75] In the original questionnaire, Q29 corresponds to S66 and Q09 to S83.

Italian, Spanish, Swedish and German patterns can occur in Q09 but not in Q29. The Turkish *-Iyor-* pattern is different from all the rest as it occurs in Q29 but cannot occur without the future marker in Q09.

4:9 Turkish (Turkic)

a) *Ann yarın gid-iyor.*
Ann tomorrow go-PROG.3
'Ann is leaving tomorrow.'

b) [If you come at 8 o'clock,]
hala yemek yap-ıyor ol-acağ-ım. [Come a little later, please.]
still food do-PROG be-FUT-1SG
'[If you come at 8 o'clock,] I will still be cooking. [Come a little later, please.]'

Bertinetto (2000:588) points out that it is difficult to understand how the future use in contexts such as Q29 has arisen. He notes that there is a connection between the proximative, which has a near future meaning, and the futurate use. A difference between futurate and proximative uses is that futurate uses often involve motion events, whereas proximative uses mainly involve events viewed as achievements that could indicate that these uses have different paths of developments. Motion verbs are a typical source for the development of future grams (Bybee et al. 1994:267). If futurate uses of progressives initially arise from motion events, it may well be that futurate ongoing uses have an explanation similar to the general grammaticalization of future grams evolving from movements. The source meaning of futures based on movements is suggested by Bybee et al. (1994:268) to be 'the agent is on a path moving toward a goal'. The explanation provided by Bybee et al. (1994:269) for futures derived from movements is that spatial movements entail movement in time.

> The temporal meaning that comes to dominate the semantics of the construction [future derived from a movement construction] is already present as an inference from the spatial meaning. When one moves along a path toward a goal in space, one also moves in time. The major change that takes place is the loss of the spatial meaning. (Bybee et al. 1994:269)

Another explanation could be that futurate uses of progressives are inherited from or arise by analogy to such uses of the simple present. In Persian, the simple present is often referred to as the non-past as it also has future function (see, e.g., Jahani 2008). The *dāštan* construction may then have developed such uses in analogy to the simple form, especially since the progressive builds on the simple forms.

Table 21 summarizes the results, also the *dāštan* construction is included in the table.

Table 21. Proximative, iterative and futurate uses

LANGUAGE	GRAM	PROX USE	ITER USE	FUTR USE
English	*be* + GER	yes?	yes	yes
Finnish	*olla* + V.NMLZ[76]	yes	yes, with DUR ADV	yes
French	*être* + *en train de* + INF	yes	yes	no
German	*sein* + *am* + INF	yes?	yes, with DUR ADV	yes (Q09)
Italian	*stare* + GER	yes	yes	yes (Q09)
Persian	*dāštan* + IPFV	yes	yes	yes
Spanish	*estar* + GER	yes	yes, with DUR ADV	yes (Q09)
Swedish	*hålla* + *på* + *att/och*	yes	yes	yes (Q09)
Thai	*kamlaŋ*	no	no	no (Q29?)
Turkish	*-Iyor-*	yes	yes	yes (Q29)
	-mAktA-	no	no	no

In summary, it seems that progressive grams, when they are able to combine with events viewed as achievements, typically give rise to proximative readings, and when they are able to combine with iterated events, give rise to an iterative ongoing reading. Some of the grams mentioned do not combine with achievements or semelfactives. The Thai *kamlaŋ* and Turkish *-mAktA-* are such grams. In English and German, the proximative use is very restricted. We can then say that in cases where an event is ambiguous as to whether it is an accomplishment or an achievement, these progressives disambiguate the event to be an accomplishment, rendering the ongoing reading only. The progressive grams also differ in their ability to occur in futurate uses. Most notably, many grams are not used in contexts such as Q29.

In the literature, other readings such as temporary readings, as well as various types of subjective readings are mentioned. We will now turn to these.

4.3 Temporariness and subjective uses

Many scholars have discussed and described the reading of *temporariness* for the English progressive, it is for example already mentioned in Jespersen (1924:279) who notes that the English "expanded form" denotes a transitory phase in contrast to a permanent state. Such a reading is also available in the examples given in 4:10 for Persian where b) can have a temporary sense while a) is more neutral. One

[76] Also referred to as the third infinitive.

consulting native speaker comments that the difference between the simple present in a) and the progressive in b) is not necessarily that big, but that the living in London in b) is perhaps somewhat more unexpected.

4:10 Permanent vs. temporary uses

a) *u* *tu* *landan* *zendegi* *mi-kon-e.*
(s)he in London life IPFV-do.PRS-3SG
'(S)he lives in London.'

b) *u* *dār-e* *tu* *landan* *zendegi* *mi-kon-e.*
(s)he have.PRS-3SG in London life IPFV-do.PRS-3SG
'(S)he is living in London.'

Temporariness is often discussed with verbs such as *live* and *stand* (Comrie 1976:37), but as noted by Ljung (1980:46), in English, temporariness is a more general feature of the progressive and is also a reading available with verbs such as *run*, *read* and *fall*. I understand temporariness as a reading that arises as a direct consequence of focality: a simple present *I run* is a generic statement in English (and is ambiguous for generic or ongoing reading in other languages such as, e.g., Persian or Swedish), the progressive *I am running* is FOC, which means that it refers to a specific event occurring right now. This in turn means that it will not go on forever. In this sense, progressive utterances typically refer to events that are temporarily valid. I believe that with stative verbs denoting location, such as *live* and *stand*, the lack of dynamicity, or the lack of possibility to change such verbs into more dynamic ones, results in that these verbs cannot be 'ongoing' in the same sense as *run*: the speaker cannot be in the midst of living or standing in the same way as (s)he can be in the midst of running, but the events of standing, living and running can occur at a punctual reference time. Understanding ongoingness as something that applies to dynamic events, we can say that the stative locative utterances are FOC, whereas *I am running* or *I am working* are both ongoing and FOC.

In Chapter 3, section 3.4.2.1, it was noted that the *dāštan* construction in DUR contexts has the implication that the subject is engaged in the event, and/or adds intensification to the event and/or has some sort of emotive effect. Interestingly, other progressive grams have been reported to also have similar readings in DUR contexts. The English progressive is noted to have emotive effect, as in *she's always buying far more vegetables than they can possibly eat* as opposed to *she always buys far more vegetables than they can possibly eat* (Comrie 1976:37). And De Wit et al. (2013:853) note that the English and French (*être en train de* + INF) present progressives carry notions of irritation and surprise when occurring in DUR and habitual contexts. Additionally, it is noted that the Finnish *olla* +

V.NMLZ progressive has an emphatic function, especially when used in perfect and pluperfect contexts (Tommola 2000:659).

In example 3:29 in Chapter 3, two utterances with subjective nuances were given. In these utterances, both *yek riz* 'over and over' and *hamintour* 'this way/repeatedly' add a sense of irriation. In these examples as well as others with subjective nuances, it is not easy to distinguish between subjective uses that arise due to the use of the progressive and subjective uses that arise due to other elements in the clause or in the context. Often, the impression is that it is the use of the progressive as well as other items that give rise to these readings.

In order to isolate the subjective nuance, the examples in 4:11 were constructed. Here, the progressive sentence in b) has a subjective nuance, as if referring to an atypical event or similar. As such, it can suggestively be used as a reproach or complaint or as an answer to someone assuming the opposite (e.g. *You never do anything around here*). The example in a), on the other hand, sounds more like a generic and objective statement. Thus, while example 4:11 is not FOC but only ongoing due to the explicit DUR context, example 4:10 is not ongoing but only FOC due to the verb semantics.

4:11 DUR contexts

a) *man* *har* *hafte* *lebās* *mi-šur-am.*
I every week clothes IPFV-wash.PRS-1SG
'I wash every week.'

b) *man* *har* *hafte* *dār-am* *lebās* *mi-šur-am.*
I every week have.PRS-1SG clothes IPFV-wash-1SG
'I am washing every week.'

FOC ongoing uses were also noted in Chapter 3 to have emotive effects with the *dāštan* construction. The link between the emotively charged FOC utterances and progressive grams may be straight-forward: if an event is marked as happening at one point in time, it is implicit that it does not happen generally or typically. In contexts that are explicitly not FOC, then, this pragmatic sense of atypicality may have triggered the use of this pattern. An alternative explanation would be that the sense of focality, which is typical for contexts in which the *dāštan* construction is used, is re-interpreted as intensification in explicitly non-FOC contexts. Whether or not these suggestions can be applied also to other languages is not clear and needs further investigation.

But not all DUR contexts have subjective readings. Consider the examples in 4:12 and 4:13. The Swedish simple past sentence in 4:12a) has a more objective sense than the sentence containing the past progressive in b). The progressive sentence can perhaps be uttered in opposing a statement implying that the subject has not done anything or similar. A Swedish informant comments that the

sentence containing the progressive emphasizes the subject's occupation with the event in comparison to the use of the simple form.

4:12 Swedish DUR contexts

a) *Han hacka-de lök hela dag-en.*
he chop-INF onion all day-DET
'He chopped onions all day.'

b) *Han höll på att hack-a lök hela dagen.*
he hold on INFM chop-INF onion all day-DET
'He was chopping onions all day.'

We can have a similar analysis of the Persian examples in 4:13: while 4:13a) is a neutral utterance that states what the subject does between one and two, b) has a sense of marking the engagement of the subject in the event: between one and two, the subject is busy working. A consulting informant notes that this sentence only works in specific contexts, as an answer to the question of what one does between 1 and 2 for example. In c), the event is FOC ongoing: at this very moment, the subject is in the midst of working. In my opinion, the Swedish and Persian examples differ in that the Swedish progressive emphasizes the engagement of the agent in the event to a greater extent than the *dāštan* progressive.

4:13 Persian, DUR vs. FOC contexts

a) *beyn=e sā'at=e yek-o do kār mi-kon-am.*
between=EZ hour=EZ one-and two work IPFV-do.PRS-1SG
'I work between 1 and 2 o'clock.'

b) *beyn=e sā'at=e yek-o do dār-am*
between=EZ hour=EZ one-and two have.PRS-1SG

kār mi-kon-am.
work IPFV-do.PRS-1SG
'I am working between 1 and 2 o'clock.'

c) *dār-am kār mi-kon-am.*
have.PRS-1SG work IPFV-do.PRS-1SG
'I am working.'

Anthonissen et al. (2016) note that the German progressive can have a sense of *evasiveness*, one German example is given in 4:14b). One example of an evasive reading is also found in the data from Chapter 3, given in 4:14a), which is uttered

by the subject as she is visited in prison in referring to her situation. What can be noted here is that in Persian, the simple present counterpart to the progressive has future time reference rather than present time reference, i.e. it means *I will get used to it*, while the *dāštan* construction has the meaning *I am getting used to it*. This shows again that the counterpart of the progressive is not always the simple present or past imperfective. Similar to the English *I will get used to it*, the simple present with future time reference in Persian is a more definite statement, while the progressive utterance has a less certain reading. I suggest that this has to do with the detelicization of the event expressed with the *dāštan* progressive: the simple present with future reading gives a more certain reading since it also includes the endpoint, i.e. the state where the subject has become used to her situation, while the progressive utterance gives no insurance of whether the endpoint will be reached or not. The German example is different since the event used is atelic. Anthonissen et al. (2016:22) note that the use of the simple present would sound harsher and more definite, and suggest that these evasive readings can be explained through the "contingent quality" of the progressive in which "the speaker is less committed to the full realization of this situation". The verb *überlegen* 'to consider, to think over' is one of the most common verbs found with evasive reading in their data.

4:14 Evasiveness

a) 'I am a mother', 01:04:20
hameči *xub=e,* *dār-am* *ādat* *mi-kon-am.*
everything good=COP.3SG have.PRS-1SG habit IPFV-do.PRS-1SG
'Everything is fine, I am getting used to it.'

b) German, (Anthonissen et al. 2016:22)
Anker […] will sich […] nicht in die Karten schauen lassen.
"Wir *sind am Überlegen*," so Schuster zurückhaltend.
'Anker […] is playing its cards close to its chest. "We are thinking about it," said Schuster, aloof.'

A short and preliminary comment is made here on the use of the *dāštan* construction with typically performative predicates, since such cases were discussed in Chapter 2, although no such examples were found in the data of Chapter 3. The *dāštan* construction may occur in such cases but has a sense of being uttered counter to what the listener believes. This means that it is a comment on what has previously been said rather than a performative. Naturally occurring data is needed to investigate these uses further.

4:15 'To promise'

a) *be to ġoul mi-d-am.*
to you promise IPFV-give.PRS-1SG
'I promise you.'

b) dār-am be to ġoul mi-d-am.
have.PRS-1SG to you promise IPFV-give.PRS-1SG
'I am promising you.'

In Chapter 2, it was shown that posture verbs were the most common type of statives with which progressive grams combine. The Turkish (Turkic) *-Iyor-* pattern, which occurred more than three times as often as the English progressive in the TED corpus in Chapter 2, may also combine with stative verbs such as 'know' and 'love'. The simple present is, however, also still in use, verbs like 'know' and 'love' can, for example, occur with both the simple present form and with *-Iyor-*, but the uses differ from one another. While the sentence *Herşeyi bili**yor*** '(s)he knows everything' can refer to both a specific matter and have generic meaning ('because (s)he is smart'), the simple present *Herşeyi bilir* '(s)he knows everything' only has generic sense. With 'love', the simple form and the *-Iyor-* form seem to have different meanings: while *onu sevi**yor*** means '(S)he loves her/him', *onu sever* instead means '(s)he likes him/her (as a person)' (Hatice Zora, p.c.). An example of this is given in 4:16 from the movie Son Hıçkırık (1971). In this classical love scene, the man pronounces his love to the woman on the night on which she is getting married to another man, to which she answers that she likes him too, but as a sister loves a brother. As shown, his utterance contains *-Iyor-*, while hers is in the simple present.

4:16 Turkish (from the movie Son Hıçkırık, at approximately 16 minutes)

- *Sev-iyor-um sen-i.*
love-PROG-1SG you-ACC

- *Ben de sen-i sev-er-im.*
I too you-ACC love-AOR-1SG

Kardeş-in=im sen-in.
younger.sibling-POSS.2SG=COP.1SG you-GEN

'- I love you.
- I like you, too. I am your younger sister.'

The combination of the English progressive with stative predicates such as 'be angry' or 'be polite' has certain subjective nuances. Ljung (1980:42) notes that such stative predicates have "covert properties", i.e. non-observable properties, whereas stative predicates with "overt properties" such as 'be tall' and 'be short' do not combine with the progressive, at least not as often. In discussing examples such as *John is being angry* and *John is being polite*, Ljung (1980:40) refers to the feature of "dynamicness" and points out that these utterances, which require an agentive subject, say something about the behavior of John, a behavior that is displayed and observable. This means that the meaning of *John is being angry* can be said to be 'John is behaving in an angry way right now' and the meaning of *John is being polite* can be said to be that John "displays- 'acts'- politeness at the moment of speech" (Ljung 1980:41). In both cases, the sentence with the progressive is saying something about the behavior of the person referred to, i.e. something that they *do*. The reason why the progressive combines more easily with stative predicates that have "covert properties" could perhaps have to do with the ease of making these stative predicates more dynamic; it is more uncommon, as well as difficult, to behave in a tall way than in an angry way. These examples suggest, that the English progressive, on the path of grammaticalizing into a general imperfective, first combines with stative predicates that can be interpreted dynamically. In this process, the English progressive requires an agentive subject – similar to the general grammaticalization path of locative to progressive as proposed by Bybee et al. (1994:136).

De Wit et al. (2013), De Wit & Brisard (2014) and Anthonissen et al. (2016) show that (present) progressive grams in English, French and German are often, but not always, used in contexts that have subjective nuances, such as contexts where the speaker expresses surprise, irritation, emphasis, or events that have an "atypical status". They show that the replacement of the progressive with the simple present tense decreases such notions. Due to this, they suggest the "core meaning" of these grams to be "epistemic contingency". As pointed out in Chapter 1, I understand their use of "epistemic contingency" to include a range of uses, such as ongoing, habitual, temporary and subjective readings. Putting all these together, in my view, "epistemic contingency" includes all uses that a simple present has except for the generic ones. The findings of the *dāštan* progressive in Persian point towards this pattern also being used in contexts with subjective nuances, I have talked of these contexts as having emotive components. In my opinion, however, the emotive readings cannot be assigned to the use of the progressive only, but are often accompanied by tone, stress or other components in the clause. Thus, emotive components seem to increase the likelihood that the *dāštan* progressive is used, but is not necessarily part of the meaning of the gram.

In my opinion, focalization and dynamicness may at times be enough in explaining emotive readings of the English progressive. Anthonissen et al. (2016:20) discuss subjective uses of the English progressive in similar examples as Ljung (1980). They note that while a sentence such as *John is silly*, meaning

'John is a silly person' has a more neutral sense, a sentence such as *John is being silly*, meaning 'John is behaving as a silly person at the moment' has "emotional overtones of irritation" (Anthonissen et al. 2016:19). They explain these subjective readings as arising due to the "core meaning" of "epistemic contingency" of the English progressive. Although I agree that the uses of progressive grams exclude generic readings, I believe that the observed subjective nuance in these examples are not necessarily part of the progressive meaning, but rather likely uses or interpretations of 'John is behaving as a silly person at the moment' or 'John is behaving as a polite person at the moment'. In these examples, the use of the progressive enables a more dynamic interpretation of the stative predicate, but does not add subjectivity to the utterance as such. In theory, it may be possible that subjective nuances with time become associated with the use of a progressive in certain contexts. However, it is not clear to me that this has happened in English.

What is more, in my opinion, the suggested "core meaning" of "epistemic contingency" in the English, French and German (present) progressive grams and possibly also other progressives, is too general. The uses discussed by the authors include the meaning of ongoing, habitual, temporariness and intensification as well as various types of subjective uses (e.g. irritation, surprise and evasiveness). As far as I can tell, this means that the notion of "epistemic contingency" will need further specification in each specific type. Or put differently: even though it is true that the different uses found with progressives cross-linguistically are all non-generic, the assignment of non-genericness as the "core meaning" of the progressive will not explain why and how it is that the progressive gram gives rise to the reading of 'ongoing at the reference time' in most contexts, and temporariness, proximative, subjective uses, etc. in other contexts. I agree that the exclusion of generic meaning is an important description of progressive grams. But it is not the only or main meaning that the progressive contributes to the utterance. In fact, in the majority of cases, the progressive is simply used to refer to an ongoing event at the reference time. This use was also shown above to directly or indirectly explain those temporary and subjective uses discussed here.

What is more, I suggest that the meaning that the progressive gram adds to the utterance is dependent on the level of maturation of the gram. Looking at the number of occurrences of progressives in the TED corpus, we see that the English progressive occurs 24 100 times, the French progressive occurs 898 times and the German progressive 27 times. We must then expect the English progressive to be much more general in its use with a much less specific and homogeneous meaning than the German and French progressives. Especially when talking of progressive grams with expanded uses, it may no longer be possible to talk of only one core meaning, unless this meaning is very general (as in non-genericness). Such a general core meaning is in turn not concrete enough for describing a gram at the beginning of a grammaticalization process.

Finally, I will comment on the fact that the data from the *dāštan* construction as well as other studies presented in this chapter indicate that the subjective readings arise in contexts where usually the simple forms are used, i.e. where the use of the progressive gram is not very common, such as in DUR contexts, with stative verbs, in performative contexts and so on. Perhaps then, it is in new, expanded uses that these subjective readings arise. This would also explain why we find such uses in a highly grammaticalized gram such as the English progressive as well as in less grammatical grams such as the German progressive. This impression however, needs further investigation. As the grammatical shift from progressive to imperfective takes place, we can expect any subjective readings to be lost, a process that is referred to as *rhetorical devaluation*.

> In my opinion, an essential characteristic of grammatical maturation is that the pattern spread leads to a decrease in the rhetorical and/or informational value of the pattern or its component expressions – what I call rhetorical devaluation. (Dahl 2004:121)

Most subjective uses discussed are in the present tense. It is unclear whether subjective uses are restricted to or mainly found in the present and to what extent they exist in the past.

4.4 The negation restriction of the *dāštan* gram in cross-linguistic comparison

It was already mentioned that the *dāštan* construction is typically not negated. In the data from five Iranian movies in Chapter 3, one case of a negated *dāštan* construction is found, nevertheless. In this example, the wife hears her husband talking on the phone about someone dying. As she asks about it, she uses an utterance containing the *dāštan* construction. He dismisses her beliefs by repeating a negated form of her utterance. The example is given in 4:17. Thus, in example 4:17, the speaker objects to the presupposition that someone is dying. This resembles what in the literature has been called *metalinguistic negation*. According to Horn (1985:121), metalinguistic negation is not a logical operator but rather "a metalinguistic device for registering objection to a previous utterance (not proposition) on any grounds whatever, including the way it was pronounced". It then follows, that in metalinguistic negation, usual restrictions on what can be negated do not necessarily apply. Interestingly, a somewhat similar but perhaps more conventionalized situation is reported for the progressive construction in the Iranian language Tajiki formed with the past participle + *istodan* 'to stand, be standing', which can be negated when contradicting an assertion, although such uses are not very frequent (Perry 2005:225).

4:17 ‘Felicity land’, 23:54

- ki	*dār-e*	*mi-mir-e?*		
who	have.PRS-3SG	IPFV-die.PRS-3SG?		
- yek	*ġavās*	*hāl=eš*	*bad=e,*	*hamin,*
one	diver	state=3PC.SG	bad=COP.3SG	that’s.it
kasi	*dār-e*	*ne-mi-mir-e.*		
someone	have.PRS-3SG	NEG-IPFV-die.PRS-3SG		

‘- Who is dying?
- One diver is ill, that’s all, no one is dying.’

Since negation of the *dāštan* construction is typically not possible, one may claim that to negate a *dāštan* construction one uses the negated simple present or negated imperfective past. Thus, the negated form of ‘I am eating’ using the *dāštan* construction can be said to be ‘I don’t eat’, given in example 4:18a), creating an asymmetry in the language. In comparison, there is no restriction on negating the *dar hāl=e* ‘in state of’ progressive nor the *mašġul* ‘busy’ construction where the negative copula is available, as seen in b) and c).

4:18 Negated form *dāštan*, *dar hāl=e* and *mašġul* constructions.

a) *man* *ġazā* *ne-mi-xor-am.*
I food NEG-IPFV-eat.PRS-1SG
‘I don’t eat.’

b) *man* *dar* *hāl=e* *ġazā* *xordan* *nist-am.*
I in state=EZ food eat.INF NEG.COP.PRS-1SG
’I am not eating’

c) *man* *mašġul=e* *ġazā* *xordan* *nist-am.*
I busy=EZ food eat.INF NEG.COP.PRS-1SG
‘I am not eating.’

The restriction of negation on the *dāštan* construction is not morphologically motivated since the negation prefix is neither inhibited from *dāštan* nor from the main verb. For some reason, 4:19a), where the *dāštan* verb is negated, sounds somewhat more ill-formed than negating the main verb in b). In 4:19 both sentences are intended to mean *I am not eating*. Note that the negation marker in example 4:17 is also on the main verb.

4:19 Ill-formed negated sentences with the *dāštan* construction

a) *man* ***na****-dār-am ġazā mi-xor-am*.

b) *man dār-am ġazā* ***ne****-mi-xor-am*.

Negation asymmetries involving progressive patterns have been attested in Miestamo (2005) and Miestamo & van der Auwera (2011:71), where languages such as Indonesian (Austronesian), Cantonese (Sino-Tibetan), Kabardian (Northwest Caucasian), Nasioi (East Bougainville), Ogbronuagum (Niger-Congo) and Uzbek (Turkic) have negation asymmetries in relation to their progressive patterns. Adding to this list, the Mazandarani (Indo-European) progressive formed with the locative copula *da*(*r*)- is typically not negated. This language and other Iranian languages are discussed in Chapter 5. There are, however, many languages that allow progressives to be negated, English being one obvious example. According to Miestamo (2005), symmetric negation is the outcome of language internal analogy where the negated structures copy the affirmative structures. Asymmetric negation, on the other hand, is functionally motivated. Miestamo & van der Auwera (2011:76) suggest a general motivation for the restriction on negation in interaction with imperfective and perfective aspect: negation is expressed in contexts where the affirmative is under discussion (*Oh, my wife is not pregnant* would sound odd if no one assumed that she was), meaning that negation is typically not used for bringing new information to the discourse. In such contexts, categories such as tense, aspect, person, etc. do not need to be specified. Explanations for why some languages show such restrictions while other do not have to be investigated for each language. Applying this explanation to the *dāštan* construction, and possibly also other progressives, it could perhaps be said that for something that did not happen, the specification that the event was ongoing at the reference time is excessive.

4.5 Summarizing Chapter 4

In this chapter, the uses of several progressive grams with different types of events were investigated. It was shown that progressive grams when applied to different types of events result in different uses and readings. A typological probe as well as findings in previous studies suggests that it is possible that cross-linguistically, progressives, if possible to apply to events viewed as activities, accomplishments, achievements and semelfactives, will have the uses as given in Table 22. The progressive as applied to events viewed as activities have the meaning of ‘ongoing at reference time’. When the progressive applies to events viewed as accomplishments, in addition to, or rather as a consequence of, the meaning of

'ongoing at the reference time', telos is detelicized. Events viewed as achievements, which can be seen as consisting of change with an implicit pre-stage, will give rise to proximative readings as the progressive focuses on the pre-stage and makes the event itself potential rather than actual. In this sense, then, the event is no longer viewed as instantaneous but is viewed as somewhat more stretched out. Some grams, such as the English and German progressives seem more restricted in rendering the proximative reading as they typically cannot apply to achievements. Iterated events where several punctual events are repeated will render a use similar to that of activities: the repeated event is ongoing at the reference time. These uses can all be explainable through the basic assumption that progressives refer to an ongoing event at the reference time, in doing so, the attention is turned away from any potential endpoint.

Table 22. Hypothesis of uses that arise when progressives apply to different types of events

EVENT VIEWED AS	TYPE OF USE	MEANING
activity	ongoing	Ongoing at the reference time.
accomplishment	ongoing	Ongoing at the reference time. Telos of event deactualized.
achievement	proximative	The (whole) event deactualized, focus on the pre-stage leading up to the event.
repeated punctual (semelfactives or achievements)	iterated ongoing	The repeated event is ongoing at the reference time.

The progressive grams were noted to differ in their ability to have futurate uses. The type of futurate uses where most progressives showed restrictions were uses where the progressive grams function as a marker of future rather than of ongoingness, such as in *I am lecturing tomorrow*.

Several subjective and temporary uses were also discussed. Most of these could be directly or indirectly derived from the meaning of FOC ongoing: the marking that something is happening at one point in time implies that it does not generally happen. This implication can give rise to both temporary as well as senses of atypicality. The chapter also suggests that due to the difference in level of grammaticalization of progressive grams cross-linguistically, all progressives may not have one and the same "core meaning".

Stative situations are not included in Table 22. It has already been noted that the English progressive in combination with stative predicates may give rise to

subjective interpretations. Some subjective nuances of mainly the English, but also other, progressives arise in contexts not typical for the progressive gram, such as uses with stative verbs, verbs expressing location and in DUR contexts. Whether this observation holds cross-linguistically is something that I leave for future research.

5 Caspian progressives in contact

5.1 Introduction

This chapter deals with progressive patterns found around the Caspian Sea in northern Iran and southeastern Azerbaijan. The main languages discussed are Mazandarani, Gilaki, Tati and Taleshi, but other languages in contact with these are also mentioned. The present tense and past imperfective patterns are also investigated for these languages. This chapter investigates the patterns found in these varieties in contact situations both structurally and functionally.

The chapter starts with an introduction in section 5.1, followed by the presentation of the main languages discussed in this chapter in section 5.2. The presentation and discussion of data is divided into two parts. The first part, section 5.3, presents and discusses two progressive gram families in Mazandarani, Gilaki, Tati and Taleshi, which are referred to as the DAR and KAR gram families, with some patterns constituting fusions between the two. The second part, section 5.4, presents a comparison between the progressive grams found among these varieties and the present tense and the past imperfective. In this section, the assumption that there are more periphrastic patterns than inflectional ones is investigated by comparing the present progressive and past progressive patterns to present tense and past imperfective patterns. Section 5.5 concludes this chapter.

The data has been collected from language descriptions. Three forthcoming texts need to be mentioned here: Stilo (in press), which looks at progressives in which the division between what in this work is referred to as the DAR and KAR progressives and the mixing of the two is presented; Stilo (forthcoming a.), which is a survey of Mazandarani; and Stilo (forthcoming b.), which deals with gender in the Tati varieties Kafteji and Kelasi.

5.1.1 Two gram familes

In this chapter, certain patterns found in the Northwestern Iranian varieties spoken in 50 villages or cities in northern Iran and southeastern Azerbaijan will be referred to as the DAR and KAR *gram families*, as defined in Chapter 1, section 1.2.1.1. In some varieties, DAR and KAR patterns are combined. Generally, these patterns can be regarded as belonging to the progressive gram type, but as will be seen, some of these patterns are better analyzed as present tenses or, at

times, also proximatives. The two gram families are found in the languages Mazandarani, Gilaki, Tati and Taleshi, to my knowledge, patterns that can be regarded as belonging to the DAR and KAR gram families are not found elsewhere.

Postulating two gram families would mean, following the definition in Chapter 1, that the grams in each family are either borrowed from one another or inherited from a common proto-language. However, I will use the notion of gram family in a less strict sense, referring to a possibility rather than a confirmed relation. This precaution is due to the fact that the investigation is based on synchronic data and on structural and phonological similarities as well as on geographic closeness of the varieties.

Examples of DAR and KAR constructions are given in 5:1 and 5:2, respectively.

5:1 Rashti, Gilaki (Rastorgueva et al. 2012:336)[77]

či	*kud-ən*	*dər-i?*
what	do.PST-INF	DAR.PRS-2SG

'What are you doing?'

5:2 Kelasi, Tati (Stilo forthcoming b.)

kǽræ	*me-šé-m*	*baɣ.*
KAR	IPFV-go.PST-1SG	garden[78]

'I was going to the garden'

The constructions that are members of the DAR gram family involve an element that is phonologically close to *dar*/*dər*/*da* and that is identical in shape to an element that has a locative meaning, but the specific function of which may vary. In Mazandarani and Gilaki, for example, the locative DAR element is a locative copula, whereas in Tati it is described as a postposition with locative meaning used in various functions (Yarshater 1969:119–120, 125–126, 128–130). In Taleshi varieties, the elements are described as locative suffixes (Paul 2011:114). In order to better understand the erstwhile locative markers used in the progressive constructions, examples of non-progressive locative sentences will also be provided. Despite the similarity in form, the *dāštan* construction is not included in this chapter, the possibility of influence between Persian and the Caspian varieties is instead discussed in Chapter 6.

In a couple of cases, the DAR element in the DAR construction is realized as =*na*= or -*u*-. Paul (2011:114), referring to a forthcoming publication of Stilo, states that the locative suffixes -*da*- and -*na*- originate from -*anda*. In Noorlander

[77] The progressives in the appendix in Rastorgueva et al. (2012) have all been glossed as 'have.AUX', but the element is clearly the locative copula (Rastorgueva et al. 2012:144–145).
[78] Orig. PROG DU-went-1S garden.

& Stilo (2015:442), both *dœ* and *dar* are stated to derive from older Iranian **andar* 'in'. Interestingly, in the Harzani [T] variety we find a locative element *œndœ*, which is often reduced to *-d-* or *-de-*, used in the DAR construction (Stilo in press). Due to these circumstances, all the above markers are hypothesized to be cognates with a locative meaning and are glossed as DAR in the present chapter.

There is one locative marker used in a progressive pattern that is phonologically different from the rest and is therefore not regarded as a cognate, namely the Chali [T] *-u-* 'from, in with'.[79] The constructional schema of the progressive pattern, including the locative *-u-*, is V-INF-*u*-COP. This schema is identical or almost identical to other DAR patterns. Therefore, it is assumed to have arisen through contact with these varieties and consequently to be part of the DAR gram family. It is noted, however, that the glossing of *-u-* as DAR is not optimal.

The KAR gram family involve elements that are realized as *kar*/*kər*/*kərə* or similar. These elements likely originate from *kă̆r* 'work, doing' (Windfuhr 1989a:256). In Rashti [G], given in 5:1, the present progressive construction is formed with the combination of the infinitive verb and a locative copula *dər* in the present tense and person and number marking. The infinitive is formed through the addition of the infinitival marker *-ən* to the past stem. In Kelasi [T], given in 5:2, the past progressive is formed with the marker *kɛ́rœ* together with the past imperfective construction, which is formed through the addition of the imperfective prefix *me-* to the past stem with person number marking.

In this chapter, I will use 'the DAR/KAR constructions' for the whole construction, and 'the DAR/KAR elements' for the elements in the construction most often realized as *dar*/*dər*/*da* and *kar*/*kər*/*kərə*, respectively.

5.1.2 Method and practical issues

For this chapter, grammatical descriptions of the various Caspian varieties were consulted. In a couple of cases, data was collected using the PROGQ (see Chapter 3). Also, relevant Turkic, Neo-Aramaic (Afro-Asiatic) and Nakh-Daghestanian varieties were included for comparison. In Appendix E, a list of all varieties discussed in this section is provided together with the place where they are spoken, whether they have a DAR or a KAR pattern, and references. When using sources from different authors, inconsistencies in, e.g., terminology and definitions have to be dealt with. In some cases, authors do not provide glossed examples, in which case I have glossed and/or translated the sentences following the information given in the grammatical description. In many varieties, there are different sets of personal endings but not all authors have chosen to show the different sets in the glossings, and in cases where the glossing are mine I have also chosen not to identify the sets in the glossings. Some authors also consistently show the difference between clitics and inflectional markers while others do not.

[79] In Chali, locative -(*e*)*ndu* 'in, with' is a separate postposition (Yarshater 1969:119).

The morpheme boundaries glossed by me are all glossed as inflectional, unless there are explicit statements in the grammatical description that they are clitical boundaries.

Authors often vary in their glossings of the DAR and KAR elements, using LOC, 'be_in', PROG and the like. In order to be consistent, elements that are regarded as belonging to a DAR construction within the DAR gram family have been glossed as DAR, likewise, elements that are regarded as belonging to the KAR construction within the KAR gram family have been glossed as KAR. In such cases, if a glossing is provided by the author, the original glossing is given in a footnote. As a rule, for the rest of the sentence, glossings provided by authors are kept.

In this chapter, several maps will be used presenting the data. There are, however, several problems with maps especially relevant when dealing with areal linguistics, some of which will be mentioned here. First, they give the impression that linguistic features belong to a certain place rather than being part of speaker's language knowledge, speakers who are mobile, multilingual, part of a language-shift process and so on. Second, it gives the impression of the existence of sharp boundaries between linguistic features, which is not the case. Third, only the available data can be presented. What is more, only cases where the search was successful are shown. If no data is given for a city or village, this could indicate that there is no data available or that that area was searched but no such feature was found. When searching in descriptions for Mazandarani, Gilaki, Tati and Taleshi varieties for progressives, however, only one variety was found for which there was an explicit claim that there exists no DAR or KAR (or other relevant) pattern, namely the Tati variety Vafsi (Stilo in press).

5.2 Mazandarani, Gilaki, Tati and Taleshi

In this section, the classification, sociolinguistic situation and verb forms in the Mazandarani, Gilaki, Tati and Taleshi varieties are presented. The Mazandarani varieties are mainly found in the province of Mazandaran, the Gilaki varieties in the province of Gilan, the Tati varieties where the provinces of Zanjan, Gilan and Ardabil meet and the Taleshi varieties in Gilan, Ardebil and the bordering provinces of Azerbaijan. The location of the Iranian varieties that are introduced in this chapter are given in Appendix F. Figure 7 provides an overview of the relevant provinces.

As is customary in Iranian linguistics, the varieties discussed here may be referred to as either 'variety of X spoken in Y', or by the name of the village or city with an additional *-i*. Thus, the Mazandarani variety spoken in the village of Ziarat may also be referred to as *Ziarati*. This is not a claim that Ziarati is a language on its own, simply that there is a location, Ziarat, where a certain variety is spoken. A special case is the Mazandarani dialect spoken in Sari, which is

referred to as *Saravi*. In what follows, the term *variety* will be used to refer to a language as spoken in a specific place, but will also be used more generally in order to avoid having to distinguish between language, dialect or subgroup of dialect. The term language will rather refer to the whole group of varieties that go under a specific language name, such as, e.g., Taleshi, when this is relevant.

Figure 7. Provinces of northern Iran

5.2.1.1 Classification

The classification of the varieties discussed here is given in Figure 8. The figure is adopted but simplified from Glottolog[80], whose sub-classification is based on Stilo (1981). According to this classification, all the 50 varieties having DAR and KAR patterns belong to the Northwestern branch of the Iranian language genus. As can be seen, Mazandarani and Gilaki fall under the Caspian sub-group. Stilo (1981), who discusses the relationship between Taleshi and Tati, groups Taleshi as Tatic (e.g. Ethnologue groups Tati varieties under Taleshi). Notably, Glottolog does not have the subgroup Central Caspian (see below). Persian and Tat (which will be discussed in section 5.3.4.2) are given for comparison, as can be seen, they are classified as Southwestern languages.

It should be noted that this classification differs from more traditional classifications. In addition, the traditional division of Iranian into a Western and an Eastern Iranian branch as well as that of Western Iranian into a Northern and Southern sub-branch has also been questioned (see, e.g., Korn 2016). However, since many of the varieties discussed here are from data provided by Don Stilo, with in fact some varieties being only mentioned in his work, I will follow his classification, which assumes a Northwestern branch.

[80] http://glottolog.org/resource/languoid/id/iran1269

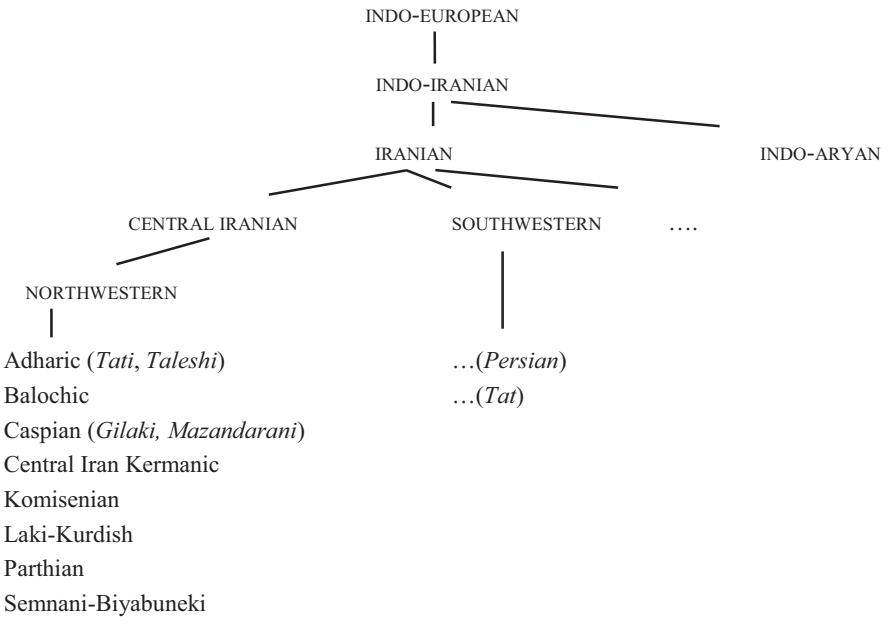


Figure 8. The Northwestern Iranian branch according to Glottolog

Borjian (2004:295) notes that there are transitional varieties between Gilaki and Mazandarani that are better understood as a separate, intermediate group. Stilo (forthcoming a.) refers to this group as Central Caspian. In this sample, three varieties of this kind are included and discussed in sections 5.3.1.1 and 5.3.1.2. Stilo (2008:364) also points out that the branches of the Northwestern Iranian languages originated out of one or several Old Iranian dialects that were already different from Old Persian, a Southwestern Iranian language. This split is likely to have taken place approximately 2800 years ago.

When referring to specific varieties of Mazandarani, Gilaki, Taleshi or Tati, an abbreviation of the language to which it belongs is given in square brackets: [M] for Mazandarani, [G] for Gilaki, [TAL] for Taleshi and [T] for Tati. Thus, Rashti [G] refers to the Gilaki variety spoken in the city of Rasht. Apart from these four main languages, a few varieties classified as Central Caspian and Semnanic are also included among the 50 varieties with DAR and KAR patterns.

A few varieties are classified as 'Tatoid' by Stilo and are therefore given as [Tatoid]. *Tatoid* is explained in the following way:

> [Tatoid refers to] languages of the Tatic family which, under the influence of other groups with which they are in heavy contact, have lost all the characteristic morphology of Tatic languages […] and have retained only the (more or less) original Tatic lexical composition. (Stilo forthcoming b.)

The varieties Kalasuri [T/TAL] and Xoynarudi [T/TAL] will be classified as both Tati and Taleshi since Yarshater (2005:269) notes that they may be seen as varieties of Northern Taleshi or a bridge between Taleshi and Tati, although speakers of these varieties view themselves as Tati.

In what follows, the varieties discussed will be referred to as Mazandarani, Gilaki, Tati, Taleshi and, at times, also Central Caspian, Semnanic and Tatoid, following the descriptions.

5.2.1.2 Sociolinguistic situation

This section briefly discusses the sociolinguistic situation for Mazandarani, Gilaki and Taleshi/Tati in turn. In Chapter 1, section 1.2.3, several sociolinguistic parameters in language contact that are crucial for the process and outcome of language contact were presented. As will be apparent below, we only have restricted information on some of these issues.

The map in Figure 9 shows the population density in the area where the most densely populated areas are, apart from Tehran, the cities and surroundings of Baku, Astara, Rasht and Sari.[81]

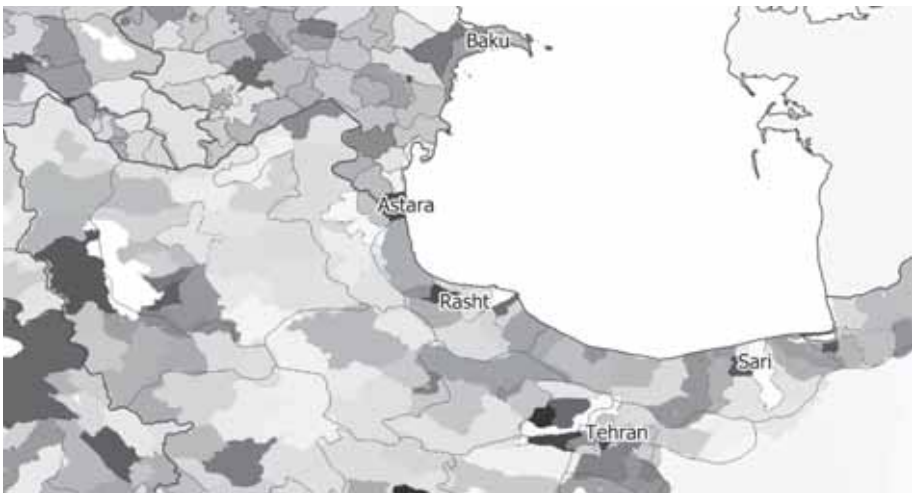


Figure 9. Population density in the area surrounding the Caspian Sea

Stilo (forthcoming a.) notes that the Caspian languages consist of a long and uninterrupted continuum of dialects without clear borders between Mazandarani, Central Caspian and Gilaki varieties, making up a total distance of 550 kilometers starting from Gorgan in the east to Enzeli and Fuman in the west. In this area,

[81] Map from Center for International Earth Science Information Network - CIESIN - Columbia University & Centro Internacional de Agricultura Tropical - CIAT (2005). The white areas are probably areas for which there is no data on population density.

speakers understand their closest neighbors without difficulty, but as distance grows mutual intelligibility decreases.

5.2.1.2.1 Mazandarani

The province of Mazandaran is home to almost four million people, 60% of which live in rural areas and 40% in urban areas. Mazandarani is the main language spoken by the inhabitants but due to colonialization, language contact and migration, other languages such as Persian, Azerbaijani, Balochi, Turkmen, Zaboli, Kurdish, Gilaki, Tati, Arabic, Godari, Russian, Katuli, Aftari, and Luri, are also reported to be spoken in the province (Shahidi 2008:23). The influence of Persian on Mazandarani is especially heavy in the urban areas, in some cases Persian has replaced Mazandarani (Shahidi 2008:28). Shahidi's (2008:295–297) study on the shift from Mazandarani to Persian among the Mazandarani population concludes that the demise of Mazandarani is probable, especially in the urban areas where new generations do not speak Mazandarani fluently.

According to Ethnologue, Mazandarani is spoken by over two million speakers mainly in the province of Mazandaran. It is one of the few Northwestern languages that has a written history (Borjian 2004:291; Shahidi 2008), although it is not a written language today. Mazandarani can be divided into three main varieties: the standard or eastern variety which is spoken in Sari, the capital of Mazandaran, the central variety, and the western variety. Speakers of the different varieties may also communicate in Persian. It is also noted that almost every Mazandarani locality has its own sub-dialect, with varieties of neighboring villages exhibiting differences in phonological systems, while the lexicon is fairly uniform (Borjian 2004:395).

Shahidi (2008:21–22) reports that the Mazandaran province is one of the most important agricultural areas of Iran and that the occupations of most of the inhabitants are in agriculture, cattle-raising, fishing, and in the industry.

5.2.1.2.2 Gilaki

Gilaki is spoken mainly in the province of Gilan. Stilo (2001:660) reports that there are possibly three million people who speak Gilaki as a first or second language, while Rastorgueva et al. (2012:1–2) report the Gilaks to be approximately 700 000 in number. In the cities, many Gilaki speakers are bilingual and speak Persian as well. In the rural districts, however, many Gilaki speakers, especially the women, do not speak Persian. There have been unsuccessful attempts at conventionalizing written Gilaki.

5.2.1.2.3 Taleshi

Taleshi is spoken along the western parts of the Caspian coast and up into the mountain areas in Iran and the Republic of Azerbaijan. The Taleshi language is divided into three dialects; southern, central and northern. The language is under the influence of Persian and Azerbaijani. Although the speech communities of

these three dialects define themselves as one unified ethnicity, Stilo (2008:364) argues that they are structurally and lexically so different, and that the mutual intelligibility between the dialects is so low, that one can see them as separate languages. In fact, the Taleshi dialects are in some cases closer to central and northern Tati dialects. There are also transitional dialects between Taleshi and Tati.

Ethnologue estimates that there is a total amount of 900 000 Taleshi and Tati speakers, out of which 112 000 Taleshi speakers in Iran. Regarding the Taleshi population in Azerbaijan, Tiessen (2003) reports close to 500 000.

Clifton et al. (2005:4), who study the northern dialects of Taleshi spoken in Azerbaijan, report the Taleshi language community being both multilingual and multicultural. The impact off Azerbaijani is great since education, television and media are primarily in Azerbaijani. During the Soviet period (1922-1991), there was also Russian influence on northern Taleshi, although Azerbaijani remained the main language of communication especially within the area of economics and politics. The Taleshi area was of importance to the Soviet Union as it provided a large amount of produce. The area was also of importance from a military point of view. Clifton et al. (2005:5) notes linguistic differences between the speakers of the mountain area, who are Sunni Muslims, and the speakers of the lowland area, who are mostly Shiite. The economic situation in the lowland area is stronger due to access to better transportation network. This means that the lowland population has more contact with other communities as well, such as Russian and Azerbaijani communities. Clifton et al. (2005:5) also report that many lowland areas that have previously been homogeneous are now becoming ethnically mixed. In addition, Azerbaijani has gained even higher prominence in this region.

According to Stilo (1981:143), the border between the Gilaki and the Taleshi is abrupt with no transitional dialects. The languages are also not mutually intelligible. He notes however that Gilaki and Taleshi share many linguistic features.

> They [Gilaki and Taleshi] coincide, however, in the greater part of their phonological system, if not all, and share many grammatical patterns, some of which are uniquely characteristic to them and do not exist in Iran outside of this geographic area. One possible explanation is that these common unique features are the result of a mutual influence from a previous substratum language. (Stilo 1981:143–144)

Throughout history, seasonal work-related migration of Taleshi and Tati speakers to the province of Gilan have created multilingualism in the area (Windfuhr 1989b:248). Paul (2011:320) reports that the influence from Gilaki on the Taleshi language is patchy, as it only concerns those speakers (typically male) that trade with the Gilaki population. Regarding Taleshi in Iran, Paul (2011:320–321) concludes tentatively that Taleshi remains the main language among the Taleshi-speaking community for those older than 25, that the younger generation has

started to shift to Persian, and that although language attitude towards Taleshi is generally positive most parents prefer to speak Persian to their children, as Taleshi is seen as disadvantageous.

5.2.1.2.4 Tati

The sociolinguistic situation of the Tati varieties is not well documented. It is noted that Tati varieties are often influenced by Azerbaijani, and Yarshater (1969:21) notes that speakers in Tati villages are often trilingual in that they speak their Tati variety, Persian and Azerbaijani.

5.2.1.3 Verb morphology

For the reader unfamiliar with Iranian languages, a short overview of relevant verb forms in Mazandarani, Gilaki, Tati and Taleshi is given here. In these varieties, tense distinctions are most often expressed through different sets of verbal stems. The two sets may go under the name of present and past stems, non-past and past stems or stem I and stem II. Historically, the two stems reflect present and past tense, where the present stem is inherited from Old Iranian unless it is a novel formation, and the past stem is derived from the 'verbal adjective' in *-ta-* (Korn 2017:38), but synchronically the temporal distinction is not necessarily consistent. In most varieties, there are ways to derive one stem from another with an often rather large group of exceptions.

Tense, aspect and mood categories are built on the two verbal stems. These stems are however not used for the same things in these languages. In the Taleshi varieties, such as, for example, Anbaran Ardebil, Asalemi and Masal-Sandermani, stem I, which historically was the present stem, is used for forming the present tense and past imperfective, and stem II is used for forming past perfective (Paul 2011). In Southern Tati varieties, Ziarati and Gilaki, the present stem is used to form the present tense, while the past stem is used to form the past perfective and past imperfective (Shokri et al. 2013; Rastorgueva et al. 2012). However, in other Mazandarani varieties such as Babolsari, Khatirabadi, Amoli and Saravi, the present stem is used to form the present tense but also the past tense together with past prefixes and suffixes (Stilo forthcoming a.). In the examples of this chapter, I have kept the glossing or description provided by authors, but it is noted that a term such as 'present stem' is not necessarily restricted to forming present tense patterns, and that reference to stem I and II does not provide information other than the distinction between two stem forms.[82]

[82] I am aware of the fact that the glossing of stems as present and past, versus stem I and stem II, is inconsistent. One option would be to change all glosses so that stem I would be glossed as present and stem II as past, or the other way around. However, it is my impression that in those varieties where authors talk of present and past stems, these stems are used to a greater extent to form present and past tenses, whereas in varieties where authors speak of stem I and stem II

In Gilaki, spoken in Rasht, verbs have a present and a past stem. The present stem is used to form the present-future tense, the imperative mood and the "the present-future tense of the subjunctive mood" (Rastorgueva et al. 2012:160). The past stem is used for the past tense (both perfective and imperfective), the participle and the infinitive (Rastorgueva et al. 2012:118, 160). The past perfective is formed with *bV-* or preverbs prefixed to the past stem, while the past imperfective is formed with a suffix *-i* suffixed to the past stem (Stilo 2001). The verb 'do', for example, has a present stem form *kun-* and a past stem form *kud-*. Additional preverbs[83] and suffixes may attach to the verb. For forming the infinitive, the infinitive marker *-ən* is attached to the past stem, as in *kud-ən* 'to do'.

The verbal stems take different sets of personal endings. There are three sets in Rashti [G]; one used when forming the present-future tense, one used for the aorist and the past tense, and a third used for the past imperfective (Rastorgueva et al. 2012:120). As seen in examples 5:3a), b) and c), the verbal stems as well as the personal endings of 3SG differ in the present-future, the past and past imperfective, respectively.

5:3 Rashti, Gilaki (Rastorgueva et al. 2012:303, 317, 356)

a) Present-future

ti	*amara*	*čuǰur*	*rəftår*	*kun-e.*
2SG.GEN	with	what.way	behavior	do.PRS-3SG.SET1[84]

'How does he treat you?'

b) Past

[…] *mara*	*bəɣəl=a*	*kud-ə* […]
1SG.ACC/DAT	embrace=PFV	do.PST-3SG.SET2[85]

'…he embraced me…'

c) Past imperfective

[…] *ammå*	*nəsim=ə*	*xunək=i*	*va=ze-i.*
but	breeze=EZ	cool=IND	PVB=hit.PST-IPFV.3SG.SET3[86]

'[…]but a cool breeze was blowing.'

---

this is less so. In addition to this, in some varieties, tense is marked by items other than stems. For these reasons, the glossing of stems in this chapter is kept close to the one given in the descriptions.

[83] Authors use the term *preverbs* to refer to verbal prefixes, which can be both derivational and grammatical markers.

[84] Orig. do.PRS-3SG.

[85] Orig. do.PST-3SG.

[86] Orig. PVB=hit.PST-IPFV-3SG.PST.

In Mazandarani, Gilaki, Tati and Taleshi, personal endings may be used to mark, in addition to person and number, temporal differences, aspectual nuances or transitivity. In the Taleshi of Anbaran Ardebil, for example, there are three sets of personal endings used for different purposes named SET1a, SET1b and SET2. The present tense is formed with stem I, a locative marker *-na-* (which I will gloss as DAR, see section 5.3.1.3), and SET1b personal endings. The past imperfective is marked through the verbal prefix *a-*, stem II and SET1a. There is also a past perfective that differs between intransitive and transitive sentences so that intransitive constructions are formed by stem II and SET1a personal endings, and transitive constructions formed by stem II, SET2 personal endings and a marker *-e*. Personal endings in Taleshi varieties are clitics and may float. The present tense, past imperfective, past perfective intransitive and past perfective transitive are given in 5:4a), b), c) and d), respectively.

5:4 Anbaran Ardebil, Taleshi (Paul 2011:123 mg, 129, 136 mg, 137)

a) Present
ža=na=m.
hit.I-DAR-1SG.SET1B
'I hit'

b) Past imperfective
gândəm devan a-k-im.
wheat scythe PVB-do.I-1SG.SET1A[87]
'I was scything the wheat.'

c) Past perfective intransitive
š-em.
go.II-1SG.SET1A
'I went.'

d) Past perfective transitive
bavə=ru=m *i-tka* *xuruš* *pât=e*.
3SG.IO=for=1SG.SET2 a-little stew cook.II=TR[88]
'I cooked a little stew for him.'

The issue of transitivity is complex, often it is not clear if or how transitivity interacts with the progressive. It is not my impression, however, that transitivity affects the forms of progressive constructions and, therefore, it will be put aside from here on.

[87] Orig. AUG-do-IPFV.1SG.
[88] Orig. cooked=TR.

A note on the *a-* prefix, which seems to function as an imperfective marker in Taleshi varieties, is made here. Authors have chosen to call this marker *augment* which suggests that it is preserved from Early Old Iranian. Since this relation is not proven (Agnes Korn, p.c.), the marker is simply glossed as verbal prefix in this chapter. *Augment* is a verbal prefix in Proto-Indo-European that is best attested in Sanskrit and has the function of past tense (see, e.g., Fortson IV 2014:101).

As we observed above, in Rashti [G] and Anbaran Ardebil [TAL] it is quite common that varieties have an aspectual distinction dividing the verb forms formed on the past stem into one imperfective form and one perfective form. To have different verb forms for the past imperfective and the past perfective is in fact a general tendency in Mazandarani, Gilaki, Taleshi and Tati. In some varieties, the distinction between perfective and imperfective in the past is marked through verbal prefixes. For example, in Mazandarani spoken in Babolsar, Khatirabad, Amol and Sari, the constructional schema V.PRS-(*n*)-PN is used for the present tense, V.PST-PN for the past imperfective and *bá*-V.PST-PN for the past perfective.

5.3 The DAR and KAR gram families

In this section, two groups of constructions used for marking the progressive gram type are presented and discussed. These groups are referred to as DAR and KAR gram families. They are found in northern Iran and southeastern Azerbaijan in the varieties of Mazandarani, Gilaki, Tati and Taleshi from which data from 50 villages or cities have been collected. The location of these villages is shown in Appendix F. Some additional Iranian and non-Iranian varieties are also discussed in section 5.3.4. All the varieties discussed in this section are given, with references in Appendix E.

Stilo (in press), notes that there are three types of progressives among the Caspian and Tatic varieties, one locative type, which corresponds to what I will call the DAR gram family, one *kǎr* 'work, doing' type, which corresponds to what I will call the KAR gram family, and one where the DAR and KAR elements are combined. The types provided in this section build on his analysis but group the data from a grammaticalization perspective. Further varieties are also added to his data.

In what follows, certain varieties may have more than one of the mentioned progressive constructions. Also, certain varieties may have progressive constructions that do not fall into this grouping. In this section, only the DAR and KAR constructions with their various merges and functions are dealt with.

Section 5.3.1 introduces the data on the DAR gram family and section 5.3.2 the data on the KAR gram family. Section 5.3.3 presents varieties where both DAR and

KAR are found. Section 5.3.4 discusses some non-Iranian varieties that may be related to the DAR gram family. Section 5.3.5 discusses and concludes section 5.3.

5.3.1 The DAR gram family

For orientation, an overview of the outcome of section 5.3 is already presented here. The data presented in this section divides the DAR constructions into four main types depending on the structural and functional features presented in Table 23. The functional coding, e.g. DAR:PROG, means that we are dealing with a DAR construction that has the progressive as its main function.

The map in Figure 10 shows that the geographical spread of Types 1 to 4 covers the Caspian Sea from Mazandaran up to the southeastern parts of Azerbaijan.

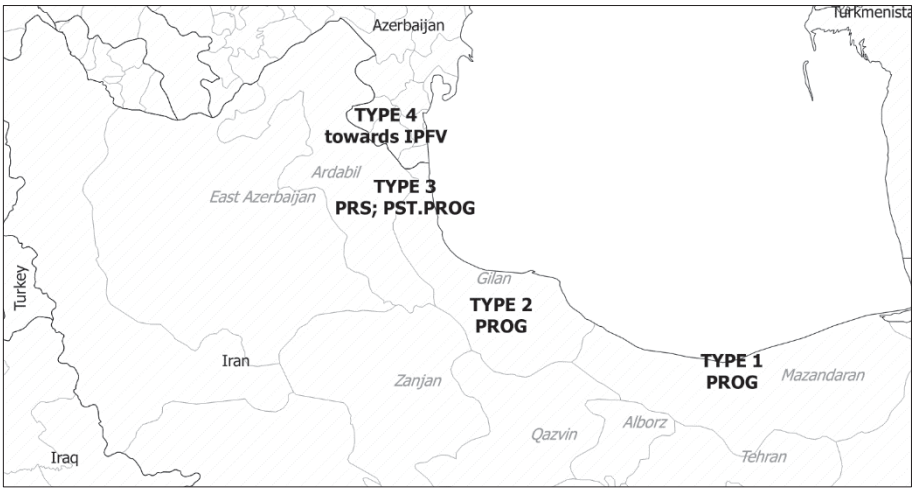


Figure 10. Locations of Types 1, 2, 3 and 4 of the DAR gram family

As can be seen in Table 23, Type 1 and 2 constructions have the same function span, while Types 2 and 3 share the structural feature of using a postposed DAR element, in varying degree of inflectionality with a non-finite element. Type 3 and 4 constructions share the same schema, which differs in function.

Table 23. Function and structure of DAR constructions

TYPE	FUNCTION	CONSTRUCTIONAL SCHEMA	STRUCTURAL FEATURE	VARIETY
Type 1	PRS PROG PST PROG	DAR.PRS-PN […] V.PRS-PN DAR.PRS/PST-PN […] V.PST-PN	periphrastic, preposed and tensed DAR, finite main verb	Mazandarani, Central Caspian
Type 2	PRS PROG PST PROG	V-INF(-)DAR.PRS-PN V-INF(-)DAR.PST-PN	inflectional or periphrastic, postposed and tensed DAR, non-finite main verb	Gilaki, , Central Caspian, Tati
Type 3	PRS PRS PROG PST PROG	V-(INF=)DAR-PN - V-(INF=)DAR=COP.PST-PN	clitic, postposed non-tensed DAR, non-finite main verb	Taleshi
Type 4	PRS PRS PROG PST PROG/PST IPFV	V-INF-DAR=PN - V-INF-DAR=COP-PST-PN	inflectional/ periphrastic, postposed and non-tensed DAR, non-finite main verb	Taleshi varieties Leriki and Shuvi.

Structurally, the Type 1 schema differs radically from the others as it is periphrastic, has a preposed DAR element and involves a finite form of the main verb. Functionally, the Type 3 schema is most often a marker of the general present in the present tense and the past progressive in the past tense. In Type 4, the schema in the present tense is a general present, while the schema in the past is a past progressive with some additional uses that are typical for imperfectives. In this area, the constructions within the DAR gram family change from periphrastic to inflectional, from preposed to postposed, and from functioning as progressives to marking the present or past imperfective. The synchronic pattern constitutes an *areal grammaticalization cline*, as discussed in Chapter 1, section 1.2.3.

In the upcoming sections, Types 1, 2, 3 and 4 are presented in that order.

5.3.1.1 Type 1: Mazandarani & Central Caspian

Constructional schemas of Type 1 are marked by a locative copula *dar-* which most often has a present and a past form. These patterns are found in Mazandarani and occasionally in Central Caspian. As can be seen, the locative copula in Babolsari [M] is invariable for person number in 3SG.

5:5 Babolsari, Mazandarani (Stilo forthcoming a. mg)

a) *dar-ɛ* *šúmmɛ.*
DAR.PRS-3SG go.PRS.1SG
'I am going.'

b) *dayy-ɛ* *ší(i).*
DAR.PST-3SG go.PST.2SG
'You were going.'

Similarly, the varieties Khatirabadi [M], Qa'emshahri [M], Yushi [M], Velatrui [M], Dikin Maraqei [M] (Stilo in press; Stilo forthcoming a.) and also, occasionally, Saravi [M] (Stilo forthcoming b.) have a progressive construction with the locative copula *dar-* which shows tense but not person and number, instead all forms show third person singular. Stilo (in press; forthcoming a.) refers to this pattern as the "true" Mazandarani and views the conjugation of person number as a Persian influence.

In Ziarati [M], Behshahri [M], Saravi [M], Kelardashti [Central Caspian], Kalarestaq [Central Caspian] and also the variety Gurani [Tatoid][89], the locative copula conjugates for person and number and may show tense (Borjian 2010; Shokri et al. 2013; Stilo forthcoming a.). A similar pattern is also found in

[89] Windfuhr (2009:12) classifies this variety as Upper Zagros and Central Plateau Group. Stilo (forthcoming b.) refers to this variety as Tatoid.

Kandelusi [Cantral Caspian] although it is unclear whether the copula conjugates for person and number. In the Ziarati [M] present progressive construction, the present tense form of the main verb takes the imperfective *me-* marker. In the past, the DAR element combines with the past imperfective with the structure (*m*(*e*)-)V.PST-PN[90], preverbs may also attach to the verb forms (Shokri et al. 2013:44). In this Mazandarani variety, both the present and past progressive are formed with the present form of the locative copula. The present and past progressive forms are shown below in 5:6, as well as an example of *dar-* used in a locative, non-progressive, construction.

5:6 Ziarati, Mazandarani (Shokri et al. 2013:22, 46–47)

a) *emā* *dar-im* *nāār* *me-xor-im.*
we DAR.PRS-1PL[91] lunch PREF.IPFV-eat.PRS-1PL
'We are having lunch.'

b) […] *dar-em* *lālā* *kārd-im.*[92]
DAR.PRS-1SG[93] lullaby do.PST-1PL
'[…] I was singing a lullaby.'

c) *alān* *de=tā* *xod=em-e* *vač-a*
now two=CL self=PC.1SG-GEN child-PL

dar-en *palu=š.*
DAR.PRS-2PL[94] near=PC.3SG
'Two of my sons work (lit. 'are') with him now'

Varieties Dodanga'i [M] and Kordkheyli [M] have two constructions, one where the locative copula is inflected for person-number and one where the locative copula is invariant for third person singular (Stilo forthcoming a.). Moving towards the capital, in Amoli [M], we do not find a locative copula, but the *dāštan* 'have' progressive, identical to the Persian pattern (Stilo forthcoming a.). This pattern is not included in the DAR gram family (but see Chapter 6).

The DAR construction is also found in Sangesari, a Semnanic variety located south of Mazandaran, to the east of Tehran close to the city of Semnan. The

[90] Examples indicate that the *m*(*e*)-marker is optional in the past
[91] Orig. be.PRS-1PL.
[92] In Mazandarani, similar to other Iranian varieties, it is not uncommon to refer to oneself in the plural. The peculiarity of using 1SG on the locative verb and 1PL on the main verb, however, may be due to stress in the interview situation (Guiti Shokri, p.c.). No other example of a past progressive is available.
[93] Orig. be.PRS-1SG.
[94] Orig. be.PRS-2PL

construction is formed through the "resultative form of *dab* 'be in'" plus the imperfective, e.g. *dare šuonde* 'you are going' and *dabiye mišuye* 'you were going' (Azami & Windfuhr 1972:118–119).

5.3.1.2 Type 2: Gilaki and Tati

The Type 2 schema combines a non-finite form of the verb followed by a DAR element, unlike the *dar-* copula found in the Mazandarani varieties, which is finite and precedes the verb. Both western and eastern Gilaki, represented by Rashti [G] and Lahijani [G], respectively, as well as Langerudi [G], Ramsari [G/Central Caspian], Tonekaboni [Central Caspian] and Tutkaboni [Tatoid] have this pattern (Stilo in press). An example is shown in 5:7. In Gilaki, the *dər-* copula combines with the infinitive form of the verb, which is formed with the past stem taking the infinitive marker *-ən*/*-an*/*-en*/*-on* (Rastorgueva et al. 2012:134; Stilo in press). Stilo (2001:663) also provides a past progressive form, as seen in 5:7b). 5:7c) is an example of the locative copula being used in locative function. However, Rastorgueva et al. (2012:134) only encounter the independent locative copula in the third-person singular in the present tense.

5:7 Rashti, Gilaki (a) and c): Rastorgueva et al. 2012:336, 134 mg; b): Stilo 2001:663, mg and transl.)

a) *či* *kud-ən* *dər-i?*
what do.PST-INF DAR.PRS-2SG
'What are you doing?'

b) *bíšt-án* *dubu*.
fry.PST-INF DAR.PST.3SG
'S/he was frying.'

c) *mizə-ru* *du-ta* *kitåb* *dər-ə.*
table-on two-CL book DAR.PRS-3sg
'On the table there are two books.'

For Type 2, authors vary in their orthographic representation of the DAR locative copula as attached or not attached to the infinitive verb. Looking at the data in Rastorgueva et al. (2012) and Stilo (2001; in press), it looks as if the DAR element always follows the infinitive verb, i.e. that nothing can intervene between the verb and the DAR element. This is interesting, since the Gilaki constructions are found between the periphrastic Mazandarani progressives, in which elements may intervene between the DAR locative copula and the main verb, and the enclitic postposed *=da* locatives found in the Taleshi varieties.

5:8 Lahijani, Gilaki[95] (Stilo 2001:666 mg; Stilo in press)

a) *gitə́-dər-əm.*
take.INF-DAR.PRS-1SG
'I am taking.'

b) *mu yeza xord-é dər-əm.*
I food eat-INF DAR-1SG[96]
'I am eating.'

Interestingly, the Lahijani progressive construction can be negated, which is not possible for the Mazandarani varieties of Behshahr and Sari. There is no information regarding negation for the other Mazandarani varieties. The Gilaki varieties also have other progressive patterns, these are discussed in 5.3.3.1.

As already mentioned in section 5.1.1, in the Southern Tati variety Chali [T], the locative element in the progressive construction is realized as *-u-*. Yarshater (1969:119) comments that *-u-* has a wide range of functions with its most frequent meaning probably being 'in, within, inside'.[97] In Chali the progressive construction is formed with the infinitival verb + *-u-* + copula. In example 5:9, we see *-u-* used in the progressive pattern, followed by an example of a locative use.

5:9 Chali, Tati (Yarshater 1969:225 mg, 120 mg)

a) *ešta tete xord-an-u-ind.*
your daughter.PL eat. PST-INF-DAR-COP.3PL
'Your daughters are eating.'

b) *ceme jíf-u.*
me/my pocket-LOC
'In my pocket.'

Apart from Chali [T], the Tati variety Khoini [T], Lerdi [T] and what is referred to as Northern Tati (Stilo in press) are also reported to have a DAR progressive pattern. Unfortunately, for these varieties, no information regarding past progressive DAR patterns is available.

[95] In Stilo (2011:660), Lahijani is refered to as Eastern Gilaki.
[96] Orig. BE4-1SG. The subscripted 4 indicates a type of copula.
[97] There is also a *-ku* 'from, in' postposition in Chali, which is used for persons. The *-ku* postposition may be related to the locative case *=ku* in Taleshi varieties, which can be realized as *=u* in the variety Anbaran Ardebil, where it loses the initial consonant /k/ (Paul 2011:161). The *=(k)u* locative case is, however, not used in the progressive construction of Anbaran Ardebil.

5.3.1.3 Type 3: Taleshi

The Type 3 schema includes an often non-finite verb with an enclitic postponed DAR element. We find this schema in some Taleshi varieties such as Anbaran Ardebil [TAL], Anbaran Mahalle [TAL], Viznei [TAL] and Jokandani [TAL]. In these varieties, the present tense schema marks the general present tense, while in the past, the schema (including an auxiliary) is used for past progressive. This asymmetry will be analyzed as a result of the present progressive having grammaticalized into a general present. An example is given for Anbaran Ardebil in 5:10 to illustrate how a Type 3 pattern is used also with stative verbs.

5:10 Anbaran Ardebil, Taleshi (Paul 2011:152)

avün zən=na *nə=b-in* […]
3PL know=DAR[98] NEG=AUX-3PL
'They did not know…'

In Anbaran Mahalle [TAL], Viznei [TAL] and Jokandani [TAL], the DAR element is realized as a locative *=da=* marker (Paul 2011:155). Note that the DAR element and the personal endings are clitics and may float, which is evident in example 5:11a) given below. According to Stilo (2008:373), similar patterns are also found in non-Iranian languages such as Armenian, Azerbaijani, Aramaic and Udi varieties, some of which will be discussed in section 5.3.4.2. In 5:11, examples of the present, past progressive and simple past from Jokandani [TAL] are given, in that order. The simple (perfective) past is formed through adding the prefix *bə-* to verbs that do not have preverbal elements.

5:11 Jokandani, Taleshi (Paul 2011:156–157)

a) *əm-e* *həye gəla* *xəc=in* *hard-e=da.*
DEMP-PL three CL pear=3PL eat-INF=DAR[99]
'Three of them are eating pears.'

b) *a* *va=na…* *š-e=da=b-e.*
DEMD direction=with … go-INF=DAR=AUX.PST-3SG[100]
'He was going in that direction.'

c) *bimi=šun* *kumak* *bə-kard=e.*
3SG.IOP=3PL help PST-did=TR
'They helped him.'

[98] Orig. know=LOC.
[99] Orig. eat-INF=LOC.
[100] Orig. go-INF=LOC=AUX-3SG.

In Anbaran Ardebil [TAL], the locative marker is instead realized as *=na=* (Paul 2011), which was discussed in section 5.1.1 as likely to be a cognate to the other DAR elements and assumed to originate from *-anda*. An example of *=anda* and *=da* used as locative markers in non-progressive constructions are given in 5:12 and 5:13, respectively. Interestingly, in example 5:13, there is no additional verb in the last clause, which may suggest that *=da* can be analyzed as a locative copula.

5:12 Anbaran Ardebil, Taleshi (Paul 2011:164)

šünapapü	*i-la*	*vər=anda*	*təktək*	*a-k-i*
woodpecker	a-CL	place=DAR[101]	pecking	PVB-do-IPFV.3SG[102]

'In one place a woodpecker was pecking.'

5:13 Anbaran Mahalle, Taleshi (Paul 2011:359)

hərdan-en	*daivard-in*	*bə-š-in*	*b-a*	*taraf;*
child-PL	passed.by-3PL	PST-go-3PL	to-DEMD	direction

hanuz=an	*a*	*merd*	*hala*	*əštan*	*du*	*bən=da*
still=also	DEMD	man	still	self	tree	beneath=DAR[103]

'The children passed by and went in that direction; still the man stayed under the tree.'

In Anbaran Ardebil [TAL], the patterns in the present and past does not include an infinitive marker. In Viznei [TAL], the pattern in the present does not have an infinitive marker whereas the past has. In both these varieties the present form of the pattern functions as a general present whereas the past form functions as a past progressive.

5:14 Viznei, Taleshi (Paul 2011:156)

a) *vind=əš=e* *kə* *əm-e* *xəc=in* *har(d)=da.*
saw=3SG=TR CMPL DEMP-PL pear=3PL eat=DAR[104]
'He saw that they are eating pear(s).'

[101] Orig. place=LOC.
[102] Orig. AUG-do-IPFV.3SG.
[103] Orig. beneath=LOC.
[104] Orig. eat=LOC.

b) *ila* *suk-a* *sas=i* *uma-i=da=b-e…*
a cock-LNK voice=IND came-INF=DAR=AUX-3SG[105]
'A cock crow was resounding…'

Stilo (in press) notes that there is also a DAR pattern in Karani [T] and Asalemi [TAL], which is mainly used with proximative function. No example is given, however. Since the structure of this pattern is not identified, it is not included in Table 23.

5.3.1.4 Type 4: Leriki [TAL] and Shuvi [TAL]

In the Taleshi variety Leriki, similar to the Taleshi varieties mentioned in section 5.3.1.3, a construction with a non-finite verb and a *-dæ* element is used for the general present. The equivalent schema is also found in the past where it marks the past progressive but also has uses typical for past imperfective. In the past, however, there also exists a 'proper' past imperfective pattern. The past DAR pattern is noted to often be interchangeable with the past imperfective. This means that the past DAR pattern has not taken over the imperfective function entirely. Example 5:15 shows the DAR construction in Leriki with ongoing and habitual use, respectively. Stilo (in press) gives the *-dæ* element with inflectional boundaries rather than clitical ones. In the Taleshi varieties discussed so far the boundaries are clitical. The inflectional boundaries of *-dæ* could be a sign of further grammaticalization.

5:15 Leriki, Taleshi (Stilo in press)

a) *ayïl* *vít-dæ=b-e* *bæ* *di* *mašin-í.*
child run-DAR=AUX.PST-3SG$_1$[106] to after car-OBJ
'The child was running after the car.'

b) *penj* *sor* *vaxt* *doy-dæ=b-in.*
five year time give-DAR=AUX.PST-3PL$_1$[107]
'They used to give 5 year leeway.'

A similar situation is found in the variety of Shuvi [TAL], spoken in the same area as Leriki [TAL],[108] which has a general present pattern V-INF-*da*-PN and an 'analytic imperfect' marked as V-INF-*da* AUX-PST-PN, these are given in 5:16a) and b). In the past, there also exists another synthetic imperfect with "strong

[105] Orig. came-INF=LOC=AUX-3SG.
[106] Orig. run-LOC=AUX.PST-3SG1, where 1 refers to set of endings.
[107] Orig. give-INF-LOC=AUX.PST-3PL1, where 1 refers to set of endings.
[108] Schulze (2000:6) does not give an exact location of Shuvi but rather presents an area in which it is found.

modal connotation" (Schulze 2000:23, 48) marked with the verbal element *a-*, attached to the verbal stem as shown in 5:16c). Similar to Leriki, the past DAR pattern in Shuvi is most likely a former past progressive that is now taking over the past imperfective function.

5:16 Shuvi, Taleshi (Schulze 2000:46)

a) *s-é-da-m.*[109]
carry-INF-DAR-1SG[110]
'I am carrying, I carry'

b) *om-e-da* *b-i-m.*
come-INF-DAR[111] AUX-PST-1SG
'I was coming.'

c) *a-vot-i-m.*
IPFV-say-PST-1SG[112]
'I was saying.'

5.3.2 The KAR gram family

Figure 11 shows the geographic spread of the KAR gram family. Here, the name and the genealogical sub-category of the variety and the function of the KAR construction are given.

Structurally and functionally, the KAR gram family is more homogeneous than the DAR gram family and is found in fewer varieties. The KAR gram family is also in most cases limited to the meeting point between the Iranian provinces Gilan, Ardabil, East-Azerbaijan and Zanjan. In all cases but two, the KAR construction is a progressive pattern. There is one instance of the KAR construction functioning as a general present tense and one instance of the KAR construction functioning as a proximative. Structurally, in most varieties found, the KAR element precedes the main verb, Kajali [T] and Karnaqi [T] are exceptions where the KAR element follows the main verb.

[109] According to Stilo (2008:374), the infinitive is based on the past stem. The example could then be glossed 'carry.PST-INF-LOC-1SG'.
[110] Orig. carry-INF-PRS-1SG.
[111] Orig. come-INF-LOC.
[112] Orig. AUG-say-PST-1SG.

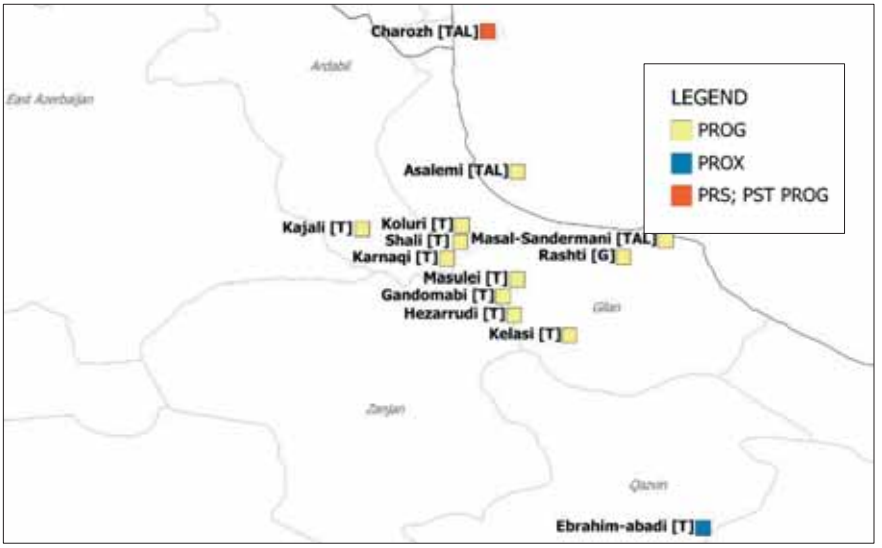


Figure 11. Function and sub-categories in the KAR gram family

As can be seen, most of the varieties in the KAR gram family are Tati, although we also find KAR progressives among Gilaki and Taleshi varieties. In fact, most Tati varieties in the sample have a progressive with a KAR progressive, although some also have DAR constructions, as was seen in 5.3.1, or a mix of the two, as will be seen in section 5.3.3.

Table 24 presents all the KAR constructions in the sample. Hybrid KAR-DAR patterns, i.e. patterns containing both a KAR and a DAR element, are also presented.

Table 24. Schemas in the KAR gram family including KAR & DAR schemas

KAR CONSTRUCTIONAL SCHEMAS	
KAR_{inv} + V_{fin}	Kelasi [T], Hezarrudi [T], Ebrahim-abadi [T], Shali [T], Masulei [TAL], Masal-Sandermani [TAL], Rashti [G], Gandomabadi [T], Koluri [T]
$PST.PTCP_{inv}$ + KAR_{fin}	Karnaqi [T], Karani [T]
V_{fin} + KAR_{inv}	Kajali [T]
KAR_{inv}(=PN) + $V_{non\text{-}fin}$	Asalemi [TAL]
KAR_{inv} + $V_{non\text{-}fin}$	Charozh[TAL]
KAR & DAR CONSTRUCTIONAL SCHEMAS	
KAR_{inv}-DAR_{inv} + V_{fin}	Kafteji [T]
V_{fin} + $KAR_{?}$-DAR_{fin}	Eshtehardi [T]
KAR_{inv} + $V_{non\text{-}fin}$ DAR_{fin}	Rashti [G]
$V_{non\text{-}fin}$=KAR_{inv}=DAR_{fin}	Lahijani [G]

As can be seen in the table, the most common constructional schema is one where an invariable form of the KAR element precedes a finite form of the main verb. In fact, in most varieties for which we have data, the KAR element is invariable. In Asalemi, however, the clitical personal endings may attach to the KAR element, and in Karani [T] and Karnaqi [T], the KAR element shows tense and follows an invariable past participle of the main verb.

In what follows, patterns used as progressives are presented first, followed by patterns with present tense function and finally those with a proximative function.

5.3.2.1 KAR: PROG patterns

Tati varieties such as Kelasi [T], Koluri [T] and Gandomabi [T] (Yarshater 1960; Stilo forthcoming a.; Stilo forthcoming b.), Shali [T] (Yarshater 1959), Hezarrudi [T] (Yarshater 1970) and also the Gilaki variety Rashti [G] (Stilo 2001) and Taleshi varieties Masulei [TAL] (Lazard 1978) and Masal-Sandermani [TAL] (Paul 2011) combine present and past forms of the main verb with a preverbal invariant KAR element to form a progressive.

The invariable KAR element has realizations such as, e.g., *kərə*, as in Masulei [TAL], or *kora*, as in Gandomabi [T], or *kærœ*, as in Hezarrudi [T] (Stilo in press), or *kerā*, as in Shali [T] (Yarshater 1959:60).

5:17 Rashti, Gilaki (Stilo 2001:665 mg)

a) *kə́ra* *gir-ə́m*.
KAR take.PRS-1SG
'I am taking.'

b) *kə́ra* *gift-i-m.*
KAR take.PST-IPFV-1SG
'I was taking.'

In Taleshi of Asalem (Paul 2011), the KAR element may take personal clitics. The personal clitics originate from a copula but has lost this sense (Yarshater 1996:101).

5:18 Asalem, Taleshi (Paul 2011:127, 128)

a) *a* *kâr=a* *hard-e.*
3SG KAR=3SG[113] eat-INF
'He is eating.'

[113] Orig. PROG=3SG.

b) *kâ=b-im* *bun-i* *sây* *â-kard-e.*
KAR=AUX-1SG[114] roof-OBL repair PVB-cause.to.be-INF
'I was repairing the roof.'

5.3.2.2 KAR: PRS pattern

In one variety, namely Charozh [TAL] which is spoken in the villages Sarak and Digadi in southeastern Azerbaijan close to the border of Iran, the KAR construction functions as the present tense, while the former present tense functions as the future tense. In the past, however, the KAR pattern is progressive which is contrasted with a past imperfective (Don Stilo, p.c.).

5:19 Charozh, Taleshi (Stilo in press, Don Stilo p.c.)

a) *čo* *sahat* *dærs* *ko=mun* *do-y.*
four hour lesson KAR=1PL[115] give-INF
'We teach for four hours.'

b) *dærs-í=ko*[116] *ko=b-im* *ome*
lesson-OBL=from KAR=AUX.PST-1SG[117] come:INF
"I was coming home from school (lit: lessons)"

Structurally, as well as functionally, this pattern is similar to the DAR patterns in the area. This matter will be further discussed in section 5.3.5.

5.3.2.3 KAR: PROX pattern

In the Tati variety of Ebrahim-abadi [T], the construction containing the KAR element, both in the present and past, is used for marking the "imminence of an action or a state" (Yarshater 1969:225), that is, it has proximative function.

5:20 Ebrahim-abadi, Tati (Yarshater 1969:225 mg)

a) *vela* *kārā* *m-ā-karia.*
flower KAR IPFV-PVB-open.PRS.3SG.F
'The flower is going to open.'

[114] Orig. PROG=AUX-1SG.
[115] Orig. PROG=1PL1.
[116] Don Stilo (p.c.) comments that this *ko* 'from' is not related to the second *ko* which comes from *kVrV* (<*kăr* 'work, doing', Windfuhr 1989: 256).
[117] Orig.PROG=AUX.PST-1S.

b) *vela* *kārā* *m-ā-kariasta.*
flower KAR IPFV-PVB-open.PST.3SG.F
'The flower was going to open.'

The Ebrahim-abadi [T] KAR construction is regarded as belonging to the KAR gram family, as the notion of gram family includes patterns that have arisen either as a result of a common parent language or of language contact, even if the meaning has changed.

5.3.3 The DAR & KAR or KAR-DAR patterns

In some varieties, two separate patterns are found, one DAR and one KAR pattern. In other varieties patterns combining DAR and KAR elements are found, I will refer to the latter as hybrid constructions. The hybrid constructions could be analyzed as a third gram family, the KAR-DAR gram family. In the present chapter, they have simply been presented as the hybrids DAR-KAR. Figure 12 shows the areal distribution of varieties that have either both a DAR and a KAR pattern or one pattern with DAR-KAR. The name and genealogical sub-category of the variety, the type DAR, KAR or KAR-DAR and the function of the construction are given. E.g. DAR: PROG, KAR:PROG means that the variety has both a DAR construction and a KAR construction, and that both these are used as progressives, and KAR-DAR:PROG means a progressive pattern containing both DAR and KAR elements.

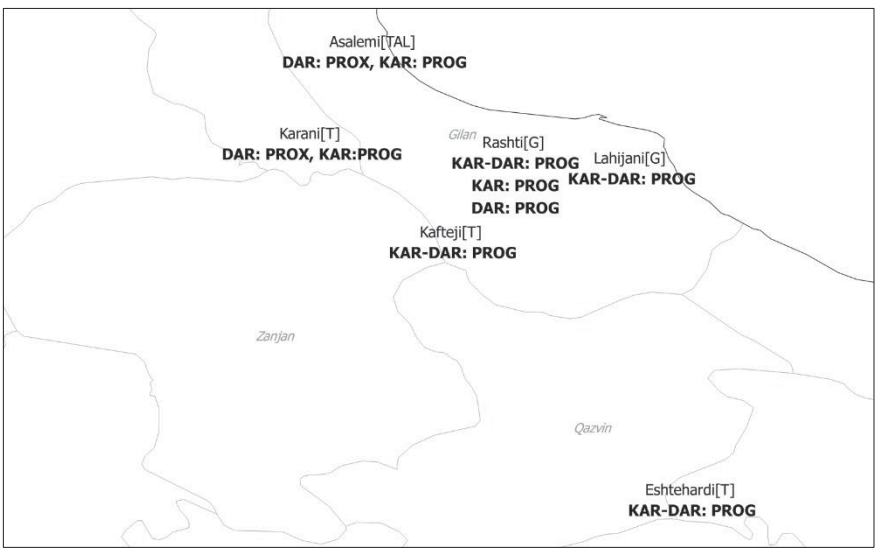


Figure 12. KAR-DAR or DAR and KAR constructions

5.3.3.1 DAR and KAR: PROG patterns

In Karani [T] and Asalemi [TAL] (Stilo in press), the constructions containing the invariable *kari* and *kárœ*, respectively, are used for the progressive, while the DAR construction is 'mostly' used with proximative/avertive function. The DAR construction in Asalemi [TAL] is not mentioned in Paul (2011), and no example is given in Stilo (in press). In Rashti [G], both a KAR and a DAR progressive pattern exist as well as a third KAR-DAR pattern. Examples of all three are given in section 5.3.3.2.

5.3.3.2 KAR-DAR: PROG patterns

In Rashti [G], there are three progressive patterns: 1) one pure KAR construction: *kərə*/*a* + finite form of the verb, 2) one pure DAR construction: a construction where *dər* immediately follows the verb as in V-INF + *dər*-PN[118], and 3) a hybrid construction: a construction where *kərə*/*a* has been added to the second pattern, i.e. *kərə*/*a* […] + V-INF + (or -) *dər*-PN. The KAR pattern is given in 5:21a), the DAR pattern in b), and the hybrid pattern in c). The examples are all in the present tense. The pattern given in 5:21a) seems to lack a past form. The pattern in 5:21b) on the other has a past equivalent given in 5:22a), where the DAR element is in the past form.

5:21 Rashti, Gilaki (Rastorgueva et al. 2012:305, 299, 431)

a) […] *mən* *kərə* *birun* *a-yəm.*
1SG.NOM KAR[119] outside come.PRS-1SG
'[…] I am going out.'

b) *koya* *šo-on* *dər-i*
where go.PST-INF DAR.PRS-2SG[120]
'Where are you going?'

[118] Or, alternatively, where *dər* is attached to the verb, as in Stilo's transcription (2001:666), recall that there are some inconsistencies regarding whether the DAR element is attached or directly following the infinitive in Gilaki.
[119] Orig. PROG.
[120] Orig. LOC.PRS-2SG.

c) *məryəm* *din-e* *ki* *iskəndər* *xeyli*
Maryam see.PRS-3SG CMPL Iskander very

nåråhət=ə *kəra* *divanə* *bost-ən*
unhappy=be.PRS.3SG KAR[121] crazy become.PST-INF

dər-ə *una* *g-e* *de* *na.*
DAR.PRS-3SG[122] 3SG.ACC/DAT.DIST say.PRS-3SG EMPH no
'Maryam sees that Iskander is very upset and is going crazy so she says to him, "Surely not".'

A third past progressive, which is structurally, but not functionally, similar to the pattern in 5:21c), also exists in Rashti. Here, an optional KAR element is followed by a DAR element in the present with an addition of the past form of the auxiliary *buon* 'to be', i.e. (*kərə*/*a*) […] + V-INF + (or -) *dər*.PRS-PN + COP.PST.PN. This pattern is given in 5:22b). According to Rastorgueva et al. (2012:155), the KAR element adds emphasis to the sentence meaning.

5:22 Rashti, Gilaki (Rastorgueva et al. 2012:154–155, mg)

a) *kud-an* *dubum.*
do.PST-INF DAR.PST.1SG
'I was doing.'

b) *kərə* *xurdən* *dər-ə* *bum* *zəng*
KAR eat.PST.INF DAR.PRS-3SG be.AUX.PST.1SG ring

bə-ze-idi.
PFV-hit.PST-2/3PL
'I had been eating when they called.'

It is, however uncertain if the pattern in 5:22b) is really a past progressive or, rather, a pluperfect progressive, given that the pluperfect is formed by adding the past form of *buon* 'to be' to past participle forms of the main verb (Rastorgueva et al. 2012:377). Due to this uncertainly, this pattern is excluded from the investigation in section 5.4.

A progressive construction with KAR + DAR exists also in the Lahijani variety of Gilaki, in Kafteji and in the Tati variety Eshtehardi. In Kafteji, the KAR and DAR elements have merged into *kærdæ*. As seen in example 5:23, it is invariable for tense, person and number.

[121] Orig. PROG.
[122] Orig. LOC.PRS-3SG.

5:23 Kafteji, Tati (Stilo forthcoming b.)

a) *čəm* *bera* *kær-dæ* *me-ší-æ* *kelas.*
my brother KAR-DAR[123] IPFV-go-3SG.M. PRN
'My brother is going to Kelas.'

b) *æzíræ* *ge* *kærdæ* *kᵉæ* *me-šé-m,*
yesterday SUB KAR-DAR[124] house IPFV-went-1SG.M,

čəm *bera-r* *vénd=əm.*
my brother-MO_2 saw=$1SG_2$
'Yesterday when I (m.) was going home, I saw my brother.'

5.3.4 Turkic, Neo-Aramaic, Nakh-Daghestanian and Tat

In this section, some progressive and/or imperfective patterns in Turkic, Neo-Aramaic (Afro-Asiatic) and Nakh-Daghestanian languages are presented. There is not enough evidence to support a claim that these patterns are part of the DAR gram family, but there may be a relation between certain patterns presented here and the members of the DAR gram family. Some Iranian varieties, such as Tat (spoken in Azerbaijan and southern Russia) and Tati/Taleshi, are also included for the discussion since they display relevant patterns. Tat is not to be confused with Tati. In the Glottolog classification, for example, Tat languages belong to the southwestern branch of the Iranian language family.

5.3.4.1 Turkic varieties

The *-mAktA-* pattern in Turkish was already mentioned in Chapters 2 and 4. An example is given in 5:24 in which *-DA-* assimilates to /k/ and is realized as *-ta-*.

5:24 Turkish (Turkic) (Göksel & Kerslake 2005:332)

Bugün *aile* *yapı-sı* *hız-la* *değiş-mek-te-dir.*
today family structure-NC speed-INS change-INF-LOC-GM[125]
'Today the structure of the family is changing rapidly.'

This pattern also exists in Azerbaijani, as *-mAKda-*, examples of which are given in 5:25. Noting the structural parallel to the Taleshi pattern, Miller (1953:146) suggests that Taleshi has borrowed this pattern from Turkish (he also mentions

[123] Orig. PROG.
[124] Orig. PROG.
[125] Orig. change-IPFV-GM.

Azerbaijani influence). Schulze (2000:47) suggests that the Shuvi [TAL] pattern is a borrowing from Azerbaijani (Turkic).[126]

5:25 Azerbaijani (Turkic) (Schulze 2000:47)

a) *al-mak-ta-yım.*
carry-INF-LOC-1SG
'I am carrying.'

b) *gel-mek-te* *i-di-m.*
come-INF-LOC AUX-PST-1SG
'I was coming.'

Simpson (1957:29) who shortly refers to this pattern calls it the 'continuous present' and Doerfer (1988), who refers to this pattern as 'durative present', also notes the dialectal variant *-AdU* "and similar forms in the dialects". Schönig (1998:254) describes this it as "a more focal present" that is similar to the Turkish *-mAktA-*. Due to these descriptions, it will be viewed as having uses towards the imperfective, similar its Turkish counterpart.

5.3.4.2 Neo-Aramaic and Tat

As was noted above, both Gilaki and Taleshi and also some Tati varieties use infinitival patterns in their progressives and related present tenses. This feature is also found in some Neo-Aramaic (Afro-Asiatic) as well as the Iranian Caucasian Tat (Indo-European) varieties in the area. This is something also noted in Noorlander & Stilo (2015), who suggest that these patterns are related to the infinitival patterns discussed in this chapter. This section provides their data, complemented by data from Khan (2008) and Yarshater (2005). The data that is provided only deals with the present tense. For a presentation of how some of these patterns relate to other patterns marking the subjunctive and future, the reader is referred to Stilo & Noorlander (2015). The map in Figure 13 shows the location and the function of the patterns discussed in this section.

[126] However, Murad Suleymanov (p.c.) comments that: "This pattern indicates the claim of 'objective reporting' (Əfəndiyeva 2005:24) on the part of the speaker. It notably does not have a negative form. Likely of recent origin (Serebrennikov & Gadžieva 1979:170), it is mainly associated with journalistic writing in contemporary Azerbaijani (Suleymanov 2015:82), a factor that makes Schulze's (2000:47) claim that the Shuvi pattern is an Azerbaijani borrowing less probable."

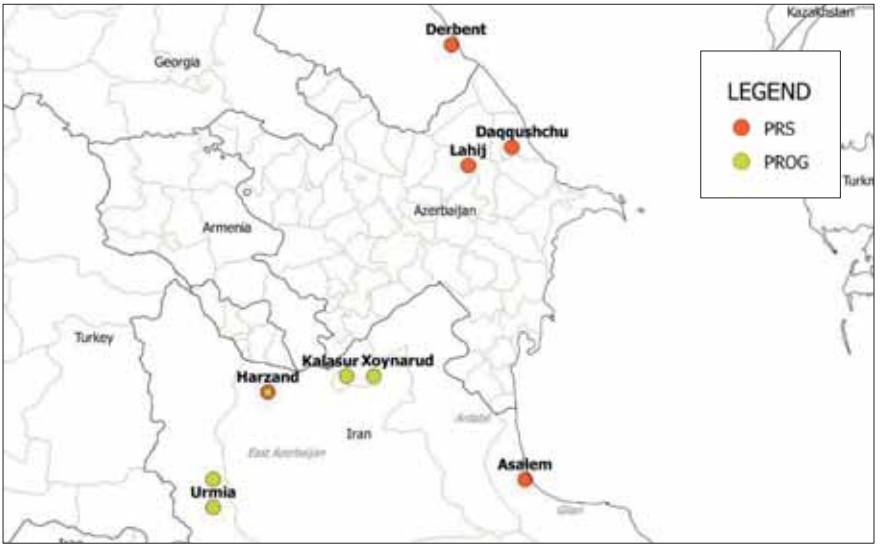


Figure 13. Infinitival patterns in some Neo-Aramaic and Tat varieties

Three main patterns are found: one pattern formed with the prefix *bV-* added to the infinitive verb and a copula; one with the infinitive verb and a copula; and one with *mV-* prefix (the same marker as the imperfective found in other Iranian varieties) and the verb in the infinitive plus a copula. The former two include both patterns with general present function as well as ongoing function, while the third is used for the present tense. It is then possible that these patterns have arisen from progressives. The *bV-* locative/allative prefix is assumed to have arisen from an adverb with the meaning 'away' (Utas 2013:257).

Examples of all three patterns are shown here. First, we see an example of the progressive construction in the Aramaic variety Christian Urmi (Afro-Asiatic) compared to the present tense construction in the Muslim Tat variety Lahiji (Indo-European) in 5:26 and 5:27, respectively.

5:26 Christian Urmi, Afro-Asiatic (Aramaic) (Hüseynova 2002:106 as given in Noorlander & Stilo 2015:441)

Progressive
+bi-graya=vin.[127]
LOC-read.INF=COP.1SG
'I am reading.'

[127] "The + sign indicates that all subsequent sounds of the word are pronounced with retraction of the tongue towards the back of the mouth, also known as backing, velarization or pharyngealization" (Noorlander 2017:195).

5:27 Lahıji, Indo-European (Muslim Tat) (Hüseynova 2002:106 as given in Noorlander & Stilo 2015:441)

Present
bœ-xund-ćen-um.
ALL/LOC-read-INF-COP.1SG
'I am reading/I read.'[128]

Second, we see an example of the constructional schema without the allative/locative marker in the Judaeo-Tat variety Juhuri (Indo-European), which also appears in Kalasuri [T/TAL]. In Juhuri, a constructional schema V-INF-COP is used for marking the present, while in Kalasuri [T/Tal] a similar schema is used as a progressive, as presented in 5:28 and 5:29, respectively.

5:28 Juhuri, Indo-European (Judaeo-Tat) (Noorlander & Stilo 2015:443)

Present
jćʕ gürd-én-üm.
fish take-INF-COP.1SG
'I catch fish.'

5:29 Kalasuri, Indo-European (Tati/Taleshi) (Yarshater 2005:278, as cited in Noorlander & Stilo 2015:443)

Progressive
olœt=em *šušt-e.*
clothes=COP.1SG wash-INF
'I am washing clothes.'

In Juhuri, a different pattern formed with a marker *ede*(*ye*)- is used as a progressive. The marker is either combined with the infinitive (*-den-* glossed as PRS), as in 5:30a, or the participle form of the verb as in 5:30b. The different verb forms probably reflect dialectal variation (Authier 2012:197).[129]

[128] Orig. 'I am reading'.
[129] Authier (2012:200) notes that Anisimov (1932) gives, as an alternative to *edey*, *edere* in brackets, which "suggests the thought" of a connection with the verb *deri* 'there is (in)'.

5:30 Juhuri, Indo-European (Judaeo-Tat) (Authier 2012:197, 194)[130]

a) *edeye* *tü=re* *xur-den-üm*.
FOCZ 2=DAT eat-INF-1[131]
'I am eating you.'

b) *imu* *ede-ym* *durun-de* *gendüm=e*.
1PL FOCZ-1PL harvest-PTCP wheat=DAT
'We are reaping the wheat.'

As mentioned in Chapter 1, several North Eastern Neo-Aramaic dialects are noted to have progressive patterns with proximative uses, in addition to the ongoing use (Noorlander 2017:195–198). An example from Jewish Urmi is given in 5:31 where the progressive pattern combines a non-finite verb form and a copula.

5:31 Jewish Urmi, Afro-Asiatic (Aramaic) (Garbell 1965:142; as quoted in Noorlander 2017:195)

+moi-t *+mitra* *ambola* *wélu-le* *dwíq-li-le*
water:PL-of rain carrying COP:3PL-him seized-OBL:1SG-him
"The rainwater was about to carry him off. [So] I seized him"

Third and finally, we see an example from the Muslim Tat of Daqqushchu (Indo-European), which forms the present tense by prefixing the *mV-* prefix to the infinitive construction. This constructional schema only differs from the Christian Urmi (Afro-Asiatic) and Muslim Tat of Lahij (Indo-European) by taking the *mV-* prefix instead of the *bV-* prefix.

5:32 Daqqushchui, Indo-European (Muslim Tat) (Grjunberg 1963:132, 234, as cited in Noorlander & Stilo 2015:443)

Present
mæn *raft-cén* *mu-xast-cén-üm.*
I go-INF IPFV-want-INF-COP.1SG[132]
'I want to go.'

[130] The original translations are: a) *Je suis en train de te manger* and b) *Nous sommes en train de moissonner le froment*.
[131] Orig. eat-PRS-1.
[132] Orig. DUR-want-INF-COP.1SG.

5.3.4.3 Note on Nakh-Daghestanian

The varieties discussed in this section are neighboring languages spoken in southern Russia or northern Azerbaijan and belong to the Nakh-Daghestanian language family. In some Nakh-Daghestanian varieties, infinitive markers are used in imperfective or progressive constructions. In Lezgian (Nakh-Daghestanian), the imperfective *-zwa*/*-zawa* is derived from the infinitive marker *-z*/*-iz* (Haspelmath 1993:130). Haspelmath (1993:140) notes that *-zwa*/*-zawa* "typically refers to progressive situations, i.e. processes going on at the time of reference". In a similar manner, the Udi (Nakh-Daghestanian) present and past imperfective forms are formed with the infinitive *-es* and an additional *-a* marker, realized as *-sa* where the *e* is lost. The progressive, however, which coincides with a future pattern, is formed with a marker *-al* (Schulze-Führhoff 1994:476–477).[133] Noorlander & Stilo (2015:443) also report an infrequent progressive formed with the infinitive and a dative marker realized as *-sax*.

Other Nakh-Daghestanian languages that were searched do not base their imperfectives or progressives on infinitives, but it is noted that many use periphrastic constructions. In Khinalug (Nakh-Daghestanian), for example, a copula auxiliary is used to "generate the overwhelming majority of tense forms" (Kibrik 1994:388); similarly, in Kryts (Nakh-Daghestanian) all tenses in the indicative (the distal past, "the concrete present"[134], general present, concrete future[135], general future), apart from the recent past, are formed with the copula *-i* (Saadiev 1994:425), and in Rutul (Nakh-Daghestanian), a durative stem combines with two types of auxiliaries for forming the present, progressive, past imperfective and general past (Alekseev 1994a:231). Similar constructional schemas were found in Budukh (Nakh-Daghestanian) (Alekseev 1994b:278). The infinitival patterns, then, exist in analogy to other periphrastic forms with imperfective functions and expanding beyond that.

There are, then, languages neighboring some of the varieties with DAR (and to a lesser extent KAR) patterns that have patterns for the imperfective domain that include infinitive forms. This could suggest influence via contact. Other Nakh-Daghestanian varieties have patterns that include copulas or auxiliaries for the imperfective domain.

5.3.5 Discussion and conclusion of section 5.3

The map in Figure 14 gives an overview of the distribution of the constructions discussed in this section. The map shows more than 60 data points including the 50 varieties having DAR and KAR patterns as well as the Iranian and non-Iranian

[133] Recall that a pattern marking both the ongoing and the future was noted for Southern Kisi (Niger-Congo) in Chapter 2, section 2.1.2.6.
[134] Examples look like progressives.
[135] Examples look like future progressives.

varieties discussed in section 5.3.4. Almost every data point on the map is a city or a village. Only a few data points refer to areas rather than locations, namely those referring to Northern Taleshi, Azerbaijani and Turkish. The map mainly aims at showing functional division between patterns in Iran and in Azerbaijan. In Figure 14, yellow indicates progressive pattern, red indicates present or past imperfective pattern, and blue indicates patterns with proximative function. Those patterns that have present tense and past progressive function are given red colour. Turkish and Azerbaijani *-mAktA-*/*-mAKda-* patterns are orange since they are described as patterns with uses towards the imperfective. Triangles are DAR patterns, squares are KAR patterns and the star indicates hybrid patterns. More specifically, the groups refer to the following:

DAR: PROG; varieties with a progressive DAR construction.

DAR_3SG: PROG; varieties with a progressive DAR construction which is invariable in the 3SG.

DAR: PRS; varieties with a DAR construction for the general present tense function and in a couple of cases past progressives moving towards past imperfectives.

DAR: PROX; varieties with a DAR construction used for proximative function.

KAR: PROG; varieties with a progressive KAR construction.

KAR: PRS; refers to one variety, namely Charozh [TAL], in which the KAR construction is used as the present tense and past progressive.

KAR: PROX; refers to one variety, namely Ebrahim-abadi, in which the KAR construction is used for the proximative function.

DAR-KAR: PROG; varieties with a progressive pattern containing both DAR and KAR elements.

IRANIAN_INF: PROG; Iranian varieties with progressive patterns that do not belong to the DAR or KAR gram family and include infinitival forms.

IRANIAN_INF: PRS; Iranian varieties with present patterns that do not belong to the DAR or KAR gram family and include infinitival forms.

NON-IRANIAN_INF: PROG; non-Iranian varieties with progressive patterns including infinitival forms.

NON-IRANIAN_INF: PRS/IPFV; non-Iranian varieties with present or imperfective patterns including infinitival forms.

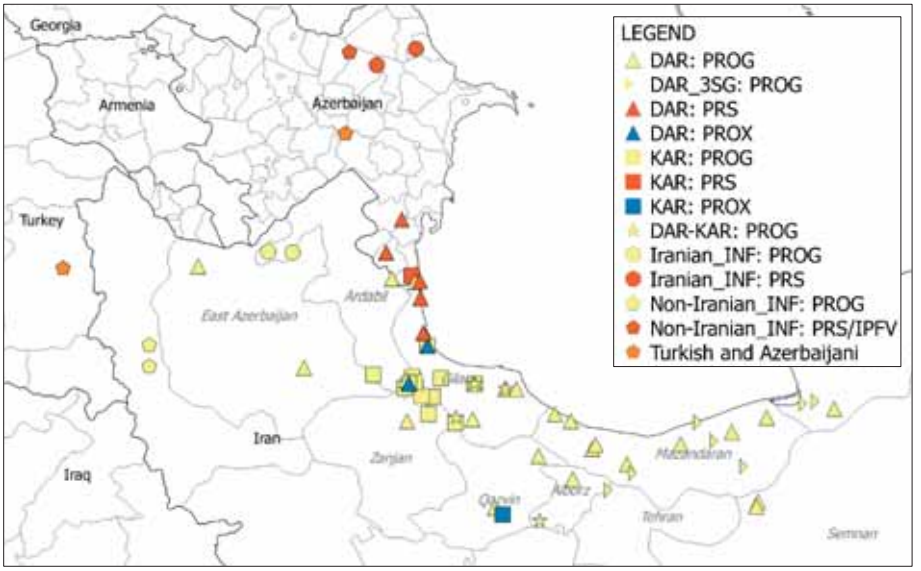


Figure 14. Iranian and non-Iranian patterns and functions

The map shows functional divisions irrespective of type of pattern or genealogical affiliation. Functionally, the varieties in and close to Azerbaijan have patterns with present tense and/or past imperfective function, while the patterns within Iran are mainly progressives. It may be that this division reflects influence from patterns existing in the dominant languages in Iran and Azerbaijan, namely the Persian *dāštan* construction which is a progressive and the Azerbaijani -(*y*)*Ir*- (the counterpart to Turkish -*Iyor*-) as well as -*mAKdA*- which are described as present tense and "a more focal present", respectively (Schönig 1998:254).

As discussed in Chapter 1, section 1.2.3, patterns in earlier stages of grammaticalization, such as periphrastic patterns, are more likely to be borrowed than more mature patterns (Dahl 2004:127). Strikingly, the data from this region shows that almost all varieties have a dedicated progressive, or former progressive, pattern. This gives the impression that the progressive construction is "highly contagious", in the sense of Dahl's cross-linguistic observation. Although many other Iranian languages also have progressive patterns, the concentration of patterns in this region is likely to be areally motivated.

> Looking at the synchronic distribution of various grammatical phenomena as the result of grammaticalization processes, a conclusion that is hard to avoid is that grammaticalization is highly contagious. In other words, while the chance that a certain morpheme or construction in a language will undergo a particular

> kind of grammaticalization is on the whole rather small, the probability increases dramatically if a neighboring language undergoes the process in question. In the majority of such cases, the languages involved are more or less closely related, but if the external conditions are the right ones, also totally unrelated languages may be affected. (Dahl 2001:1469)

Thus, the process of grammaticalization in one variety may inspire a similar process, and/or contribute to the retaining of a similar pattern and/or speed up an ongoing grammaticalization process of a similar pattern in a neighboring variety. Also, as noted in Chapter 1, section 1.2.3, patterns that are shared among languages and that are in contact increase in frequency or productivity.

I cannot draw any certain conclusions regarding the origin of the DAR patterns, i.e. which of the varieties that is most likely to have grammaticalized the pattern first, nor regarding the direction of borrowing between these varieties. For the KAR patterns, the data suggests that their origin is Tati, as discussed below. In what follows, I will also discuss some matters related to contact-induced change, grammaticalization of the progressive as well as the problem of establishing an origin and direction of borrowing among these patterns in light of the data presented in the map in Figure 14.

5.3.5.1 The DAR gram family

5.3.5.1.1 The areal cline

The map in Figure 14 shows an areal grammaticalization cline in the DAR gram family. Looking at the distribution of DAR constructions, we observe progressive patterns in the east and general imperfectives in the west. It is as if the more westwards we move, the further along the grammaticalization cline of the DAR construction we get: the DAR construction changes from periphrastic to inflectional, from unbound to bound, and from functioning as the progressive to functioning as the general present and past imperfective, and thus becoming obligatory. The fossilized third person singular DAR element is an eastern Mazandarani feature.

At one pole of our areal cline, we find the Mazandarani periphrastic progressive pattern, and at the other pole we find the most mature construction in the Taleshi varieties. It is tempting to assume that one of the Taleshi varieties is the source of this areal cline due to its level of maturation, the most striking one being that it is the obligatory present tense construction and on its way of becoming a past imperfective in some Taleshi varieties. However, the most mature pattern is not necessarily found at the location where the development started. It is, for example, possible that the development to present tense function of the DAR patterns has been triggered by patterns in the surrounding varieties.

5.3.5.1.2 Taleshi DAR patterns

The data suggests that the Taleshi present tense patterns originated as progressives, which was previously noted by Windfuhr (1989b:256). This is evident in the data presented in this chapter in several ways. Most notably, the functional division between present and past for the same constructional schema in various Taleshi varieties suggests that the pattern originated as a progressive. This means that in Taleshi varieties the shift from progressive to imperfective seems to have happened, or be happening, in the present tense first and then spread to the past. For the Taleshi varieties Leriki and Shuvi, the shift from progressive to imperfective has also partly happened in the past. In addition, we know that locative sources, especially those in combination with non-finite verb forms, are cross-linguistically common patterns for marking the progressive (see, e.g., Bybee et al. 1994:128). This further strengthens the argument that these are former progressives. This, however, does not mean that all DAR patterns in Taleshi originated from progressives, since it is possible that one pattern already grammaticalized to the general present in one variety is borrowed as such, or that the shift in meaning occurred at the same time as the borrowing. The grammaticalization to imperfective (or shift to imperfective in the borrowing process) may also have been influenced by the functions of the Azerbaijani -(*y*)*Ir*- and *-mAKdA-* patterns.

5.3.5.1.3 Origin of DAR gram family

Table 25 is a shorter version of Table 23 and summarizes the types within the DAR gram family. Here, Types 3 and 4 are merged in the last row referring to all Taleshi varieties. As can be seen, in the Mazandarani and some Central Caspian varieties, we observe periphrastic DAR constructions, while in Gilaki, Taleshi and some Central Caspian we instead have inflectional or clitic DAR constructions. In the Mazandarani varieties, the DAR element is preposed and can be separated from the main verb by several elements. In the other varieties, the DAR element is postposed, always directly following the verb. The Mazandarani DAR constructions involve finite main verbs, while the Gilaki, Taleshi and Tati DAR constructions (almost always) involve infinitive forms of main verbs.

Table 25. Short version of Table 23

FUNCTION	SCHEMA	VARIETY
PROG	DAR + V_{fin}	Mazandarani, Central Caspian
PROG	V-INF(-)DAR	Gilaki, Central Caspian, Tati
PRS, PST PROG/PST IPFV	V-(INF=/-)DAR(=COP)	Taleshi

In what follows, I will discuss contact-induced change likely to have given rise to DAR patterns, as well as the source and direction of such changes, and some problems that arise when one attempts at establishing such.

As a start, we can rule out Tati as the source of the DAR pattern since the Tati varieties mainly exhibited KAR patterns and are a minority within the DAR gram family. Regarding the Taleshi V-(INF=/-)DAR(=COP) schema and the Gilaki/Central Caspian V-INF(-)DAR schema, it is likely that V-INF(-)DAR and V-(INF=/-)DAR(=COP) have arisen as a result of borrowing or calquing, due to the similarity in form and since we know that some of these varieties are in contact. Regarding Taleshi and Gilaki, also inheritance from a mutual substrate language is perhaps a possibility. The alternative assumption that each village grammaticalized its own pattern, which then emerged as both structurally and functionally similar to the surrounding variety's patterns, is not plausible. It is, however, not possible to tell what has been borrowed or calqued. As was noted in Chapter 1, when it comes to related languages, it may not be possible to distinguish between matter and pattern borrowing since both yield the same result.

The schema DAR + V_{fin} in Mazandarani and Central Caspian and the rest may also have arisen as a result of contact, since we know that Mazandarani, Central Caspian and Gilaki are in contact, and the DAR elements in these patterns are built on locative copulas. If so, however, due to the constructional differences, they cannot be assumed to have arisen through borrowing or calquing in the same way as the Gilaki/Central Caspian and Taleshi patterns. Instead, the development must have been somewhat more independent. In this scenario, these patterns must have arisen as a consequence of contact-induced grammaticalization of some sort. In Chapter 1, section 1.2.3, it was noted that speakers can only replicate paths of grammaticalization using similar components as the model language if such components are still transparent in the model pattern. The locative copulas in Gilaki, Central Caspian and Mazandarani are such transparent components.

One complicating factor is the synchronic status of the Gilaki locative copula. Rastorgueva et al. (2012:134) only found the locative use of the locative copula in the third person singular present tense form ("there is") in their Gilaki data. Also, a consulting native speaker of Rashti does not recognize the locative use of the locative copula, not even the "there is" use but only its use in the progressive pattern. Thus, the independent locative copula use does not seem to be very common. If Gilaki has grammaticalized its pattern modeled on the Mazandarani and/or Central Caspian patterns, it must have done so at a time when this locative copula was used in its original meaning more frequently, or alternatively, Gilaki must have revitalized the locative copula for the creation of the progressive under the influence of the surrounding varieties. Or, as a third alternative, the Gilaki pattern is the source of these patterns and the progressive was created as the locative copula was still in use. The last scenario is not completely unlikely, as the Gilaki pattern has tighter boundaries than the Mazandarani/Central Caspian patterns, which could indicate that it is older. In addition, it is situated in the center

of the area where DAR patterns are found, which could add to the probability of it being the source variety from which the DAR pattern have spread upwards to the Taleshi varieties, eastwards to the Mazandarani varieties and to the surrounding Tati varieties. Note also that, from the perspective of Mazandarani speakers, the locative copula in the Gilaki progressive pattern could easily be recognizable, since their own locative copula is still used, while from the Gilaki perspective the locative copula is no longer transparent.

The Mazandarani progressive is somewhat peculiar since it consists of two finite forms. In fact, this pattern is strikingly similar in its constructional schema to the Persian *dāštan* construction. I will discuss the possibility of influence between Persian and Mazandarani in detail in Chapter 6. Here I will mention that even if the Mazandarani pattern is a replication of the Persian *dāštan* pattern, it is still possible that it was also influenced by the surrounding varieties in some way. We can assume that if neighboring varieties have a progressive built with a locative copula *dər-*, this will increase the probability that Mazandarani creates a progressive built with an (identical) locative copula, even if the main model is the Persian *dāštan* construction. The frequency of use as well as the likelihood of spread and preservation of the Mazandarani pattern is also increased if the surrounding varieties have a progressive.

The Mazandarani varieties have patterns that vary between one where person-number is marked on the locative copula and one where the locative copula is frozen, invariable, in the third person singular. Also, in, e.g., Ziarati [M] the DAR element is always in the present tense. Stilo (forthcoming a.) assumes the pattern with the invariant 3SG to be the 'true' Mazandarani pattern meaning that the invariable pattern is the original, older pattern, while the marking of person number is a more recent development under Persian influence. Interestingly, Rastorgueva et al. (2012:134, 145) list the past forms of the Gilaki locative copula *dər-* as invariant for third person singular, i.e. *dərə*.

The situation is then complicated. I do believe, nevertheless, that among these varieties it is unlikely that the borrowing or grammaticalization process was not influenced by the neighbors in one way or another.

5.3.5.2 The KAR gram family

Most varieties with a KAR progressive are Tati, which indicates that the origin of the KAR gram family is Tati. Similar to the DAR grams, the KAR grams show structural and functional features which can be assumed to have been influenced by language contact. One example is the KAR pattern in Charozh [TAL] spoken in Azerbaijan, which has a present tense function unlike other KAR patterns. In the past, this pattern is a past progressive. All other varieties with KAR patterns are concentrated around the meeting point of Gilan, Zanjan, East Azerbaijan and Ardabil. The Charozh and Asalemi [TAL] KAR $V_{non\text{-}fin}$ patterns are also structurally close to DAR patterns in the area as they combine with a non-finite form of the verb, as will be illustrated in Table 26. In addition, the hybrid patterns

discussed in section 5.3.3.2 are the outcome of borrowing between DAR and KAR patterns. It is not possible to determine the direction of influence, nor if there was only one direction of influence. Nor is it possible to determine why the spread of the KAR gram family has not been as successful as the spread of the DAR gram family.

5.3.5.3 Proximative patterns

In three cases presented in sections 5.3.2.3 and 5.3.3.1, the DAR and KAR patterns have proximative function rather than being progressives. The proximative patterns are found around the meeting point of the DAR and KAR gram families. In Asalemi [TAL] and Karani [T], the DAR patterns have a proximative function, while another KAR pattern exists functioning as the progressive. In the borrowing process, the two patterns have taken on different functions, dividing the labor as it were. The fact that most of the varieties with DAR patterns, including those surrounding Asalemi and Karani, are progressives suggests that the proximative is not the original function of DAR constructions. It is then likely that the proximative function has arisen as a shift in meaning from the progressive in the borrowing process. As presented in Chapter 1, section 1.2.3, it has previously been noted that borrowed or calqued patterns may have slightly different meaning than the model pattern. It was also shown in Chapters 3 and 4 that there is a functional link between the progressive and the proximative in that progressive constructions typically have proximative uses when applied to events viewed as achievements. This overlap in function could explain why new patterns built on progressives assume the function of proximative in a borrowing process.

5.3.5.4 Infinitival patterns

Moving beyond the DAR and KAR gram families, and including the Tat, Aramaic and Turkic varieties discussed in section 5.3.4, we observed infinitival patterns that are formed in similar ways with progressive or imperfective functions. It is not certain that these patterns are related since using the infinitive for marking progressive is common cross-linguistically, or if they are, in what way.

Table 26 illustrates DAR and KAR patterns with non-finite forms that are used for the present and/or past imperfective and which have a structural equivalent in the present and/or past progressive in Iranian, Aramaic and Turkic. The Nakh-Dagestanian varieties are not included since data on their patterns is very limited. In this table, stem alternations and person-number marking are disregarded.

The table could be an indication of the imperfective patterns in Tat, Aramaic and Turkic having a progressive origin, but these developments do not need to be related to the Gilaki, Tati and Taleshi varieties. However, similar to the DAR and KAR patterns, neighboring varieties may still have influenced each other in some way. As already discussed, generally speaking, we know that if surrounding languages have a particular pattern for marking a particular function, the

neighboring variety may borrow a similar pattern (from that or from a different variety), or create an equivalent pattern, or influence an existing pattern with a similar function so that it is (further) grammaticalized, retained and/or so that it increases in frequency and productivity. Thus, in one way or another, contact will increase the likelihood of finding similar patterns in neighboring varieties.

Table 26. Infinitival schemas used for imperfective and progressive

INFINITIVAL SCHEMAS	USED FOR THE PRS IPFV AND/OR PST IPFV IN VARIETY	USED FOR THE PRS PROG AND/OR PST PROG IN VARIETY
bV-V-/.INF-/(=COP)	Kalasuri [T/TAL] Xoynarudi [T/TAL] Asalemi [TAL] Lahiji (Muslim Tat)	Christian Urmi (Aramaic)
mV-V-INF-COP	Daqqushchui (Muslim Tat)	
V.INF-/=COP	Juhuri (Judaeo-Tat)	Jewish Urmi (Aramaic)
=COP V-INF		Kalasuri [T/TAL] Xoynarudi [T/TAL]
V-INF-DAR	Jokandani [TAL] Anbaran Mahalle [TAL] Shuvi [TAL]	Rashti [G] Lahijani [G]
V-INF-LOC(*da*)	Turkish (Turkic)[136] Azerbaijani (Turkic)	
V-INF-DAR-COP	Shuvi [TAL]	Chali [T] Jokandani [TAL] Viznei [TAL] Anbaran Mahalle [TAL]
V-INF(-)KAR(-)DAR		Lahijani [G] Rashti [G]
KAR V-INF-DAR		Rashti [G]
KAR V-INF	Charozi [TAL]	Charozi [TAL] Asalemi [TAL]
damand=PN V-INF		Asalemi [TAL]

[136] The Turkish and Azerbaijani patterns have previously been analyzed as patterns having uses towards the imperfective.

5.4 Progressive vs. present and past imperfective

This section further examines grams within the imperfective domain in Mazandarani, Gilaki, Tati and Taleshi, more specifically, it examines the present tense, the past imperfective, the present progressive and the past progressive grams, in their affirmative and indicative forms. The imperfective domain may contain other grams in certain languages, such as habituals, ingressives or proximatives, these are not included in this section. Also, in the languages examined here, such grams are not often mentioned in the available descriptions.

This section includes 21 Mazandarani, Gilaki, Tati and Taleshi varieties. The remaining varieties included in section 5.2 are not included since they lack the necessary grammatical descriptions. This section only addresses varieties that have a dedicated progressive pattern, meaning that Leriki [TAL] and Shuvi [TAL], which do not have a pattern mainly used as a progressive, are excluded. In Appendix G, all varieties and the constructional schemas under examination are given in alphabetical order.

The data presented in section 5.3 show great variation among the progressive patterns, and it was concluded that the progressive is, as expected, highly contagious. In addition, several varieties have more than one progressive pattern. We have also seen examples of how constructional schemas are mixes of two patterns. The impression is that the variation of marking the progressive is great in several ways. As discussed in Chapter 1, section 1.2.3.3, previous studies have postulated that, synchronically, we expect to have more competing patterns in the beginning of a maturation process than at later stages (see illustration in 1:17). In addition, periphrastic patterns were noted to be more borrowable than more mature patterns (Moravcsik 1975:110; Dahl 2004:127), which enhances the number of periphrastic patterns in comparison to inflectional ones synchronically. Therefore, in what follows, I will answer the following questions: is the variation of the marking of the progressive gram in the Mazandarani, Gilaki, Tati and Taleshi varieties greater than that of the present tense and past imperfective grams? This is done by comparing the number of progressive grams in the present and past to the number of present and past imperfective grams, as shown in 5:33, both across varieties and within varieties. The data presented in this section will show that the answer to this question is affirmative. I will also investigate how the schemas found are divided in their covering of the imperfective domain across these languages, i.e. whether a schema can be a progressive in one variety but an imperfective pattern in another. In addition, a general discussion on the structure of these patterns is also made.

5:33 Comparison of grams within the imperfective domain

PRS + PST IPFV vs. PROG (PRS & PST)

Among the varieties studied here, it is often the case that the present/non-past tense and past imperfective are based on the same constructional schema, which varies for tense. In Chali [T], for example, the present tense pattern is *me*/*mi*-V.PRS-PN, whereas *me*/*mi*-V.PST-PN expresses the past imperfective. We can then view *me*/*mi*-V as an imperfective gram with present and past forms. Although this is the most common case among these varieties, there are varieties that have different constructional schemas for these functions: in Rashti [G], for example, the present-future schema is V.PRS-PN, whereas the past imperfective schema is V.PRS-*i*- PN.

Often, patterns that are used for the present tense are also used for referring to the future, regardless of whether there also exists a separate future pattern. Therefore, grammars may refer to these forms as non-past or similar rather than present tense. In what follows, the possible future use of these grams is ignored, and I will refer to these as present tense grams.

A pattern used for the present tense includes the function of the present progressive, and a pattern used for the past imperfective function includes the function of the past progressive, but not the other way around. Also, a present tense patterns in these languages is the same as a present imperfective pattern. In the past, however, the varieties discussed here have a division between the past imperfective and a past perfective.

In section 5.2, only DAR and KAR progressives were looked at but the variation is greater if we include other progressive patterns as well, these additional patterns are included in the present section. In the Taleshi variety spoken in Asalem, for example, the present progressive is marked by *kâ(r)* + the verb in the infinitive as seen in 5:34a). In the past, the same constructional schema is found with the addition of a past copula, seen in 5:34b). But an additional past progressive constructed with *damand* 'stay in' and the infinitive also exists along with the past KAR construction, where *damand* is analyzed as consisting of the preverb *da*- 'to, into' and *mande* 'stand' (Paul 2011:126, 267). This construction, given in 5:34c), does not exist in the present tense. In addition to these patterns, there is a DAR pattern that is used mainly with proximative function. Unfortunately, no example is available for this pattern.

5:34 Asalemi, Taleshi (Paul 2011:126, 128, 129)

a) *a* *kâr=a* *hard-e*.
3SG KAR=3SG[137] eat.PST-INF
'He is eating.'

[137] Orig. PROG=3SG.

b) *əm* *ruj* *səb-i=râ* *cə=b-iš* *kâ*
DEMP day morning-OBL=LOC what=AUX-2SG KAR[138]

kard-e?
do-INF
'What were you doing this morning?'

c) *damand=a* *əštan* *a* *zua* *nava=râ*
PROG=3SG self DEMD boy grandchild=for

lailai *vât-e.*
lullaby say-INF
'She was singing a lullaby for that grandson of hers.'

In Ziarati [M], both a DAR gram (involving a DAR (*dar*-) element invariable for tense and person and number) and a *dāštan* 'have' construction exist. What is more, one speaker seems to use an invariable form of the *dāštan* verb, i.e. *dār-e* 'have-3SG', as if mixing the Mazandarani and Persian pattern in building on the Mazandarani model, where the DAR element is invariable in the 3SG, and also on Persian in using the *dāštan* verb instead of the locative copula *dar*-, thus creating a new pattern in her idiolect.

5:35 Ziarati, Mazandarani (Shokri et al. 2013:205 mg and transl.)

man *dar* *palu=ye* *in* *kāri* *ke* *dār-e* *mi-kārd-am* […]
I in beside=EZ this work that have-3SG IPFV-do.PST-1SG
'Along with the work I did […]'[139]

The comparison will not include verb forms with preverbs if such forms differ from non-preverb verb forms since there is not sufficient data on these verb forms especially on how they are formed in progressive constructions. This is unfortunate since, in a few varieties, the verb forms taking preverbs may differ quite a bit from forms that do not. In Asalemi [TAL], for example, the present tense without preverb is formed as *ba*-VII=PN1b whereas the present tense with preverbs is formed as PREV-PN1b-*a*-VII (Paul 2011:124).

Before moving on, a brief note on the marking of stative verbs is given in section 5.4.1. This is followed by the data presentation in which non-progressive and progressive grams with imperfective function in Mazandarani, Gilaki, Tati and Taleshi are given in 5.4.2. Section 5.4.3 summarizes and concludes this section.

[138] Orig. PROG.
[139] Perhaps a better translation is 'Along with the work I was doing […]'.

5.4.1 Note on stative verbs

In the investigation of the present and past imperfective patterns, the marking of stative verbs was excluded. There were several reasons for this. First, cross-linguistically, stative verbs tend not to combine with the progressive and would therefore create an asymmetry in the data collection. Second, often stative verbs, or certain stative verbs, have reduced morphological marking in certain paradigms in comparison to dynamic verbs. Also, as it turns out, the inclusion of constructional schemas used for stative verbs would not change the outcome of our investigation, meaning that the marking pattern used for stative verbs is always found elsewhere either within the same or in another variety. The inclusion of stative verbs in the investigation would increase the number of constructional schema for the present and past imperfective *within* a variety, but not *across* varieties.

The reduced morphology can at times be explained by the resistance to take on a new imperfective form. As a new present construction takes over functions of older presents, certain statives may resist this structural change. Such examples have previously been noted cross-linguistically (see Haspelmath 1998). It has previously been mentioned in Chapter 3, that in Persian, the verb *dāštan* 'have' has preserved the older Iranian structure V.PRS/PST-PN in that it does not take the imperfective *mi-* prefix, which is obligatory for other verbs (Lazard 1963, cited in Haspelmath 1998:43-44). In the Mazandarani variety Ziarati, we see the same structure where the verb 'to have' does not take the imperfective *mi-* prefix (Shokri et al. 2013:21). In example 5:36, the present tense forms of three verbs in Anbaran Ardebil are compared. While most verbs, such as the dynamic verb 'to eat' but also the stative verb 'to know', are formed with what has been shown to be a former progressive construction, the verb of existence does not take any marking; instead, a bare stem is combined with a floating copula clitic carrying person and number marking.

5:36 Anbaran Ardebil, Taleshi (Paul 2011:124, 170, 210)

a) *av* *ângivin=yə* *hâ=na.*
3SG honey=3SG eat=DAR[140]
'He is eating honey.'

[140] Orig. eat=LOC.

b) *avün zən=na nə-b-in ki*
3PL know=DAR[141] NEG-was-3PL CMPL

əm âmbu dəzdi-anin=e.
DEMP pear stolen-NEC=COP.3SG
'They were not aware that these pears are stolen'

c) *ca ləng-ə piu yâra=y hest.*
POSSD.3SG[142] leg-OBL up wound=COP.3SG exist
'There is a wound on his leg.'

5.4.2 Presentation of patterns within the imperfective domain

In this section, the patterns of present tense and past imperfective are compared to the patterns for present progressive and past progressive in Mazandarani, Gilaki, Tati and Taleshi.

5.4.2.1 Mazandarani

The varieties of Mazandarani make use of different progressive constructions while often sharing the constructional schema for the present and past imperfective. In several varieties, the general present is formed with the present stem, the marker *-n-* (which can be dropped under certain circumstances) and personal endings, i.e. V.PRS-(*n*)-PN, whereas the past imperfective is formed with the past stem followed by personal endings only, i.e. V.PST-PN (Stilo forthcoming a.). Thus, the past imperfective is unmarked and contrasted with the 'preterite', i.e. past perfective, which takes a prefix *bá-* (unless a preverb is present). The present tense is given for Dodanga'i [M] and the past imperfective for Kordkheyli [M] in 5:37 and 5:38, respectively.

5:37 Dodanga'i, Mazandarani (Stilo forthcoming a. mg)

de-n-e.
give-PRS-3SG
'S/he gives'

[141] Orig. know=LOC.
[142] It is unclear what the D in POSSD refers to.

5:38 Kordkheyli, Mazandarani (Stilo forthcoming a. mg)

pɛla pát-ɛnɛ.
rise cook.PST-3PL
'They would cook rice.'

In Amoli [M], instead of a DAR pattern, the *dāštan* construction is used. In Babolsari [M], a DAR construction is used invariable for 3SG, whereas in Saravi [M], the DAR construction can either be invariable or show person and number.

5:39 Amoli, Mazandarani (Stilo forthcoming a. mg)

dār-mə šú-mɛ.
have-1SG go.PRS-1SG
'I am going'

5:40 Babolsari, Mazandarani (Stilo forthcoming a. mg)

dar-ɛ šúmmɛ.
DAR-3SG go.PRS.1SG
'I am going'

5:41 Saravi, Mazandarani (Stilo forthcoming a. mg)

dar-me šúmmɛ.
DAR-1SG go.PRS.1SG
'I am going'

In Ziarati [M], the present tense and past imperfective take the imperfective *me-* prefix, *me*-V.PRS-PN and *me*-V.PST-PN, respectively. For expressing the progressive there are two constructions, at least in the present: the *dāštan* construction and the DAR construction.

Putting these patterns together, among the Mazandarani varieties we see three progressive constructional schemas, each with a present and past form, two present tense constructional schemas and two past imperfective constructional schemas. Also, in Ziarati [M] and Saravi [M], the *dāštan* construction and the DAR construction coexist, whereas no variety has two ways of marking the present or past imperfective. We then note greater variation when comparing the constructions used as progressives to the constructions used for the present and past imperfective, both across the Mazandarani varieties and within single varieties.

5.4.2.2 Gilaki

As was shown in 5:21, three constructional schemas exist for the progressive in Rashti [G]. In Lahijani [G], shown in 5:42, we find two constructions, a), which is structurally identical to 5:21b)[143], and b) which is similar to but slightly different from 5:21c). Stilo (Stilo in press) comments that *ka* is a nominal form of *kar*. In the Gilaki varieties, then, we have a total of four different progressive patterns.

5:42 Lahijani, Gilaki (Stilo 2001:663 mg)

a) *xord-á-dər-ə.*
eat-INF-DAR.PRS-3SG
'S/he is eating.'

b) *xord-e-ká-dər-ə.*
eat.PRS-INF-KAR-DAR.PRS-3SG
'S/he is eating.'

For the present tense, both Gilaki varieties employ the constructional schema V.PRS-PN. The past imperfective is formed with the constructional schema V.PST-*i*-PN for Rashti [G] and V.PST-PN for Lahijani [G].

5:43 Rashti, Gilaki (Rastorgueva et al. 2012:298, 312)

a) *šime zak-an či kun-idi?*
2PL.GEN child-PL what do.PRS-1/3PL
'What are your children doing?'

b) *ita duxtərbəčə dər-ə biǰa båzi kud-i-∅.*
A little.girl door-3SG side game do.PST-IPFV-3SG.PST
'A little girl was playing by the door.'

For the Gilaki varieties, then, the number of progressive constructions is greater within each variety as well as across the two varieties in comparison to the marking of present tense and past imperfective.

5.4.2.3 Tati

Yarshater (1969) looks at several dialects referred to as Southern Tati. In these varieties, the present tense is formed by adding the prefix *me-* to the present stem, followed by person-number marking, i.e. *me*-V.PRS-PN. In the past imperfective, the analogous constructional schema using the past stem is found, i.e.

[143] As noted in section 5.3.1.2, authors are inconsistent regarding the inflectional or non-inflectional marking of the Gilaki progressive pattern.

me-V.PST-PN. In two of the dialects that Yarshater examines, progressive constructions are found but not with the same constructional schemas: in Chali [T], the infinitive is followed by the locative marker *-u-* and a copula; in Eshtehardi [T], the present tense form is followed by KAR and DAR prefixed to the copula. There is no information about a past progressive construction in these varieties.[144]

5:44 Chali, Tati (Yarshater 1969:225 mg)

ešta tete *xord-an-u-ind.*
your dauthter.PL eat-INF-DAR-COP.3PL
'Your daughters are eating.'

5:45 Eshtehardi, Tati (Yarshater 1969:225 mg)

mi-vaz-em *kăr-dar-ima.*
IPFV-run.PRS-1SG KAR-DAR-COP.1SG.M?
'I am running.'

In both Hezarrudi [T] and Kajali [T], the present tense and past imperfective are marked in the same way as in the Southern Tati dialects, that is, as *m*(*e*)-V.PRS-PN and *m*(*e*)-V.PST-PN, respectively (Yarshater 1960:281–282; Yarshater 1970:462–463). Their present progressive patterns differ, however, in that the freestanding KAR element precedes the present tense verb in Hezarrudi [T], while it follows the present tense verb in Kajali [T]. Unfortunately, there is no information on the existence of a past progressive in these two varieties.

5:46 Kajali, Tat (Yarshater 1960:282 mg)

me-daji *kore.*
IPFV-ache.PRS.3SG KAR
'It is aching.'

In the Tati variety spoken in Shal, the constructional schema V.PRS-PN.PRS is used for the present (Yarshater 1959:59–60). The present tense has its own set of personal endings. In the past imperfective, the present stem is used but with a set of past imperfective endings, i.e. V.PRS-PN.IPFV. Unlike the other Tati varieties discussed here, Shali present and past imperfective do not take *m*(*e*)-. Tense in Shali, at least when it comes to the patterns under discussion, is thus expressed through person-number marking. The present and past progressive constructions

[144] There is, however, a mentioning of a progressive pluperfect (Yarshater 1969:233). This pattern falls outside the scope of this investigation.

are formed through the invariable *kerā* which is attached to the present and past imperfective forms, respectively.

In Kalasuri [T/TAL] and Xoynarudi [T/TAL], what is referred to as stem II is used for the present and past imperfective as well as the present progressive and past progressive (Yarshater 2005:274). Yarshater (2005:274) comments that stem II has replaced stem I for many verbs. The present tense is formed by adding *be*/*ba*- to stem II taking the infinitive marker -*e* followed by one of three sets of person-number endings, forming the constructional schema *be*/*ba*-VII-INF-PN. The past imperfective is formed in the same way, but takes an enclitic 'to be' serving as person-number marking, yielding the following constructional schema: *be*/*ba*-VII-INF=be.PST.PN. The only difference between the present and the past imperfective form is then in the marking of personal endings versus the enclitic copula. The present and past progressive are formed through the freestanding 'to be', in the present and past, respectively, and stem II + the infinitive marker -*e*. In 5:47, the present, past imperfective, present progressive and past progressives are shown in a), b), c) and d), respectively.

5:47 Kalasuri, Tati/Taleshi (Yarshater 2005:275, 277, 278 mg)

a) *ba-vrit-e-m*.
PREF-run.2-INF-1SG
'I run.'

b) *be-het-e=um*.
PREF-sleep.2-INF=be.PST.PN
'I used to sleep'

c) *enem* *gāndem* *kust-e*.
be.PRS.1SG wheat thresh.2-INF
'I am threshing wheat'

d) *enum* *olat* *šušt-e*.
be.PST.1SG clothes wash.2-INF
'I was washing clothes'

Looking at the present progressive, for which we have information for all varieties mentioned, it becomes clear that all varieties, apart from Kalasuri [T/TAL] and Xoynarudi [T/TAL], have their own unique constructional schema. Looking at the present tense and the past imperfective, constructional schemas are shared, either between Kalasuri [T/TAL] and Xoynarudi [T/TAL], or between Hezarrudi [T], Kajali [T], Eshtehardi [T] and Chali [T]. Shali [T] partly shares the same constructional schema with the latter group. A greater constructional variation is then evident when marking the progressive than in the rest of the present tense and past imperfective.

5.4.2.4 Taleshi

The constructions used for the imperfective function are somewhat more diverse among the Taleshi varieties. This is, at least partly, due to progressive patterns having grammaticalized into imperfectives. In some Taleshi varieties such as Anbaran Ardebil, Anbaran Mahalle, Viznei and Jokandani, the DAR construction is used for the general present tense and the past progressive. In varieties such as Masal-Sandermani [TAL], Masulei [TAL] and Asalemi [TAL], the present tense and past imperfective constructions contain neither a DAR nor KAR element, and the progressive construction is a KAR pattern. In Taleshi of Masal-Sanderman, the present tense and the past imperfective are based on what is referred to as stem I, which combines with different personal endings for the two patterns (Paul 2011:112, 124). The progressive is formed by adding *kərâ* to the present tense or past imperfective construction. In 5:48, the present tense, past imperfective, present progressive and past progressive constructions are given in that order.

5:48 Masal-Sandermani, Taleshi (Paul 2011:124 mg, 131, 127, 129)

a) *vin-i.*
see-2SG
'You see.'

b) *ziri* *cə* *šur-iri?*
yesterday what wash-2SG.IPFV
'What were you washing yesterday?'

c) *a* *kərâ* *har-ə.*
3SG KAR[145] eat-3SG
'He is eating.'

d) *ama* *kərâ* *râ=dəre* *š-imun.*
1PL KAR[146] way=SRCE go-1PL.IPFV
'We were going along.'

The Taleshi spoken in Masule differs slightly from Masal-Sandermani in that the past imperfective takes an *-i-* suffix (Lazard 1978). In Asalemi [TAL], the present tense takes an "allative prefix" *ba-* attached to stem II, and a special set of personal endings, i.e. *ba-*vII=PN (Paul 2011:124). Yarshater analyzes the *ba-* morpheme as a fused combination of *bə-* and the durative affix *-a*. Also, he notes that the personal endings are copulas that have lost their copulative sense in these patterns

[145] Orig. PROG.
[146] Orig. PROG.

(Yarshater 1996:101).[147] Paul (2011:130) provides a past imperfective form that is formed with the prefix *a* and stem I, as in *a*-vI-PN. Yarshater (1996:100), on the other hand, considers the past imperfective to be formed through the adding of *a*- to the verbal stem followed by *-i-*, i.e. *a*-vI-*i*-PN.[148] There are three different progressive patterns: one present progressive and two past progressive patterns. The present progressive is formed with *kâ(r)* as in *kâ(r)*=PN vII-INF. In the past, there is a past progressive that is similar to the present progressive, *kā* COP-PN vII-INF, and one that is formed with an element *damand*, as in *damand*=PN vII-INF. As already mentioned, *damand* does not have a present progressive correspondent. The present tense, past imperfective and the three progressive constructions are given in 5:49 in that order (examples 5:49d) and e) are repeated from 5:34b) and c), respectively).

5:49 Asalem, Taleshi (Paul 2011:142, 106, 128, 128, 129)

a) *a* *bamən* *pul* *ba-dâ=y*.
3SG 1SG.IO money PRS-give=3SG
'He gives me money (e.g. today).'

b) *zua-e* *per-a-p-in*.
boy-PL PVB-PVB-get.up-IPFV.3PL[149]
'The boys used to get up.'

c) *cərâ=š* *kâ* *bəraməst-e?*
why=2SG KAR[150] cry-INF
'Why are you crying?'

d) *əm* *ruj* *səb-i=râ* *cə=b-iš* *kâ*
DEMP day morning-OBL=LOC what=be.PST-2SG[151] KAR[152]

kard-e?
do-INF
'What were you doing this morning?'

[147] Paul (2011:112) comments that in both Asalemi and Anbaran Ardebil, the set 1b enclitics can also be used as copulas. However, since it is unclear to which other varieties this analysis also applies to, the set 1b endings are not marked as copulas in this thesis.
[148] In the comparison in section 5.4.3, both patterns a-v and a-v-i have been included, a-v is also found elsewhere.
[149] Orig. PVB-AUG-get.up-IPFV.3PL
[150] Orig. PROG.
[151] Orig. what=AUX-2SG.
[152] Orig. PROG.

e) *damand=a* *əštan* *a* *zua* *nava=râ*
PROG=3SG self DEMD boy grandchild=for

lailai *vât-e.*
lullaby say-INF
'She was singing a lullaby for that grandson of hers'

In the Taleshi spoken in Anbaran Ardebil, the present tense has the constructional schema VI=*na*=PN1b. The past imperfective is marked with a verbal prefix *a-* attached to stem II, as in *a*-VI-PN1a, but with a different set of pronouns (Paul 2011:129). The progressive, which only occurs in the past, has the schema =COP-PN1 V=*na* (Paul 2011:125). Paul (2011:155) comments that the morphology of Taleshi varieties such as Anbaran Mahalle, Jokandani and Viznei only differs from Anbaran Ardebil in their past progressive and simple past forms.[153]

The situation in Taleshi is more complex, as we can note more constructional variation for the progressive gram than for the present tense and past imperfective among certain Taleshi varieties, but not all. The upcoming section will compare all patterns discussed and conclude section 5.4.

5.4.3 Discussion and conclusion on section 5.4

The varieties discussed in this section and the constructional schemas used in these varieties for the present, past imperfective and present and past progressive are all given in Appendix G. The same patterns are presented in a more schematic form in Table 27. The schematization is done in a similar manner as for Table 26 but needs some further explanation:

a) Stem alternations, whether described as temporal distinctions or stem I and II are disregarded. This is motivated by the fact that stem alternations cut across the grams for present and past imperfective as well as present and past progressive.

b) The imperfective markers realized as *me-*, *m(e)-*, *mi-* are given as *mE-*. The prefix *be*/*ba-* is given as *bÄ-*. The past marker *-i/y-* is simplified to *-I-*.

c) In certain Mazandarani varieties, there is a present tense marker *-n-*. Since there are many exceptions where *-n-* is not applied, *-n-* is disregarded, absorbed by the verbal stem as it were. The keeping of *-n-*

[153] The example given for present tense in Jokandani and Anbaran Mahalle, however, seems to add an infinitive marker to the verb (Paul 2011:156), rendering the pattern V-INF=DAR. Since this pattern already exists elsewhere in the data, it does not affect the outcome in section 5.4.3.

would yield one more constructional schema, V-(*n*), in the left most column, and two more, DAR V-(*n*) and HAVE V-(*n*), in the right most column, thus 'favoring' the progressive count.

d) Since all constructions have person and number marking, it is disregarded (this includes the difference between using invariable 3SG or variable person-number marking in Mazandarani varieties).

e) Items glossed as the verb 'be' are given as COP.[154]

Table 27. Constructional schemas used in the imperfective domain

Constructional schemas for the present tense and the past imperfective		Constructional schemas for the present progressive and the past progressive	
both tenses	V	both tenses	DAR V
	mE-V		DAR *mE*-V
only present tense	V-DAR		V-INF-DAR-COP
	V-INF-DAR		V-INF=DAR=COP
	bÄ-V-INF		V-INF-KAR-DAR
	bÄ-V		V-INF-DAR
only past tense	V-*I*		KAR V
	a-V		KAR V-INF-DAR
	a-V-*I*		COP V-INF
	bÄ-V-INF=COP		HAVE V
		only present tense	*mE*-V KAR-DAR
			mE-V KAR
			KAR *mE*-V
			KAR V=INF
			HAVE *mE*-V
		only past tense	=COP V=DAR
			KAR V-*I*
			KAR COP V-INF
			damand V=INF
			V-INF KAR DAR COP

[154] The difference between personal endings and copulas is not always clear. However, in some varieties the distinction is important, for example, in Kalasuri and Xoynarudi where the difference between the present tense and past imperfective forms is that the former takes personal ending suffixes, while the latter takes copula suffixes. For such reasons, the copulas are kept as such in the table.

The difference between inflectional and clitical boundaries, where these are explicitly mentioned, has been kept in the table, but it needs to be noted that it is uncertain if all authors have intended to make this distinction.

Looking at Table 27, we can note that 10 different constructional schemas are used for the present and past imperfective among the varieties examined, and 20 constructional schemas are used for the present and past progressive. In addition, for four varieties we lack information about past progressive constructions, such data could have added progressive schemas to the table. Also, four Taleshi varieties, namely those of Anbaran Ardebil, Anbaran Mahalle, Jokandani and Viznei, lack present progressives most probably due to the fact that their present progressive constructions have evolved into general presents, an issue discussed in section 5.3.5. That is, the present tense markers V-DAR, V-INF-DAR, and *be*/*a*-V-INF and the past imperfective marker *be*/*a*-V-INF-COP are assumed to be former progressive patterns. We can also note that only in one case do we have a constructional schema, namely V-INF-DAR, which occurs in both the imperfective and the progressive domains.

Only four out of the twenty constructional schemas in the rightmost column, i.e. the progressive patterns, are not periphrastic. In contrast, almost all patterns in the rightmost column include inflectional boundaries. This is expected since periphrastic patterns are predicted to tighten as they grammaticalize into more general functions. Thus, the data shows that generally, the structural verbosity, i.e. phonetic weight that exceeds what would be minimally necessary (Dahl 2004:53), is greater among progressive constructions than among imperfective constructions. Exceptions are also found in Kalasuri [T/TAL] and Xoynarudi [T/TAL], which use a schema *be*/*a*-V-INF(-COP) for the imperfective and COP V-INF for the progressive. Conversely, the imperfective constructions have more uses in that they are typically used for the ongoing, habitual and generic functions. There is then a shift in load from phonetic material to function as patterns grammaticalize from progressives to the imperfectives.

Looking at the patterns used for the imperfective function several things can be noted. The affixes *a-* and *-i* are only used here and only in the past tense. However, they are not past tense markers but rather markers of past imperfective. In Rashti, for example, the V-*i* past imperfective pattern is contrasted with the past perfective *bV*(*/preverb*)-V pattern, and in Asalemi the *a*-V past imperfective pattern is contrasted with the V pattern used for the past perfective intransitive and V-*e* for the past perfective transitive.

It is also evident that the progressive constructions easily change in that we observe various combinations of the elements V, DAR, INF, KAR and COP. Mixes can also be observed among the imperfective patterns as in the patterns V-*i*, *a*-V and a-V-*i*.

We can further note that none of the KAR patterns in the table have generalized into imperfectives although we did see one such development in section 5.3.2.2 from Charozh [TAL] (this pattern was not included in this section due to lack of

data). There are also constructional schemas that are solely used for imperfective function, such as, e.g., *m*(*e*/*i*)-V, V, V-*i* and *a*-V-*i*. In some cases, these patterns combine with a DAR or KAR element in forming a progressive construction. Thus, the constructional schemas used for progressives are not typically new creations but are often built on constructions already existing for the present tense or past imperfective. This is exemplified with Shali [T] present tense and past progressive in 5:50.

5:50 Shali, Tati (Yarshater 1959:59–60 mg)

a) *š-i.*
go.PRS-2SG
'You go.'

b) *te* *kerā* *š-i.*
you KAR[155] go.PRS-2SG
'You are going.'

Looking at the variation within one and the same variety, we note that no variety marks the function of the present or past imperfective with more than one constructional schema.[156] However, some varieties use two or three competing constructional schemas for the present or past progressive. The data is thus consistent with the claim that there are more competing grams in the beginning of a maturation process. We have then observed more progressive gram types both among varieties and within one and the same variety.

5.5 Concluding Chapter 5

Both investigations in sections 5.3 and 5.4 in this chapter present data supporting the borrowable and volatile nature of the progressive. Section 5.3 confirms that the progressive is highly borrowable among the varieties Mazandarani, Gilaki, Tati and Taleshi. It was argued that the synchronic data suggests that most of these patterns have arisen through language contact. It was shown that progressives are often borrowed or calqued and typically undergo structural change, including cases where patterns are combined. The data also presents an areal grammaticalization cline for the DAR gram family. It was shown that as the structural pattern of progressives becomes more grammaticalized, the functional

[155] Orig. PROG.
[156] Some varieties, however, have different imperfective constructional schemas depending on transitivity. As already noted, there is no data on how transitivity interacts with the progressive patterns.

span of the pattern becomes more generalized. This investigation also shows that when it comes to volatile constructions such as progressives, areal typological studies are most fruitful when detailed dialectal information is available, since one cannot assume that two neighboring villages or cities that speak varieties often assumed to be the same will mark their progressive gram in the same way, or that the same constructional schema will mark the same function.

The data presented in section 5.4 shows that there are more progressive patterns than imperfective ones within one variety as well as among all the varieties examined, and confirms that there are more competing patterns in the beginning of a grammaticalization cline than later in the maturation process

6 The origin of the *dāštan* construction

The Persian periphrastic *dāštan* construction, mainly used in FOC ongoing function, was the topic of Chapter 3. In this section, in light of the data presented in Chapter 5, a possible borrowing process between this pattern and the patterns of the DAR gram family is discussed.

It has been pointed out in the literature that we have no data on the source of the *dāštan* construction in Persian (Dehghan 1972; Jeremiás 1993; Davari & Naghzguy-Kohan 2017; Nematollahi in press). Both Dehghan (1972:202) and Keshavarz (1962, cited in Nematollahi in press) have found no attestation of this pattern before the 19th century. Around the end of the 19th century, however, it is found in Persian, a dialect spoken close to Isfahan and in Gavrouni/Zoroastrian Dari (Nematollahi in press). Dehghan (1972:204), thus concludes that the construction is a fairly recent innovation. The earliest attestation reported by Dehghan, quoting Zhukovski (1888:376–377), is found in a popular folk song from the late 1870's, given in 6:1a). The next attestation of the *dāštan* construction appears in the satirical text *Čarand Parand* from 1906 by Mirza Ali Akbar Qazvini Dehkhoda[157], where examples in 6:1b) and c) are found.

6:1 Earliest attestations of the *dāštan* construction (Dehghan 1972:201, 203, mg)[158]

a) Taken from Zhukovski (1888:376–377)

poles [sic]	*sāz*	*mi-zan-e*	*dār-e*
police	instrument	IPFV-play-3SG	have.PRS-3SG

mi-raġs-e.
IPFV-dance-3SG
'The police are playing *sāz*, and are (in the process of) dancing'

[157] Dehkhoda (ca. 1879-1956) was a Persian scholar, poet and social critic living in Tehran whose parents were Caspian migrants from Qazvin (Saidi Sirjani 1994).
[158] Translation by Dehghan, my glossing and transcription.

b) Taken from Dehkhoda (1907)
ādam-hā=ye *āġā* *dār-and* *mi-āy-and.*
servant-PL=EZ gentleman have.PRS-3SG IPFV-come.PRS-3PL
'The gentleman's servants are (in the process of) coming.'

c) Taken from Dehkhoda (1907)
čubdār-hā *dāšt-and* *gusfand=e* *ziād=i*
sheep.dealer-PL have.PST-3SG sheep=EZ lot=IND

be šahr *mi-āvard-and.*
to city IPFV-bring.PST-3PL
'The sheep dealers were bringing a lot of sheep to the city.'

To my knowledge, the first mentioning of this construction in a grammatical description is in Phillott (1919:265), who notes that there exists a pattern of 'continuative and present signification' in the colloquial language and provides the examples given in 6:2a) and b). This suggests that the construction was well established in the language at least at that time.

6:2 Phillott (1919:265)[159]

a) *hanuz* *dār-ad* *mi-āy-ad*.
still have.PRS-3SG IPFV-come.PRS-3SG
'He is now coming.'

b) *dāšt-and* *mi-āmad-and.*
have.PST-3PL IPFV-come.PST-3PL
'They were coming along.'

As was noted in Chapter 2, examples with the verb 'come' can have a proximative or futurate reading. Whether the examples in 6:1b) and 6:2 can be understood in this way is unclear.

Nematollahi (2014), looking at the occurrences of this pattern in colloquial prose, notes an increase in the use of this pattern. Her table of results is reproduced in Table 28. Looking at present, past and what she refers to as evidential forms (e.g. *dāšte mixaride*) of the pattern, her data confirms the clear preference for present tense. There were no occurrences of evidential progressive forms in her data.

[159] Translation by Phillott (1919), my glossing and transcription.

Table 28. Re-production of table of results from Nematollahi (2014)[160]

COLLOQUIAL PROSE	YEAR OF PUBLICATION	GENRE	NR OF WORDS	NR OF *DĀŠTAN* CONSTRUCTION	FREQUENCY PER 1000 WORDS
Three Persian Plays	1890	play	23500	0	0
Čarand parand	1907-1908	satirical essay	25500	2 (1 PRS, 1 PST)	0,08
The collection of Jamālzāde's works	1921-1974	short story	169000	64 (41 PRS, 21 PST)	0,37
Two plays by Yaʿqubi	1998, 2010	play	21000	79 (66 PRS, 13 PST)	3,76

Scholars are not in agreement regarding the origin of this construction. Pistoso (1974) suggests that this pattern has come about due to Mazandarani influence, similar observations have also been made more recently (Jahani 2017:264). Interestingly, Nematollahi (in press) notes the occurrence of both the locative pattern and the *dāštan* pattern in Mazandarani texts from 1889 provided in Borjian (2008). Davari & Naghzguy-Kohan (2017), as well as Nematollahi (2014), however, suggest that this pattern has arisen through internal grammaticalized in Persian.

The historical data do not conclusively support either hypothesis. The first hypothesis has the advantage of assuming a cross-linguistically well attested grammaticalization path, in fact, by far the most common one, but has the disadvantage in that there is no data showing that the pattern existed in Mazandarani before being attested in Persian. I will look at each assumption in turn, starting with the first one.

6.1 Persian pattern modeled on Mazandarani pattern

The first hypothesis concerns change that has arisen due to structural as well as phonetic similarity between the locative elements included in the DAR gram family and the verb 'have', which is used for the progressive in Persian. For some varieties, the difference between the DAR element and the 'have' verb in the present tense is distinguishable only by the quality of the vowel, e.g. in Mazandarani *dār-*/*dāš-* 'have', which is identical to the Persian verb, in comparison to the Mazandarani *dar-*/*dayy-* locative copula. Tati varieties typically

[160] Re-produced with permission from Negar Nematollahi.

lack a possessive verb and can be ruled out as source varieties. The matching between the Persian 'have' verb and the Mazandarani locative copula in the present tense are given in bold in 6:3.

6:3	Persian	***dār-***	'have.PRS-'	*dar*	'in'
	Mazandarani	*dār-*	'have.PRS-'	***dar-***	LOC COP.PRS

Although supporting the other hypothesis, Nematollahi (in press) notes that if the pattern was replicated from Mazandarani into Persian this most probably happened in the present tense as the phonetic similarity is greater in the present tense than in the past. Interestingly, the progressive pattern in Mazandarani, unlike the one in Gilaki and Taleshi, is formed with a finite form of the verb inflected for tense and person-number similar to the Persian *dāštan* construction. In most Mazandarani varieties examined, the present tense form of the main verb does not take the prefix *mi-*, an exception being Ziarati, where the present tense may take the *mi-* prefix, rendering a progressive pattern even more similar to Persian. Since this is an exception among Mazandarani varieties and is probably a copy from Persian, it will not be viewed as relevant to the present discussion.

6:4 Mazandarani, Sarivi variety[161]

a) *maryem me pali dar-e,*
maryem my side LOC.COP.PRS-3SG

dar-e ruznāme xund-ene.
LOC.COP.PRS-3SG newspaper read.PRS-3SG
'Maryem is near me, she is reading the newspaper.'

b) *maryem me pali dar-e ruznāme xund-ene.*
maryem my side LOC.COP.PRS-3SG newspaper read.PRS-3SG
'Maryam is reading the newspaper near me.'

In example 6:4a), the first *dar-* functions as a locative copula, while the second functions as the marker of progressive. A path of grammaticalization for the locative progressive pattern is easily imaginable when assuming that an example like 6:4b) has arisen through a stage where this sentence entailed 'Maryam is (in a place) reading the newspaper'. Similar to the cross-linguistic suggestion made by Bybee et al. (1994:136), the locative meaning is later lost.

The assumption that Persian has grammaticalized its progressive inspired by the Mazandarani pattern has to assume an equation between the verb 'have' and

[161] Many thanks to Guiti Shokri for providing the examples and glossings.

the locative copula. Such a process could well have been a conscious choice, initially carrying a sense of irony or association of foreigness. The fact that *dāštan* 'have' is a stative and frequent verb, with a meaning often involved in grammaticalization processes, may have facilitated the process.

One main problem of assuming that Mazandarani is the source language lies in the fact that it today is viewed as having low status in comparison to the prestigious Persian. This has however not always been the case, in fact, historically, Mazandarani had a sufficiently high status to be written and to be used as a literary language.

> Of the living Iranian dialects, Mazandarani boasts the longest written tradition, roughly matching that of New Persian. This status was achieved in the long reign of the independent and semi-independent provincial rulers, commonly known as Ispahbads, during the centuries after the Arab invasion of Iranian lands. (Borjian 2004:291)

Additionally, Carina Jahani (p.c.) comments that there are indications of the existence of many competing languages spoken in Tehran around the 19th century, out of which Mazandarani was, and still is, an important dialect used in what is now the northern suburbs of Tehran.

In a study of the linguistic features of the south-central Alborz area (north of Tehran, also including what is today the northern part of greater Tehran), Borjian (2013a; 2013b; 2013c) investigates continuity between Caspian and Persian and shows that language convergence between Mazandarani and Persian dialects goes in both directions. In Borjian (2013b), which is concerned with Mazandarani influences on Persian dialects, 50 locations surrounding Mazandaran, mainly in the south-central Alborz adjoining northern Tehran are investigated. Due to migration trends from both north and south, and seasonal migration from the Caspian provinces into the south-central Alborz area, phonological, lexical, morphological and syntactic isoglosses between Persian and Caspian are found (Borjian 2013a:33–34). Borjian (2013a:46) distinguishes, among other things, a Perso-Tabaric[162] linguistic group stretching from Tajrish (today part of northern greater Tehran) in the south to the village Ruta in the north. He notes that regarding morphology, the varieties in the Perso-Tabaric group "resemble Tehrani Persian but with a rich Caspian coloring in the verb, which is distinguished by its preverbs and past tense forms, and by non-Perside 'be' and 'become'" (Borjian 2013a:31).

Nematollahi (in press) notes that ethnographic data available for Tehran around the 19th century does not mention a large group of Mazandarani speakers, and that 19th century travel notes indicate that travelling between Mazandaran and Tehran was not easy nor frequent. Also her search through a number of Mazandarani texts from 1830-1866 does not show any locative progressive

[162] Tabari is another term for Mazandarani.

patterns. She concludes that the replication must have happened from Persian to Mazandarani. On the other hand, we must assume that Persian and Mazandarani were in contact in the 19th century since both DAR and *dāštan* progressives are found in her search through Mazandarani texts from 1889. In addition, recall that searches in earlier texts by Dehghan (1972) has also failed to find the *dāštan* construction in Persian before the 1870's.

It seems then that we do not have a clear picture of the situation in Tehran and surroundings around 1900 at this time. Due to the structural similarity between Persian and Mazandarani, I believe that we need to assume that there has been some influence at work, either from Persian to Mazandarani or from Mazandarani to Persian.

6.2 Internal grammaticalization in Persian

Davari & Naghzguy-Kohan (2017), who do not comment on the possibility of external influence, provide a different hypothesis in which the *dāštan* construction has grammaticalized in Persian. In short, they suggest "a conceptual shift, in the form of metaphoric extension, from possessing a physical object to possessing the continuum of an action in a focal point of utterance" (2017:163). More specifically they suggest the following meaning changes: to hold > possess something (to be with something concrete) > to be with an activity or in the process of an activity (Davari & Naghzguy-Kohan 2017:174).

Their suggestion for the development of possession to progressive is not completely convincing. The data provided for their analysis are synchronic constructed sentences such as that in 6:5, which they note are ambiguous between denoting possession and ongoing use, and therefore constitute so-called 'switch context' examples in Heine's (2002) sense. In my opinion, however, the possessive meaning does not imply the progressive one, which is a requirement if the progressive has arisen from possessive meaning. Rather, the ongoing and the possessive meaning constitute two different sentences: one in which we have two clauses with *dāštan* as main verb followed by *xordan* 'eat' as main verb, and one in which we have the *dāštan* progressive construction. This is quite different from the Mazandarani example given in 6:4b), which uncontroversially can be assumed to have undergone a stage where the locative as well as the ongoing meaning co-occurred, as in 'Maryam is (in a place) reading the newspaper'.

6:5 Switch context according to Davari & Naghzguy-Kohan (2017:174)

āb=i	*ke*	*dār-im*	*mi-xor-im.*
water=INDF	that	have.PRS-1PL	IPFV-eat-1PL

'The water that we have, we eat it'

They also assume a grammatical development of progressive to "prospective", where "prospective" seems to refer to what in this thesis has been called proximative. Conversely, Nematollahi (2014:112) shows that in earlier fiction (mainly data from 1921-1974) the proximative use accounts for 35% of the occurrences of the *dāštan* construction in her data, and in later prose (1998, 2010) the proximative use constitutes 14%, meaning that the use is not uncommon in earlier period. The high number of proximative uses in Nematollahi's data leads her to shortly discuss a possible internal grammaticalization from proximative to progressive originating in light verb construction, such as *bar dāštan* 'set off' (Nematollahi 2014:112). The problem with this analysis is that *bar dāštan* 'set off', as given in her examples, does not combine with main verbs taking the *mi-* prefix, but the main verbs are in the preterite or the subjunctive, e.g. *bar dāšt-and telegrāf kard-and* 'they set off, telegraphed', and also that the shift from an inchoative 'set off' to proximative 'about to' has, to my knowledge, not been attested.

In the literature on grammaticalization, it is noted that continuatives may grammaticalize into progressives. Historically, *dāštan* used to have the meaning 'to hold'. In Middle Persian, from which Modern Persian descended, this verb was used "as a sort of auxiliary [verb] denoting the duration of the act" (Nyberg 1974:60).

6:6 Middle Persian (Nyberg 1974:60 glossed by Agnes Korn)

-m	*pat*	*drīnišn*	*dār-ēt*	*u*	*drīn-ēt*
1SG.PC	to	tearing	hold.PRS-3SG	and	tear.PRS-3SG

'He keeps tearing me up.'

A similar meaning is also found when this verb combines with a verb in the past stem as in *grift dār-* 'keep bound', *winārd dār-* 'to maintain in order (lit. hold arranged)' (Henning 1934:247, as cited in Korn 2017:44). As noted by Korn (2017:44), this pattern could also have played a role in the rise of the *dāštan* construction. Similar patterns with *δāray-* 'hold, have' are also found in the extinct language Khwarezmian (believed to have died out sometime after the 14th century), where *δāray-* combined with a nominal form seems to have a continuative sense, as in *āγrāci-hi δār-eda* 'he kept him awake' (Durkin-

Meisterernst 2009:336).[163] Due to the time span, however, it is hesitant to relate these patterns to the *dāštan* construction in Persian. Jahani (2017:278) comments that since it seems as if this pattern is a more recent development in Persian, it is likely that *dāštan* had already acquired the meaning 'to have' when the progressive pattern emerged, which makes the continuative to progressive development less probable.

Returning to the possible borrowing between Persian and Mazandarani, it is noted here, that if the Mazandarani DAR pattern is a replica of the Persian *dāštan* construction without influence from the other varieties discussed in Chapter 5, it should in principle not be part of the DAR gram family. As we know that Mazandarani has been in contact with neighboring varieties discussed in Chapter 5, a scenario with no influence is perhaps not very likely. Conversely, if the Persian pattern is a replication on the Mazandarani, it could, in principle, be part of the DAR gram family – even though it does not contain a locative marker.

6.3 Possessive sources for progressives

Although possessive sources for progressive patterns are rare, they are not unattested. For example, in Chapter 3, the construction with *gir* 'seize, trap' was noted to partly overlap with the *dāštan* construction. Some Germanic languages also have progressive constructions involving items with the meaning 'hold' (Ebert 2000: 607; Heine & Kuteva 2002: 184). Similarly, the Piedmontese dialect of Italian had a periphrastic construction formed with *tenere* 'hold, keep' together with the verb in the past participle marking the "continuative-iterative" in the 16th-century (Ricca 1998). The English *keep*, as in *keep doing something*, is obviously also an example of not progressive but continuative meaning. Furthermore, Heine and Kuteva (2002: 184-185) mention several languages outside of Europe with interesting constructions: in Imonda (Border), spoken in Papua New Guinea, *ula* 'keep, hold' is a "durative/intensity marker with durative verbs" and an "iterative marker with punctual verbs"; in the Waata dialect of Oromo (Afro-Asiatic), *(harka) k'awa* 'hold (in one's hand)' marks the "continuous aspect"; and in the Muduug dialect of Somali (Afro-Asiatic) **hayn* 'keep' is believed to have grammaticalized into an "auxiliary of durative aspect". Kuteva (2001:79), quoting Narumov (1987:98), provides an interesting example of an avertive gram in Galician (Indo-European) built on the verb 'have', given in example 6:7.

[163] Sangesari is noted to share features with an Eastern-Iranian continuum that included Kwarezmian (Azami & Windfuhr 1974:15-17, 36, Durkin-Meisterernst 2009: 337). In Chapter 5, it was shown that Sangesari has a DAR gram.

6:7 Galician, Indo-European (Kuteva 2001:79)

Houbemos de caer por aquel burato.
have.1SG.PERF[164] from/of fall through this hole
'We nearly fell down in that hole.'

An example from Sekpele[165] (Niger-Congo) is of particular interest. Sekpele has developed its progressive under the influence of Ewe. Both Sekpele and Ewe were included in the gram set used in the Bible corpus in Chapter 2.[166] They belong to the Kwa branch of the Niger-Congo language family and are spoken in West Africa. The Sekpele progressive is believed to have developed on the model of the Ewe progressive, which is formed with the verb *lè* 'be.at.PRS'. In Ewe dialects geographically close to Sekpele, the verb is pronounced *lɛ́* (Ameka 2006:130–131). The Sekpele and Ewe present progressive forms are given in 6:8.

6:8 Borrowing of progressive between Ewe and Sekpele

a) Ewe, Niger-Congo (Ameka 2006:131)
Kofi le mɔ́lĭ ɖŭ.
PRN be.at.PRS rice eat.PROG
'Kofi is eating rice.'

b) Sekpele, Niger-Congo (Ameka 2006:131)
li-kpefí nə́-mə́ lɛ́ wó ambe bɔ-kpɔ́-n-kó.
CM-child CM-DET hold 3SG mother CM-fight-LIG-ASSOC
'The child is fighting with his/her mother.'

Sekpele has, under the influence of Ewe, formed a present progressive using the verb *lɛ́* 'hold', which is phonologically close to the copula verb *le* in Ewe. Aikhenvald (2006:24) refers to this process as grammatical accommodation. This process would be very similar to an assumed scenario where the *dāštan* construction in Persian is created after being inspired by the locative copula/element in some Caspian language. The Sekpele replication shows that the availability for grammaticalization or borrowing in grammatical accommodation increases if the criteria phonetic similarity is met.

As noted in Chapter 3, the Persian *dāštan* construction has been borrowed into many languages and dialects of Iran. Some Kashkay (Turkic) varieties show interesting patterns, namely Abivardi, spoken in Shiraz province, and Galūgāhi,

[164] Probably have.1PL.PERF.
[165] Or Likpe, as given in Ameka (2006)
[166] However, the Ewe progressive in Chapter 2 is the *-ḿ* verbal suffix. Ameka (2006:131) notes that the *-ḿ* progressive is a pattern that is found in some Ewe dialects.

spoken in Mazandaran (Doerfer et al. 1990:vii; Csató 2005:277). These varieties have progressive patterns copied from the Persian *dāštan* construction. Since Turkic does not have a verb 'have', they make use of *var* 'exists' followed by a possessive suffix, i.e. a possessive construction, which agrees with the person of the inflected lexical verb (Csató 2001:117). Notably, Abivardi and Galūgāhi are not neighboring varieties. This means that these possessive progressive patterns must each have been copied from Persian independently, or they must have been copied at a time when these varieties where in contact.

6:9 Kashkay varieties (Turkic) (Csató 2001:117, partly mg)

a) Abivardi
var-im *gid-er-em*.
exists-POSS.1SG go-AOR-1SG
'I am just going.'

b) Galūgāhi
va'r-am *gäl-ị'r-äm*.
exists-POSS.1SG come-AOR-1SG
'I am coming.'

According to Csató, there are indications of this pattern being a Persian loan. If this assumption is correct, we can note that Abivardi and Galūgāh have calqued the Persian *dāštan* construction by using a verb of existence that is typologically a common path of grammaticalization for progressive patterns. If the hypothesis of Caspian origin is ever proven to be correct, these examples would entail a spiral of borrowing where a locative copula is matched with a verb 'have', which, in turn, is further matched to a verb of existence. In a similar manner, if the same assumption holds, the progressive formed with the *dāštan* verb in the Mazandarani variety Amoli is a borrowing that replaces the original locative construction, the same construction that gave rise to the Persian *dāštan* construction in the first place.

6.4 Concluding Chapter 6

Given the data at hand, I conclude that we cannot draw any definite conclusions regarding the origin of the Persian *dāštan* construction. The structural similarity between the Persian and Mazandarani patterns, however, suggests that the patterns in these languages have arisen through calquing between these languages, the direction of which cannot be established. This means that it is still possible that Persian is the source language of this pattern, but data, preferably historical,

is needed for such a claim to be convincing. Simultaneously, we cannot exclude the possibility of Caspian origin. A third possibility, that neither Persian nor any Caspian language is the source of this pattern, can also not be ruled out. One of the points speaking against the Persian pattern being the original one is that, to my knowledge, no other language has grammaticalized 'have' for marking the progressive cross-linguistically, although 'hold' progressives are attested. What speaks against both Persian and Mazandarani as source languages is that there is no attestation of these patterns in earlier texts, in the case of at least Persian, this issue is complicated by the fact that the progressive is restricted to the spoken language.

7 Summary and concluding discussion

7.1 Summary of findings

This dissertation has investigated a number of progressive grams from several perspectives. The most typical and favorable uses for progressives as well as their peripheral uses were investigated. The present and past uses were investigated with regard to the ongoing uses in present and past time reference, and also with regard to the shift from ongoing meaning to imperfective as it is manifested in the present and past tenses among a number of varieties found in the Caspian region. The data from the Caspian region, the data from Southeast Asia, as well as data on the borrowing of the Persian *dāštan* progressive within Iran, showed spread of progressives likely being a result of contact.

Chapter 2 presents a typological investigation of 89 progressive grams for which the most favorable contexts were presented and discussed. It was shown that in contexts with present time reference, the most favorable contexts refer to FOC ongoing events. At times, such contexts include attention-requiring elements. Contexts with past time reference often involved narrations. Here, the most favorable contexts for progressives were backgrounding contexts. In these sentences, the event to which the progressive applies is the background to another, often FOC and telic, event that pushes the narration forward. The progressive grams in the TED corpus are mainly used in reference to topical events or new changes in the world. Chapter 2 also show that generally, progressive grams have more uses with present time reference than with past time reference. This is expected in spoken language but was also shown for the Bible corpus, where past time reference is more common than present time reference. Several grams restricted to the present were noted, as well as grams with an especially high occurrence in the past. Peripheral uses found with certain progressive grams, such as uses in habitual and performative-like contexts, uses with stative verbs, and uses in contexts with future time reference were discussed. It was suggested that some of the grams discussed were grams with extended uses towards the imperfective.

Chapter 3 presents an in-depth investigation of the *dāštan* progressive in Persian. Here, it was shown that the pattern has a main use of FOC ongoing but also has peripheral uses such as proximative uses, futurate uses, iterative uses, and may occur in DUR contexts. Often, the contexts in which the pattern was encountered were dramatic, urgent, ironic or similar.

In light of some of the findings in Chapters 2 and 3, Chapter 4 discussed and suggested some explanations for the different uses and readings that arise when progressives occur in different contexts and apply to different types of events. Most notably, almost all uses discussed were directly or indirectly linked to the meaning of ongoingness at reference time. It was also suggested that when progressives combine with events viewed as achievements, a proximative reading arises. Also, it was noted that while a FOC reading is the default one if no explicit reference time is given, the duration of a DUR reference time is given explicitly.

Chapter 5 discusses the borrowing of progressive patterns and the shift from progressive to imperfective among a number of Northwestern Iranian varieties spoken along the Caspian Sea in Northern Iran and Southern Azerbaijan. The shift from progressive to proximative in some of these varieties was also noted. It was shown that the synchronic situation suggests a grammaticalization cline where these pattern have arisen through borrowing between varieties. In addition, the chapter shows that there are more progressive patterns than present and past imperfective ones both within one variety and across varieties. This is expected given that progressives are typically periphrastic, while the other gram types are more likely to be inflectional.

Chapter 6 returned to the *dāštan* progressive and discussed its origin. More specifically, it looked at the possibility that this pattern is a replication of the DAR patterns discussed in Chapter 5, or, has evolved as internal grammaticalization in Persian. It was concluded that the origin of this pattern cannot be established, but that the similarity in constructional schema between the Persian *dāštan* progressive and the Mazandarani DAR progressive suggests replication in either direction.

In what follows, a concluding discussion including prospects for future research is presented. First, temporal asymmetries are discussed, followed by a discussion on the uses of progressives which includes a refined description of the progressive gram type. Finally, some elaboration on the grammaticalization of the progressive gram type is presented.

7.2 Temporal asymmetries

Throughout this dissertation, temporal reference has been shown to be of importance for the progressive gram type, both in terms of temporal preferences and restrictions of progressives, and with regard to the grammaticalization of progressives as presented in Chapter 5. In this section, these temporal preferences and restrictions are discussed.

7.2.1 Temporal preference

Both Chapters 2 and 3 showed that progressives occur more often with present time reference than with past time reference in spoken language. This is not surprising since we can assume that we speak more about the present than about the past. The fact that present uses of progressives are in the majority in the Bible corpus, in spite of present time reference being less frequent than past time reference in Bible texts, however, suggest that it is a general tendency for progressive grams.

The preference for present time reference is most likely linked to the general difference in frequencies of present tense, past imperfective and past perfective cross-linguistically, where the past imperfective is less common in the past than the past perfective. This was illustrated in Figure 1 in Chapter 1. Assuming that the progressive most often operates on the imperfective domain, i.e. mainly has a distribution that is also covered by the present tense and past imperfective, the outcome is expected. The present time preference of progressives is then not specifically a feature of progressives, but a more general feature of the imperfective domain where past imperfectives are relatively uncommon. This, in turn, is linked to how we speak of events in the present and past: while utterances with present time reference have a typical interpretation of being imperfective, utterances with past time reference are typically interpreted as perfective. Also, as mentioned previously, in narration, which is a frequent genre in the past, telic events are needed to create change and bring the plot forward. These characteristics explain why past imperfective grams are less frequent than past perfective grams and consequently, why we find more present uses of progressives than past uses.

7.2.2 Temporal restrictions

The typological investigation in Chapter 2 presented several progressive grams that are restricted to present time reference. I will assume that these patterns are a structural result of the general tendency of the progressive gram type to occur more often in the present. At least one clear case of a future-progressive gram was also noted in Southern Kisi (Niger-Congo), this gram does not have past uses. In some cases, languages were shown to have both a present progressive and a past progressive where the past progressive is much less frequent than the present progressive.

A past progressive gram was found in one case, *-lako-* in Jola-Fonyi (Niger-Congo). In Chapter 5, certain Taleshi varieties exhibiting past progressive patterns with no dedicated pattern for the present progressive were presented. In those cases, it was shown that the asymmetry is a consequence of the pattern in the present having grammaticalized into becoming a general present tense, while the past progressive remained a progressive. Languages with past progressives grams

and no corresponding present progressive have also been noted previously in Bertinetto et al. (2000a:525). Also in the Taleshi of Asalem, there exists an additional past progressive pattern formed with *damand* 'stay in' that does not have a corresponding present form. It is, however, not certain that these asymmetries are the outcome of a similar process as for the patterns in Taleshi varieties. An asymmetry, reverse to the Taleshi one, is also noted in Bybee et al. (1994:142) for Kui (Dravidian), where *man-* is used with ongoing meaning in the present and for ongoing and habitual function in the past.

We then observe different types of asymmetries that have probably arisen in different ways: one where we observe present progressive grams with no corresponding past progressive, which are assumed to exist due to a general tendency to use progressives more often in the present than in the past, and one where either the present or the past progressive pattern has further grammaticalized towards the imperfective in one tense but not in the other. The typological investigation indicates that it is more common to have a progressive gram restricted to the present than a progressive restricted to the past, thus confirming previously observations in Dahl (1985:94).

7.3 Uses of progressives

This thesis has been concerned with the uses of progressives in various ways. Below the most favorable contexts and the peripheral uses are discussed in turn. The section ends with a description of the progressive gram type.

7.3.1 Most favorable contexts for progressives

The most favorable contexts for progressives were shown to be contexts that include a FOC ongoing event where the FOC reference time is either implicit or explicit. In the present, it was often noted that progressive grams are used in contexts carrying emotive effect, such as emphasis, intensification, irony or hyperbolic uses. Also, utterances with progressives were often used in contexts that attempt to turn the attention of the listener towards, or make the listener aware of, an ongoing event. Progressives were also shown to relate to new and topical information and highlight the involvement and busyness of the agentive subject. Somewhat similar uses were noted, although less often, in non-backgrounding past contexts.

Contexts like these appear to have something in common, but they are not coherent enough to be explainable as one feature. The suggested notion of "epistemic contingency" (see, e.g., De Wit & Brisard 2014) is in my opinion too broad, as these utterances are more than non-generic. It is also difficult to prove that readings such as emphasis or irony, etc., are solely triggered by the

progressive, often these uses are also induced by other elements in the context, tone, etc. Therefore, I will suggest that these features instead increase the likelihood for the use of a progressive gram. Apart from ongoingness and dynamicity, which are almost obligatory, such features can be said to be those given in 7:1.

7:1 Features increasing the likelihood of the use of a progressive gram

- A FOC reference time.
- The engagement or 'busyness' of the agentive subject on the event.
- An emotive component, e.g. the event referred to is dramatic, urgent or given with an ironic tone, expresses surprise, etc.
- The desire to turn the attention of the addressee towards, or make the addressee aware of, an ongoing event.

I will suggest that, the more features that are present, the more likely it is that a progressive gram will be used. Consider, for example, the utterance that had the highest incidence of progressives in the questionnaire presented in Dahl (1985:92) as well as in PROGQ presented in Chapter 3. With a slight modification and uttered in a dramatic way with an irritated tone, it would have all the features in 7:1.

7:2 - Don't disturb me! Can't you see, *I am writing a letter*?

Most uses in the past were part of narrations. In these contexts the progressive often applies to an event that is the background of another (often FOC and telic) event. Thus, in backgrounding uses, the FOC point is not given by the event to which the progressive applies but by the telic event. In Chapter 1, it was noted that telic events are needed to create change, which pushes narration forward (Dahl 2013:70–72). A detelicised event, such as an utterance including a progressive, does not involve change in this sense. Progressive utterances in narration thus have a sense of pausing the narration or bringing it to a plateau, as it were. Therefore, in the past, unlike in the present, the event to which the progressive applies has a sense of lacking an ending, as shown in 7:3.

7:3 a) He is reading.

b) He was reading... (when suddenly the door opened).

The difference between uses of progressives with present time reference and past time reference has the slightly odd consequence that, while progressives are typically used for providing new, urgent and emphatic information in the present, they are typically used for providing backgrounded information in the (narrative) past where the sense of immediacy is much less prominent.

The results of the thesis suggest that when it comes to progressives, their relationship to FOC and DUR contexts is not equal. While FOC is the default reading when progressives are used, the duration of the reference time in DUR contexts has to be explicitly given. The default FOC reading of progressives in present time reference could be due to features of the present time reference rather than the progressive: things that are said with the speech moment as reference time have a natural interpretation of having a FOC reference point. Only generic utterances are different in this regard. This means that focality could be a feature that the progressive inherits from the speech moment in the present. In narration, FOC seems to be linked to telic events, meaning that it is not given by the event to which the progressive applies, rather, the progressive gram seems to require it. In this sense, focality as applied to progressives behaves somewhat different in the present than in the past. On the other hand, progressives disambiguate utterances with regard to focality in both the present and the past, meaning that more general and generic readings are not available. The issue of focality need further examination, especially with regard to earlier stages in the development of progressive that have been shown to involve durative contexts (Bertinetto 2000:576–581). The conclusion drawn here is that focality is something required by progressives and with which progressives are strongly associated.

7.3.2 Peripheral uses of progressive grams

This thesis has been concerned with the peripheral uses of progressive grams in different ways, where these are understood as all uses that cannot be described as ‘ongoing at the reference time’. Peripheral uses are, e.g., habitual, proximative, futurate, iterative and performative uses, uses in DUR contexts, uses with stative predicates, as well as temporary, intensified and subjective readings. These uses are of different sorts: they may be uses typically associated with further grammaticalization such as habitual uses, or uses of the progressive when combining with different types of events such as proximative or iterative uses, or uses that are interpretations such as the notions of temporariness and subjective readings.

The ability to occur with stative predicates is viewed as a sign of further expansion of a progressive gram. Chapter 2 showed that the occurrence of progressives with stative verbs almost exclusively involved posture verbs. The Persian *dāštan* progressive was shown to have all the above mentioned peripheral uses except for uses with stative predicates. As discussed in Chapter 4, the English

progressive may combine with a greater number of stative predicates, including posture verbs as well as stative predicates such as *be silly* or *be angry*. In the latter case, the use of the progressive renders a subjective and atypical reading rather than marking ongoingness. It has been shown that these predicates, when used with the progressive, refer to the behavior of the subject and are thus made more dynamic. In Turkish, *-Iyor-* can be used with statives and seems to have pushed the simple present towards generic or more specialized uses. In a language like Taleshi of Anbaran Ardebil, where a former present progressive has taken over the functions of the general present, certain stative verbs such as the verb 'to know' are marked with the new present pattern, a DAR gram, whereas the verb of existence resist this change and is unmarked. The data suggests that among the stative verbs, posture verbs are the first type of statives to combine with progressives, and that the combination with non-posture stative verbs is one of the last types of peripheral uses that progressive grams take over before fully grammaticalizing into imperfectives. For some statives, the take-over of a new present may never happen (see Haspelmath 1998:46).

The subjective readings with the *dāštan* progressive were noted to rise when the pattern occurs in contexts that are less typical for the pattern, such as in contexts with a DUR reference time. Similarly, the subjective readings in English are often discussed with stative predicates, which are also a rather infrequent combination for the pattern. In these two languages, subjective readings as arising with the use of progressives are linked to uses in contexts that are new for the patterns. It would be interesting to know if this also holds cross-linguistically.

One question that arises is how these peripheral uses relate to one another cross-linguistically as well as historically. We know that not all progressive grams can occur with (certain) stative verbs or events viewed as achievements. Progressive grams may also have, apart from the peripheral uses, uses that are remains of their original uses, such as locative ones. Taking these also into account, we may ask: are there progressive grams that only mark ongoingness at reference time and nothing else? It may well be that such grams are not very common. Both with regard to function and structure, the progressive seems unstable. Functionally, it is as if progressives are either showing signs of their origin or signs of further grammaticalization. Structurally, the data from the Caspian languages showed that at least in this part of the world, they have a strong tendency to be borrowed.

7.3.3 Refined description of the progressive gram type

Given the availability of progressives in various less typical uses, I will suggest a refinement of the description of the progressive gram type as given in 7:4.

7:4 Description of the progressive gram type

> Progressives are morphosyntactic devices primarily used for events that are ongoing at the reference time. They are not the primary choice in habitual contexts or with stative predicates and they are not available in prototypical generic contexts such as *Cows eat grass*.

From a distributional perspective, I will suggest that the members of the progressive gram type have FOC ongoing uses as their majority use. This will enable an inclusion of less grammaticalized patterns found in languages where also other patterns are used for events that are FOC ongoing, as well as more mature patterns where the progressive is the primary choice for such events.

7.4 Grammaticalization of PROG to IPFV

The findings of this thesis can be linked to the maturation process of progressive grams towards the imperfective. I suggest that they are linked in the way given in Table 29. These findings are in line with previous research and complement them. The different aspects of grammaticalization referred to are related to one another and can be said to view the grammaticalization process from different perspectives.

As seen in the table, the requirement of features given in 7:1 is weakened the more mature a progressive gram gets, a process linked to the other processes. To give an example, it was noted that the Swedish progressive with *hålla + på + att/och* has a stronger sense of engagement and busyness than the Persian *dāštan* progressive. We can then say that the Swedish progressive has a more prominent meaning, i.e. semantic content that the gram adds to the utterance, than the Persian pattern, which results in the use of the Swedish pattern having a stronger sense of pragmatic markedness than the use of the Persian pattern. The Swedish gram was also shown to have a much lower frequency in the TED corpus than the Persian pattern.

The weakening of the concrete meaning that can be assigned to a progressive is accompanied by a decrease in borrowability. New (pre-progressive) patterns need, however, to reach a certain frequency before they can spread. This means that borrowability is first low, later high and finally again low. The final stage of low borrowability applies to former progressives that have become, or are on their way to becoming, imperfectives.

One important step in the grammaticalization of progressive grams towards the imperfective is when the progressive becomes the typical choice for marking ongoingness. If continued, this process will eventually push imperfectives

towards generic, or other, more specialized, uses. This has, for example, already happened in English but not in languages such as Persian and Swedish.

Table 29. The grammaticalization of PROG to IPFV

	PROG ⟶		IPFV
FREQUENCY OF GRAM	low		high
MARKING ONGOINGNESS	PROG or IPFV for ongoing use	PROG for ongoing use	former PROG/new IPFV for ongoing use
FEATURE REQUIREMENTS (as in 7:1)	more features required	fewer features required	no features required
MEANING OF PATTERN	more prominent	less prominent	meaning of progressive gram lost
USE WITH STATIVES	restricted, posture verbs if any	less restricted	(almost) not restricted
ENCODING PATTERN	typically periphrastic		typically inflectional
BORROWABILITY	more borrowable		less borrowable

The expansion of uses has already been discussed with respect to stative predicates. Also, the preference for FOC contexts is assumed to decrease as progressives start occurring in habitual and generic contexts.

The peripheral uses are not included in the table since they involve different types of uses and readings. I will leave future research to investigate whether there is a diachronically cross-linguistic uniform order in which these uses occur, as patterns shift from progressive to imperfective.

Engagement and 'busyness' are not only a requirement for the use of a progressive in earlier stages of maturation, they were also found as original meanings of progressives among certain grams in the red cluster group in Chapter 2 where grams with meanings such as 'middle', 'in the middle', 'be engaged in/busy in' were found. Also, the *mašġul* 'busy' + INF pattern in Persian, which partly overlaps with progressives, has a similar meaning. In fact, removing the requirement of location from the original meaning of progressives suggested by Bybee et al. (1994:136), we get a meaning that can be assumed to give rise to these grams: "An agent is in the midst of an activity at reference time".

The data on subjective uses is limited, but as was discussed, it indicates that such uses are available when progressives are starting to be used in new contexts. The agency requirement is a feature typical for less grammaticalized progressives, but it may also be linked to new contexts as the expansion of the English

progressive to non-posture statives have been shown to require an agentive subject. It would be interesting to know if subjectiveness and agentivity requirement are typical features of new contexts for progressives cross-linguistically as they expand their uses. I will leave these issues to be further investigated in the future.

7.5 Final remarks

Having come to the end of this journey, one of the conclusions that I draw is that much still remains to be done for a full understanding of the uses and development of progressive grams. Some unsolved questions have been mentioned in this chapter. Adding to these, the question of why peripheral uses, readings and nuances arise is still puzzling. We still cannot fully explain why, or consistently foresee when, speakers make the choice of using a progressive. The more detailed steps in the further grammaticalization of the progressive also need to be elucidated. The complexity of these issues is partly linked to the evolving nature of progressives, which may result in the members of this gram type being quite diverse. This thesis has provided some answers to these questions in presenting features that increase the likelihood of using a progressive, discussing and suggesting explanations for some of the less typical readings of progressives and examining patterns on their way of becoming imperfectives. The empirical data presented in this thesis has contributed to answering some questions on the path to an understanding of the progressive in its full complexity.

Appendices

A. Grams in the Bible corpus

Below, the grams in the Bible corpus are given, grouped after classification. The name of the language is followed by the ISO 639-3, the search string of the gram, the total number of occurrences of that gram in the corpus and the recall and precision values. The first parenthesis gives the recall and precision for the whole gram set, the second for the cluster group in which the gram is a member. The colour term refers to the cluster group with the same name. In the search string, the hyphen indicates an optional inflectional boundary (e.g. *-sha-* searches for *sha*, *-sha-*, *-sha* and *sha-*) and *=V* means that the search string is an element annotated as a verb in the corpus. For more complex patterns, a structural schema is given.

Afro-Asiatic

Biu-Mandara

Merey [meq] *faya* 3648 (0.9, 0.649) (0.8, 0.584) yellow

Semitic

Tigrinya [tir] *'ālo-, 'ālā-, zolo-, zolā-, naye- , nére-, naba-* 1896 (0.2, 0.683) (0.75, 0.618) yellow

Austro-Asiatic

Khmer

Central Khmer [khm] *kampoung-* 174 (0.9, 0.958) (0.65, 0.8) red

Viet-Muong

Vietnamese [vie] *đương* 200 (0.8, 0.946) (0.8, 0.728) red

Vietnamese [vie] *đang* 485 (0.85, 0.927) (0.85, 0.596) red

Austronesian

Barito

Ma'anyan [mhy] *rahat* 178 (0.6, 0.994) (0.85, 0.981) red

Ngaju [nij] *metoh(-)* 400 (0.55, 0.789) (0.8, 0.545) red

Ot Danum [otd] *rahat* 180 (0.35, 0.994) (0.6, 0.925) red

Celebic

Balantak [blz] *pintanga'* 149 (0.75, 0.915) (0.9, 0.817) red

Muna [mnb] *tangasano* 124 (0.9, 0.975) (0.85, 0.942) red

Central Malayo-Polynesian

Bima [bhp] *wunga* 302 (0.85, 0.912) (0.9, 0.832) red

Greater Central Philippine

Gorontalo [gor] *donggo* 435 (0.95, 0.908) (0.95, 0.615) red

Javanese

Javanese [jav] *lagi* 223 (0.5, 0.856) (0.8, 0.66) red

Lampungic

Lampung Api [ljp] *sedang* 171 (0.4, 0.967) (0.8, 0.894) red

Malayo-Sumbawan

Achinese [ace] *teungoh* 357 (0.95, 0.965) (1.0, 0.868) red

Central Malay [pse] *dang* 190 (0.55, 0.953) (1.0, 0.89) red

Indonesian [ind] *sedang* 157 (0.5, 0.972) (0.75, 0.834) red

Jarai [jra] *hlak* 566 (0.9, 0.903) (0.85, 0.844) green

Madurese [mad] *teppana* 152 (0.65, 0.963) (0.8, 0.912) red

Minangkabau [min] *sadang* 341 (1.0, 0.964) (1.0, 0.899) red

Sasak [sas] *kenyeke* 232 (0.55, 1.0) (0.85, 0.952) red

Standard Malay [zsm] *sedang* 155 (0.45, 1.0) (0.75, 0.915) red

Northern Luzon

Eastern Bontok [ebk] *cha* 1111 (0.85, 0.77) (0.95, 0.688) green

Oceanic

Hawaiian [haw] *ke* + V + *nei* 453 (0.05, 0.918) (0.45, 0.896) yellow

Kara (Papua New Guinea) [leu] *taxa* 2009 (0.85, 0.884) (0.9, 0.819) yellow

Rejang

Rejang [rej] *gidong* 135 (0.5, 0.984) (0.8, 0.896) red

South Sulawesi

Bambam [ptu] *mahassa-* 181 (0.8, 0.911) (0.8, 0.841) red

Creoles and Pidgins

Belize Kriol English [bzj] *di* + V 1516 (0.8, 0.84) (0.7, 0.774) yellow

Morisyen [mfe] *pe* 1754 (0.95, 0.921) (0.9, 0.88) yellow

Seselwa Creole French [crs] *pe* 1410 (1.0, 0.936) (1.0, 0.876) green

Hmong-Mien

Hmong Daw [mww] *tabtom* 221 (0.95, 0.986) (0.85, 0.802) red

Indo-European

Albanian

Tosk Albanian [als] *po* 1104 (0.8, 0.803) (0.65, 0.764) yellow

Germanic

English [eng] *be* + GER 1252 (0.95, 0.917) (1.0, 0.883) yellow

Indic

Bengali [ben] V-*(c)ch(il)-* 1323 (0.65, 0.93) (0.75, 0.902) yellow

Hindi [hin] *rah-* + COP 210 (0.9, 0.733) (0.8, 0.738) green

Panjabi [pan] *rah-* + COP 1327 (0.85, 0.886) (0.75, 0.8) green

Romance

Portuguese [por] *estar* + GER 538 (0.8, 0.969) (0.85, 0.926) green

Spanish [spa] *estar* + GER 339 (0.5, 0.969) (0.6, 0.913) green

Mayan

Achi [acr] *tijin* 652 (0.85, 0.892) (0.9, 0.841) green

Aguacateco [agu] *-tzan-* 994 (0.65, 0.832) (0.9, 0.702) green

Chol [ctu] *woli-* 1306 (0.9, 0.8) (0.95, 0.671) green

Chuj [cac] *van* 581 (0.75, 0.929) (0.65, 0.892) green

K'iche' [quc] *tajin* 801 (0.95, 0.939) (1.0, 0.896) green
Kekchí [kek] *yo̲-* + *chi-* 1315 (0.95, 0.887) (1.0, 0.799) green
Popti' [jac] *lañan* 264 (0.7, 0.892) (0.65, 0.866) green
Q'anjob'al [kjb] *lanan-* 530 (0.8, 0.94) (0.9, 0.926) green
Tz'utujil [tzj] *-emjon* 520 (1.0, 0.916) (1.0, 0.857) green
Tzeltal [tzh] *yac-* 2628 (0.95, 0.677) (0.85, 0.605) yellow
Uspanteco [usp] *-tijin-* 954 (0.95, 0.923) (1.0, 0.853) green
Western Kanjobal [knj] *lalan* 551 (0.9, 0.971) (0.85, 0.961) green
Yucateco [yua] *táan* 1898 (0.75, 0.777) (0.75, 0.708) yellow

Niger-Congo

Bantoid

Koongo [kng] *-eti* 1493 (0.15, 0.793) (0.9, 0.742) yellow
Lenje [leh] *-too-* 1097 (0.1, 0.855) (0.8, 0.83) yellow
Makaa [mcp] *ŋgɔ́* 2022 (0.7, 0.806) (0.25, 0.732) yellow
Mbunda [mck] COP + *naku-* 1379 (1.0, 0.887) (0.9, 0.84) yellow
Nyoro [nyo] *n(i)-...-a* 2985 (0.9, 0.725) (0.8, 0.665) yellow
Venda [ven] *khou* 786 (0.5, 0.924) (0.25, 0.83) green
Zemba [dhm] *ma-* 2535 (0.0, 0.624) (0.65, 0.552) yellow

Cross River

Gokana [gkn] *gé* 3409 (0.75, 0.679) (0.85, 0.609) yellow

Dogon

Toro So Dogon [dts] V-*wɔ-* 3662 (0.5, 0.714) (0.65, 0.646) yellow

Eastern Mande

Busa [bqp] *-tɛn* 1846 (0.95, 0.88) (0.85, 0.831) yellow

Kru

Kuwaa [blh] V-*nù* 2422 (0.8, 0.824) (0.8, 0.765) yellow

Kwa

Ewe [ewe] V-*m* 2022 (0.95, 0.803) (0.85, 0.746) yellow
Gen [gej] *le =V* 2676 (0.8, 0.851) (0.9, 0.802) yellow
Sekpele [lip] *-lɛ* + V 972 (0.65, 0.814) (0.6, 0.76) yellow

Mel

Southern Kisi [kss] *chō* 2904 (0.05, 0.7) (0.75, 0.632) yellow

Northern Atlantic

Jola-Fonyi [dyo] *-lako-* 1084 (0.8, 0.711) (0.9, 0.596) green
Wolof [wol] *ngi* 885 (0.05, 0.9) (0.55, 0.881) yellow

Western Mande

Susu [sus] V-*fe* 1254 (0.7, 0.728) (0.75, 0.667) yellow

Nilo-Saharan

Nilotic

Kumam [kdi] *-tye* + V 777 (0.5, 0.956) (0.6, 0.892) green

Oto-Manguean

Zapotecan

Amatlán Zapotec [zpo] *ka-* 1295 (1.0, 0.688) (1.0, 0.582) green

Chichicapan Zapotec [zpv] *ca*-V 1265 (0.7, 0.807) (0.9, 0.677) green

Coatecas Altas Zapotec [zca] *ki*-V 1665 (0.85, 0.74) (0.7, 0.658) yellow

Ozolotepec Zapotec [zao] *nge-* 824 (1.0, 0.776) (1.0, 0.717) green

Santo Domingo Albarradas Zapotec [zas] *ca-* 1889 (0.95, 0.778) (0.95, 0.67) green

Quechuan

Ayacucho Quechua [quy] *-chka-* 2328 (0.85, 0.8)(0.8, 0.741) yellow

Cajamarca Quechua [qvc] *-yka-* 1500 (0.6, 0.808) (0.75, 0.675) green

Cusco Quechua [quz] *-sha-* 2398 (0.65, 0.83) (0.8, 0.76) yellow

Eastern Apurímac Quechua [qve] *-sha-* 3016 (0.6, 0.803) (0.75, 0.718) yellow

South Bolivian Quechua [quh] *-sha-* 2176 (0.85, 0.866) (0.95, 0.795) yellow

Sino-Tibetan

Chinese

Min Nan Chinese [nan] *teh* 794 (1.0, 0.926) (0.95, 0.82) green

Kuki-Chin

Bawm Chin [bgr] *liau(ah)* 435 (0.7, 0.701) (0.8, 0.614) green

Trans-New Guinea

Angan

Angaataha [agm] *taati, taamɨ, taawɨ* 659 (0.65, 0.749) (0.85, 0.661) green

Turkic

Kirghiz [kir] *žata-* 516 (0.1, 0.909) (0.65, 0.872) yellow

Turkish [tur] *-yor-* 2100 (0.15, 0.764) (0.9, 0.716) yellow

Uzbek [uzb] *-jap-, -yotgan edi-* 520 (0.1, 0.93)(0.65, 0.917) yellow

Uralic

Saami

Northern Sami [sme] COP + V.AKTIO.ESSIVE 445 (0.9, 0.847) (0.95, 0.733) green

Uto-Aztecan

Aztecan

Zacatlán-Ahuacatlán-Tepetzintla Nahuatl [nhi] V-*to-* 1259 (0.8, 0.823) (0.85, 0.663) green

Western Fly

Bine [bon] V-*eni* 1980 (0.15, 0.753) (0.8, 0.701) yellow

B. PROGQ

پرسشنامه

این پرسشنامه دربرگیرنده شکل های مختلف دستوری زبان ها و گویش های ایرانی است. در صورت تمایل شما می توانید ناشناس بمانید.

نام:

نام فامیل:

سن:

محل اقامت:

محل تولد:

به چه زبانی یا گویشی در کودکی صحبت می کردید؟

لطفا جمله های زیر را ترجمه کنید. ترجیحا به جای زبان نوشتاری، پرسشنامه را به زبان گفتاری خود پر کنید. کلمه هایی که بین [] قرار دارند بخشی از زمینهٔ جمله هستند و نیازی به ترجمه ندارند.

شکل فعل به فرم مصدر داده شده است (یعنی مثلا "خوردن"، "دیدن"، "دوش گرفتن") تا شما فرم مناسب را خودتان بنویسید. لطفا، با جملات کامل پاسخ بدهید.

سئوال نمونه:

[سر سفرهٔ غذا، نیما سراغ نوشین را می گیرد:

- نوشین کجاست؟]
- او در حمام **{دوش گرفتن}**.

پاسخ نمونه:

- توی حموم داره دوش می گیره.

۱. [پدر به بچه اش می گوید:

- لطفا مزاحم من نشو،] من **{نامه نوشتن}**.

۲. [کسی زنگ می زند و سراغ مریم را می گیرد. من جواب می دهم:

- مریم پیش من است] او روزنامه **{خواندن}**.

۳.

[- نیما چه کارمی کند؟

- من نمی دانم ولی فکر می کنم که:] او الآن باید **{غذا خوردن}**.

۴. سال گذ شته آرین ما را سه بار **{ملاقات کردن}**.

۵.

[- مریم هر شنبه چکار می کند؟]

- او خانه را **{تمیز کردن}**.

۶. [دیروز همزمان که من خوابیده بودم]

یاسمن دو ساعت برای خودش **{ بازی کردن}**.

۷. [چه کار قشنگی!] تو چه **{مهربان بودن}**.

۸.

[- من حالا بلوز آبی ام را احتیاج دارم. اون کجاست؟]

- اون روی جا لباسی **{آویزان بودن}**.

۹. اگر تو ساعت ۸ بیایی من هنوز **{غذا درست کردن}**. [لطفاً یک کمی دیر تر بیا.]

۱۰. مریم تا یک دقیقهٔ دیگر **{رفتن}**.

۱۱. [کسی زنگ می زند و سراغ مریم را می گیرد. من جواب می دهم:
- مریم پیش من است] او الآن **{کار کردن}**.

۱۲. [بعد از این همه وقت]، تو تازه حالا درس خواندن را **{شروع کردن}**؟

۱۳. [آرین از موهای مریم ایراد گرفت. مریم با تعجب به او می گوید:]
- تو چقدر امشب **{بی ادب بودن}**.

۱۴. [به آرین که روی مبل خوابیده نگاه کن!]
او **{خواب دیدن}**.

۱۵. [کسی زنگ می زند و سراغ نوشین را می گیرد. من جواب می دهم:
- نوشین پیش من است] او تز خود را **{نوشتن}**. [ولی من فکر نمی کنم که هرگز تمام شود.]

۱۶. [گفتگوی تلفنی:
- نیما الآن با تو است؟]
- نه، او [در اطاق بعدی] **{ورق بازی کردن}**.

۱۷. [هنوز امید هست] یواش یواش شرایط **{بهتر شدن}**.

۱۸. [مادر برای تنبیه دختر به او می گوید:] تو **{نرفتن}** مهمونی!

۱۹. [در تمام مدّت طول کلاس] مریم با دوست کناری اش **{حرف زدن}**. [او حتي بعد از کلاس هم به حرف زدن ادامه داد].

۲۰. اون وقت ها او هر جمعه پیاده روی **{رفتن}**.

۲۱. من درست زمانی عکس را گرفتم که آرین به طرف پنجره **{سنگ پرتاب کردن}**.

۲۲. [کسی زنگ می زند و سراغ مریم را می گیرد. من جواب می دهم:
- مریم پیش من است] او به خواهرش **{هدیه دادن}**.

۲۳. [عجله کن!] قطار **{رفتن}**.

۲۴. درست زمانی که کاپیتان **{فرمان شلیک دادن}**، پیام صلح رسید.

۲۵. [یک روز روشن تابستانی بود.] زنبورها **{وز وز کردن}**، پرنده ها **{ آواز خواندن}** و گاوها در چراگاه **{چریدن}**. [ناگهان زمین دهان باز کرد و شیطان بیرون آمد].

۲۶. [لحظه به لحظه] پلیس چیزی را که او می گفت **{یادداشت کردن}**.

۲۷. [کسی زنگ می زند و سراغ نوشین را می گیرد. من جواب می دهم:
- نوشین پیش من است] او سیب زمینی ها را **{پوست کندن}**.

۲۸. [تورو خدا] وقتی رئیس بر گشت **{کار کردن}**!

۲۹. مریم فردا **{رفتن}**.

۳۰. [بیا بریم بیرون] حالا دیگه **{باران نباریدن}**.

۳۱. او مرتب اسم مردم را **{فراموش کردن}**.

۳۲. [نگاه کن. او دوباره این کار را می کند!]

او دوباره نا خود آگاه به همسایهٔ خود**{ توهین کردن}**. [او واقعاً نمی فهمد که این کار توهین است].

۳۳. [کسی زنگ می زند و سراغ آرین را می گیرد. من جواب می دهم:

- همین الآن] او در آشپزخانه **{نشستن}**.

۳۴.[گفتگوی تلفنی:

ـ نیما الآن خانه است؟]

ـ نه، او **{ورق بازی کردن}**. [مثل همیشه].

۳۵. [همین الآن] آرین وسط درگاه **{ایستادن}**.

۳۶. در سالی که گذشت ما معمولا خانه را شنبه ها **{تمیز کردن}**. [ولی حالا جمعه ها تمیز می کنیم.]

۳۷. آب **{جوشیدن}**. [می خواهی چای دم کنم؟].

۳۸. همین الآن کوه نورد به قلهٔ کوه **{رسیدن}**.

۳۹. [نگاه کن، هوا روشنتر شده] تابستان دیگه **{شروع شدن}**.

۴۰.[کسی زنگ می زند و سراغ آرین را می گیرد. من جواب می دهم:]

او همین الآن به مسکو **{پرواز کردن}**. [تو میتوانی فردا به هتل او زنگ بزنی.]

۴۱. [من الآن نمی توانم بیایم، چونکه] نوشتن نامه ام را **{تمام کردن}**.

۴۲. [- یاسمن چه کارمی کند؟

- من نمی دانم ولی فکر می کنم:] او الآن باید **{تدریس کردن}**.

۴۳. آرین فردا **{عروسی کردن}**. کی ساقدوشش **{بودن}**.

۴۴. [کسی زنگ می زند و سراغ مریم را می گیرد. من جواب می دهم:

- مریم پیش من است] او برای گوسفند ها **{آخور ساختن}**.

۴۵. [مادر به بچه اش می گوید:] من دیگه **{عصبانی شدن}**.

۴۶. [از پنجره به بیرون نگاه کن!]

خورشید **{تابیدن}**.

۴۷. زمانی که منشی مریض است، رئیس نامه ها را خودش **{تایپ کردن}**.

۴۸. [به بیرون نگاه کن] کم کم برف زمین را **{پوشیدن}**.

۴۹. [کسی زنگ می زند و سراغ نوشین را می گیرد. من جواب می دهم:

- نوشین پیش من است] او سیب زمینی **{پوست کندن}**.

۵۰. رئیس عصبانی بود چون وقتی آمد تو نیما **{کار نکردن}**.

۵۱. دیروز زمانی که آرین آمد خانه، خیلی خسته بود چون او تمام هفته سخت **{کار کردن}**.

۵۲. [این سیب ها را دیگه باید بخوریم] آنها **{پوسیدن}**.

۵۳. [نه، هنوز شرایط بهتر نشده] او همینطور این بازی ها را **{ادامه دادن}**.

۵۴. [بفرمایید تو! همین الآن] غذا **{داده شدن}**.

۵۵. [کسی زنگ می زند و سراغ مریم را می گیرد. من جواب می دهم:
- مریم پیش من است] او **{آواز خواندن}**.

۵۶. [من خیلی خسته ام] از وقتی که پا شدم همینطور **{غذا درست کردن}**.

۵۷. پسرجوان **{مردن}** [ولی بالاخره داروی مناسب را پیدا کردند و او را درمان کردند].

۵۸. آنقدر در آفتاب دراز کشیدم که **{سیاه شدن}** [ولی خوشبختانه نشدم].

۵۹.[الآن یکدفعه] آرین این موزیک را **{دوست داشتن}**.

۶۰. [فکر کن! هم زمان که ما با هم حرف می زنیم] زمین دور خورشید **{چرخیدن}**.

۶۱. دیروز در حالی که مریم در اطاق **{درس خواندن}**، آرین در حیاط **{بازی کردن}**.

۶۲. [الآن نا گهان] آرین جواب سوال را **{دانستن}**.

۶۳. کوه ها دور و بر هوا پیما را **{گرفتن}**.

۶۴. [کسی زنگ می زند و سراغ علی را می گیرد. من جواب می دهم:
- علی پیش من است] او **{آماده شدن}**.

۶۵. [دیشب ساعت ۸] وقتی که آرین رسید، مریم هنوز **{کار کردن}**.

۶۶. [من خیلی خسته ام] از وقتی که بیدار شدم **{نان پختن}**.

۶۷. کوه نورد به قلهٔ کوه **{رسیدن}** [که ناگهان سکته کرد و مرد].

۶۸. [در طول تابستان] مجسمه در حیاط **{بودن}**.

۶۹. [کسی زنگ می زند و سراغ آرین را می گیرد. من جواب می دهم:]
او همین الآن **{بیرون رفتن}**. [می خواهی صداش کنم؟]

C. English translation of PROGQ

Questionnaire

The questionnaire is concerned with the different grammatical forms of verbs in various Iranian languages and dialects. If you wish, you may remain anonymous.

Given name:
Last name:
Age:
City of birth:
Current place of residence:
In what language or dialect did you speak during your childhood?

Please translate the sentences given below. Preferably, fill out the questionnaire in your own spoken language rather than written language. The words between [] are part of the context and need not to be translated.
The given verb forms are in the infinitive (as in, for example, 'EAT', 'SEE', 'TAKE SHOWER') so that you yourself can provide a suitable verb form. Please answer in complete sentences.

Example of question:

[At the dinner table, Nima asks about Ali:
- Where is Ali?]
- He TAKE SHOWER.

Example of answer:

- He is taking a shower.

1. [Father says to his child:
- Don’t disturb me,] I WRITE LETTER.

2. [Somebody’s calling and asking for Maryam. I answer:
- Maryam is near me,] she READ the newspaper.

3. [- What is Nima doing?
- I don’t know but I think that:] he must EAT right now.

4. Last year, Aryan VISIT us three times.

5. [- What does Maryam do every Saturday morning?]
- She CLEAN the house.

6. [Yesterday, during my sleep] Yasaman PLAY for 2 hours all by herself.

7. [What a nice gesture!] You BE KIND.

8. [- I need my blue shirt right now. Where is it?]
- It HANG on the nail.

9. If you come at 8 o’clock, I still COOK. [Come a little later, please.]

10. Maryam LEAVE in a minute.

11. [Somebody’s calling and asking for Maryam. I answer:
- Maryam is near me,] she WORK right now.

12. [After all this time], you only BEGIN to study right now?

13. [John has made a negative comment on Maryam’s hair-style. Maryam says with a tone of surprise:]
- You BE RUDE this evening.

14. [Look at Aryan on the sofa!] He DREAM.

15. [Somebody’s calling and asking for Noshin. I answer:
- Noshin is near me,] she WRITE her thesis. [But I don’t think it will ever finish.]

16. [On the phone:
- Is Nima with you right now?]
- No, he PLAY CARDS [in the next room].

17. [There is still hope] the situation BECOME BETTER slowly.

18. [Mother to daughter, whom she wants to punish] You NOT GO to that party!

19. [During the whole time of the class] Maryam TALK to her neighbor [in fact, she carried on even afterwards].

20. At that time, he GO walking every Saturday.

21. I took the photo exactly when John THROW the stone towards the window.

22. [Somebody's calling and asking for Maryam. I answer:
- Maryam is near me,] she GIVE a present to her sister.

23. [Hurry up!] The train LEAVE.

24. The pardon arrived just while the captain GIVE ORDER TO FIRE.

25. [It was a bright summer day.] The bees HUM, the birds SING and the cows GRAZE in the pasture. [Suddenly, the earth opened and the devil came out.]

26. [Moment by moment] the policeman TAKE NOTES of what (s)he said.

27. [Somebody's calling and asking for Noshin. I answer:
- Noshin is near me,] she PEEL the potatoes.

28. [For goodness sake] WORK when the boss comes back!

29. Maryam LEAVE tomorrow.

30. [Let's go out] it NOT RAIN now.

31. He FORGET people's names all the time.

32. [Look, he does it again!] Again, he inadvertently INSULT his neighbor. [He really doesn't understand that this is an insult].

33. [Somebody's calling and asking for Aryan. I answer that right now:]
He SIT in the kitchen.

34. [On the phone:
- Is Nima home right now?
- No, he PLAY CARDS [as usual].

35. [Right now] Aryan STAND in the doorway.

36. Last year we usually CLEAN the house on Saturdays. [But now we do it on Fridays.]

37. The water BOIL. [Shall I make tea?]

38. Right now the climber REACH the top of the mountain.

39. [Look, the air is brighter]. Summer BEGIN.

40. [Somebody's calling and asking for Aryan. I answer:]
- He FLY to Moscow right now. [You can call him tomorrow at his hotel].

41. [I can't come right now, because] FINISH writing my letter.

42. [- What is Yasaman doing?
- I don't know but I think that:] she must TEACH right now.

43. Aryan GET MARRIED tomorrow. Who BE his witness?

44. [Somebody's calling and asking for Maryam. I answer:
- Maryam is near me,] she BUILD a shelter for the sheep.

45. [Mother says to her child:] I GET ANGRY.

46. [Look out the window!] The sun SHINE.

47. When the secretary is ill, the boss TYPE his own letters.

48. [Look out (the window)] the snow gradually COVER the land.

49. [Somebody's calling and asking for Noshin. I answer:
- Noshin is near me,] she PEEL potatoes.

50. The boss was angry, because when he came in Nima NOT WORK.

51. When Aryan came home yesterday, he was very tired because he WORK hard all week.

52. [We have to eat these apples] they ROT.

53. [No, the situation has not improved]. (S)he unceasingly CONTINUE these games.

54. [Come in please! Right now] the food GIVE.

55. [Somebody's calling and asking for Maryam. I answer:
- Maryam is near me,] she SING a song.

56. [I am so tired] I unceasingly COOK all day since I got up this morning.

57. The young man DIE [but finally they found the right medicine and cured him].

58. I lay in the sun for so long that I BECOME BLACK(/GOT BLACK) [but fortunately I didn't].

59. [Now unexpectedly] Aryan LIKE the music.

60. [Think! As we are talking] the earth TURN around the sun.

61. Yesterday, while Maryam STUDY in her room, Aryan PLAY in the courtyard.

62. [Now unexpectedly] Aryan KNOW the answer.

63. The mountains SURROUND the plain.

64. [Somebody's calling and asking for Ali. I answer:
- Ali is near me,] he BECOME READY(/GET READY).

65. [Last night at 8 o'clock,] when Aryan arrived, Maryam still WORK.

66. [I am so tired] I BAKE BREAD all day since I got up this morning.

67. The climber REACH the top of the mountain [when suddenly he had a heart attack and died.]

68. [For the summer] the statue BE in the garden.

69. [Somebody's calling and asking for Aryan. I answer:]
He GO OUT right now. [Do you want me to call on him?]

D. Occurrences of *dāštan* construction in five Iranian movies

TIME	*dāštan* CONSTRUCTION	TAG OF USE	TENSE
'WOMEN'S PRISON' BY MANIJEH HEKMAT (2002)			
09:55	*tamām-e asāsāmun-o dāran mibaran daftar.*	FOC ONGOING	PRS
12:34	*dāri milarzi.*	FOC ONGOING	PRS
21:53	*hamejā-ro dāran mizanan.*	FOC ONGOING	PRS
26:17	*golandām dāre bačč-aš miād!*	PROX FOC	PRS
27:00	*golandām dāre mizād! zud bāšin!*	PROX FOC	PRS
49:33	*dāvod ham peyġām dāde, dāre pulā-š-o ja'm mikone jarime-m-o bede āzād šam.*	FOC ONGOING	PRS
01:31:22	*vaġt-i az birun miumadam didam dāštan mibordan-eš enferādi-ye pāin.*	FOC ONGOING?/DUR ONGOING?	PST
01:35:52	*bebin, man dāram miām birun.*	PROX FOC	PRS
'THE MUSIC MAN' BY DARIOUSH MEHRJUI (2007)			
09:10	*ġam dāre az man-o to del mibare.*	?	PRS
32:30	*jenseš dāre tamum miše az xomāri mitarse.*	PROX FOC	PRS
1:08:21	*fekr kardam dāri mibini mā va'z-emon zir-e xat-e faġr-e, mixāi ye ček-i čiz-i bekeši barā-ye mā zalilmordeha.*	FOC ONGOING	PRS
01:14:36	*inā-ro kojā dāri mibari?*	FOC ONGOING	PRS
01:22:36	*injā-ro dāran se tabaġe čār tabaġe dorost mikonan dige.*	FOC ONGOING	PRS
01:26:30	*be xodā agar in me'de-ye man lāmazhab dāre ? mikeše.*	?	PRS
01:26:41	*in me'de-ye man dāre misuze.*	FOC ONGOING	PRS
01:30:27	*dāram miram kānādā.*	FUTR ONGOING/PROX FOC	PRS
01:30:49	*man-am dāram miram nemiresam beram donbāl-eš.*	FUTR ONGOING/PROX FOC	PRS
01:40:52	*man tāze dāram jun migiram.*	FOC ONGOING? PROX FOC?	PRS
01:44:00	*bā to-am ke dāri be gerye-am mixandi.*	FOC ONGOING	PRS

01:36	*čizhā-i ke dārin migin dalāyel-e kāfi-i nist barā-ye talāġ, xānum.*	FOC ONGOING	PRS
03:24	*inhame bačče tu in mamlekat dāre zendegi mikone, ya'ni hič kudum āyande nadāran, xānum?*	FOC ONGOING	PRS
07:18	*dāram dars mixunam, xob.*	FOC ONGOING	PRS
10:33	*dāre mire vāġe'an.*	PROX FOC	PRS
15:52	*čikār dāre mikone?*	FOC ONGOING	PRS
36:27	*[- in xānom-i ke xune-ye mā kār mikone-ro nadidin?]* *- dāšt pelehā-ro mišost.* *[-key?* *- piš az zohr-i]*	FOC? ONGOING	PST
40:47	*dāšt mimord man residam.*	PROX(AVER) FOC HYP?	PST
42:08	*man dāram migam be emām-e zamān man az in dargāh-e agar yek ġadam pā-m-o gozāšte bāšam unvar.*	FOC ONGOING	PRS
44:26	*un ru-ye sag-e man-o dāri miāri bālā hā.*	PROX FOC HYP	PRS
56:26	*šomā vaġt-i miumadin man yā rafte budam yā dāštam miraftam sar-e kār.*	PROX FOC	PST
57:33	*- touhin nakon.* *- man touhin dāram mikonam hāj-aġa yā in?*	FOC ONGOING	PRS
01:03:45	*če ġad ham bābā-t negarān bud man dāram miram.*	FUTR ONGOING/PRO X FOC	PRS
01:03:50	*midunest alaki dāri miri.*	FUTR ONGOING/PRO X FOC	PRS
01:07:29	*doxtar-ešon dāšt naġāši mikešid, ye zan-o mard. [goft in bābā-m-e in māmān-am-e. Goftam aziz-am māmān-et ke enġadr čāġ nist]*	FOC ONGOING	PST
01:09:48	*pas če-rā un ruz tu un naġāši-i ke kešidi bābā māmān-et dāštan bā ham da'vā mikardan?*	FOC? ONGOING	PST
01:10:06	*diruz dāre mige man aslan nemidunestam bārdār bude.*	FOC ONGOING	PRS
01:10:08	*emruz dāre mige man aslan hol-eš nadādam.*	FOC ONGOING	PRS
01:11:00	*išun az hamun aval dāran haminjur touhin mikonan.*	INT ITER	PRS
01:18:36	*raftam begam fardā dāran miān barā tahġiġāt, havās-eš bāše.*	FUTR ONGOING/PRO X FOC	PRS
01:22:09	*ya'ni šomā al'ān dārin migin rās migan?*	FOC ONGOING	PRS
01:28:44	*miduni dāri či-kār mikoni?*	FOC ONGOING	PRS

01:28:48	*dāre tahdid mikone.*	FOC ONGOING	PRS
01:33:17	*hamun mouġe ke dāšt bā mo'alem-et tu hāl harf mizad, man az tu āšpazxune hame harfā-šuno mišnidam.*	FOC ONGOING	PST
01:42:11	*dāram mišnavam.*	FOC ONGOING	PRS
01:42:48	*hamin al'ān ke šomā dir kardin man dāštam mimordam.*	HYP PROX(AVER) FOC	PST
01:42:50	*man migam to če-ra dāri tu in mou'zu dexālat mikoni?*	FOC ONGOING	PRS
01:43:28	*in bačče tu sen-e boluġ-e, dāre zajr mikeše tu in vaziat.*	FOC ONGOING	PRS
01:43:57	*dāre zajr mikeše be-ru xod-eš nemiāre.*	FOC ONGOING	PRS
01:54:20	*čera al'ān dāri migi pas?*	FOC ONGOING	PRS
01:54:21	*pas čera al'ān dāri migi pas?*	FOC ONGOING	PRS
01:57:07	*pedar–mādar-et gozāštan be ohde-ye xod-et ke tasmim begiri, hālā ke dāran jodā mišan, az in be ba'd bā kudum-ešun mixāi zendegi koni.*	PROX FOC/FOC ONGOING	PRS

'FELICITY LAND' BY MAZIAR MIRI (2011)

04:36	*širin či-kār dāri mikoni?*	FOC ONGOING	PRS
20:07	*hamin al'ān dāram be do-tā šerkat mošāvere midam*	FOC ONGOING	PRS
23:54	*ki dāre mimire?*	PROX FOC	PRS
23:57	*yek ġavās hāl-eš bad-e, hamin, kas-i dāre nemimire*	HYP PROX FOC	PRS
30:30	*migam šivā ki-e? dāre beh-eš zang mizane.*	PROX FOC	PRS
33:06	*nešaste piš-e tahmine dāre so'āl javāb-eš mikone.*	FOC ONGOING	PRS
33:48	*tahmine, telefon-et dāre zang mixore.*	FOC ONGOING	PRS
38:58	*dāram sekte mikonam.*	HYP PROX FOC	PRS
41:05	*man migam xod-et-o bezan be bixiāli. dāram migam xod-et-o bezan be bixiāli.*	FOC ONGOING	PRS
44:42	*lāle to dāri be man doruġ migi*	FOC ONGOING	PRS
44:54	*azizam či-o dāri az man gāyem mikoni?*	FOC ONGOING	PRS
44:35	*lāle man dāram az negarāni divune mišam, mifāhmi?*	HYP PROX FOC	PRS
51:05	*ayāl dāre emšab be mā xoš migzare hā!*	DUR ONGOING?/FOC ONGOING?	PRS
53:18	*al'ān unvar-e donya nešaste ru wheelchair(?) doxtar-eš injā dāre pus-te mā-ro mikane.*	FOC ONGOING HYP	PRS
53:31	*dah sāl-e dāri hamin mozaxraf-o migi.*	DUR ONGOING INT	PRS

01:08:27	*eh dāri miri zešt-e ke!*	FUTR ONGOING/PRO X FOC	PRS
01:08:28	*injuri dāšt xoš migzašt ke!*	FOC ONGOING	PRS
01:16:26	*- rāsi bā bahrām či migofti?*	DUR ONGOING	PST
	- hiči.	INT	
	- pas nimsā'at tu bālkon dāštin či migoftin?		
'I AM A MOTHER' BY FEREYDON JEYRANI (2012)			
01:24	*dāre mire be samt-e daryā.*	FOC ONGOING	PRS
10:45	*či dāri migi?*	FOC ONGOING	PRS
14:30	*či-kār dāri mikoni?*	FOC ONGOING	PRS
14:30	*dāri ġese migi?*	FOC ONGOING	PRS
15:52	*tāze dārim miresim be zaman-e hāl.*	FUTR ONGOING/PRO X FOC	PRS
18:34	*beh-eš bego man dāram azyat mišam.*	FOC ONGOING	PRS
27:19	*dāram zendegi-m-o mikonam.*	FOC? ONGOING	PRS
30:45	*pas dāšti mirafti xābgāh beh-em zang bezan, xob?*	FUTR ONGOING/PRO X FOC	PST
35:06	*yek riz dāre so'āl mikone.*	INT ITER	PRS
46:37	*dāram mimiram, bāyad beram birun.*	PROX FOC HYP	PRS
01:00:54	*simin dāre taġāzā-ye ġesas mikone.*	PROX FOC/FOC ONGOING	PRS
01:04:20	*hame-či xube, dāram ādat mikonam.*	GRADUAL PROCESS/PROX FOC	PRS
01:09:59	*koja dāri dar miri, hā?*	PROX FOC	PRS
01:12:37	*išon dāran touhin mikonan be doxtar-e man.*	FOC ONGOING	PRS
01:18:38	*to dāri az man enteġām migiri.*	FOC ONGOING	PRS
01:37:37	*dāran miāraneš.*	FOC ONGOING	PRS

E. List of varieties in Chapter 5

The varieties discussed in section 5.1.2 are given below in alphabetical order, together with the place where they are spoken, the DAR and KAR pattern (if they have such) and references.

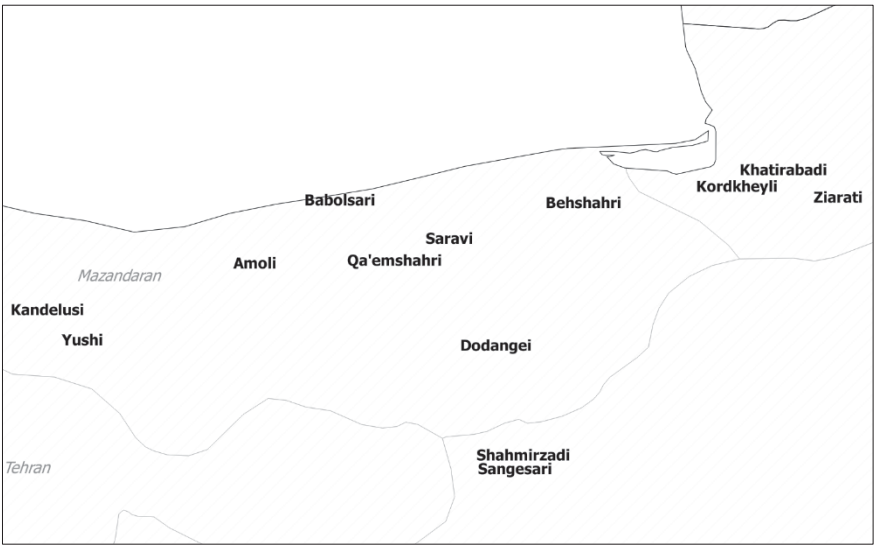
VARIETY	PLACE	PATTERN	REFERENCES
Amoli [M]	Amol	other ('have')	Stilo (forthcoming a.)
Anbaran Ardebil [TAL]	Anbaran	DAR	Paul (2011)
Anbaran Mahalle [TAL]	Anbaran Mahalle	DAR	Paul (2011)
Asalemi [TAL]	Asalem	DAR, KAR, other (*damand*)	Paul (2011), Stilo (in press)
Astara [TAL]	Astara zone	DAR	Paul (2011), Schulze (2000), Stilo (2008)
Azerbaijani, Turkic	area	other	Schulze (2000)
Babolsari [M]	Babolsar	DAR	Stilo (forthcoming a.)
Behshahri [M]	Behshahr	DAR	PROGQ
Budukh, Nakh-Daghestanian	Northern Azerbaijan	other	Alekseev (1994b)
Chali [T]	Chal	DAR	Yarshater (1969)
Charozh [TAL]	Sarak, Digadi	KAR	Stilo (in press)
Daqqushchui (Muslim Tat)	Daqqushchu	other	Noorlander & Stilo (2015)
Dikin Maraqei [M]	Dikin	DAR	Stilo (in press)
Dodanga'i [M]	Dodange	DAR	Stilo (forthcoming a.)
Ebrahim-abadi [T]	Ebrahim-abad	KAR	Yarshater (1969)
Eshtehardi [T]	Eshtehard	KAR-DAR	Yarshater (1969)
Gandomabi [T]	Gandombad	KAR	Stilo (in press)
Gurani, Tatoid (or Upper Zagros and Central Plateau Group)	Gouran	DAR	Stilo (forthcoming a.), Windfuhr (2009)
Harzani [T]	Harzand	DAR	Stilo (in press), Noorlander & Stilo (2015)
Hezarrudi [T]	Hezarrud-e Olya	KAR	Stilo (in press), Yarshater (1970)
Jokandani [TAL]	Jokandani	DAR	Paul (2011)
Juhuri, Judaeo-Tat	Derbent	other	Authier (2012), Noorlander & Stilo (2015)

Kafteji [T]	Kabateh	DAR-KAR	Stilo (forthcoming b.), Stilo (in press)
Kajali [T]	Kejal	KAR	Stilo (in press), Yarshater (1960)
Kalarestaqi, Central Caspian	Rudbarak	DAR	Borjian (2010)
Kalasuri [T/TAL]	Kalasur	other	Noorlander & Stilo (2015), Yarshater (2005)
Kandelusi, Central Carspian	Kandelus	DAR	(Borjian 2016)
Karani [T]	Karin	DAR, KAR	Stilo (in press)
Karnaqi [T]	Karnaq	KAR	Stilo (Stilo in press)
Kelardashti, Central Caspian	Keleardasht	DAR	Stilo (forthcoming a.)
Kelasi [T]	Kelas	KAR	Stilo (forthcoming b.), Stilo (Stilo in press)
Khatirabadi [M]	Khatirabad	DAR	Stilo (forthcoming a.)
Khinalug, Nakh-Daghestanian	Northern Azerbaijan	other	Kibrik (1994)
Khoini [T]	Xoin, Khvoin	DAR	Stilo (in press)
Koluri [T]	Kolur/Kolor	KAR	Stilo (in press), Yarshater (2005)
Kordkheyli [M]	Kord 'e' Kheyl	DAR	Stilo (forthcoming a.)
Kryts, Nakh-Daghestanian	Northern Azerbaijan	other	Saadiev (1994)
Lahijani [G]	Lahijan	DAR, DAR-KAR	Stilo (2001), Stilo (in press)
Lahiji (Muslim Tat)	Lahij	other	Noorlander & Stilo (2015)
Langerudi [G]	Langarud	DAR	Stilo (in press), Stilo (forthcoming a.)
Lerdi [T]	Lerd	DAR	Stilo (in press)
Leriki [TAL]	Lerik	DAR	Paul (2011), Stilo (in press)
Lezgian, Nakh-Daghestanian	Southern Russia	other	Haspelmath (1993)
Masal-Sandermani [TAL]	Masal	KAR	Paul (2011)
Masulei [TAL]	Masouleh	KAR	Lazard (1978), Stilo (in press)
Northern Tati [T]	area	DAR	Stilo (in press)
Qa'emshahri [M]	Qa'emshahr	DAR	Stilo (forthcoming a.)
Ramsari, Gilaki/Central Caspian	Ramsar	DAR	Stilo (in press), Stilo (forthcoming a.)
Rashti [G]	Rasht	DAR, KAR, DAR-KAR	Rastorgueva (2012), Stilo (2001)
Rutul, Nakh-Daghestanian	Northern Azerbaijan	other	Alekseev (1994a)

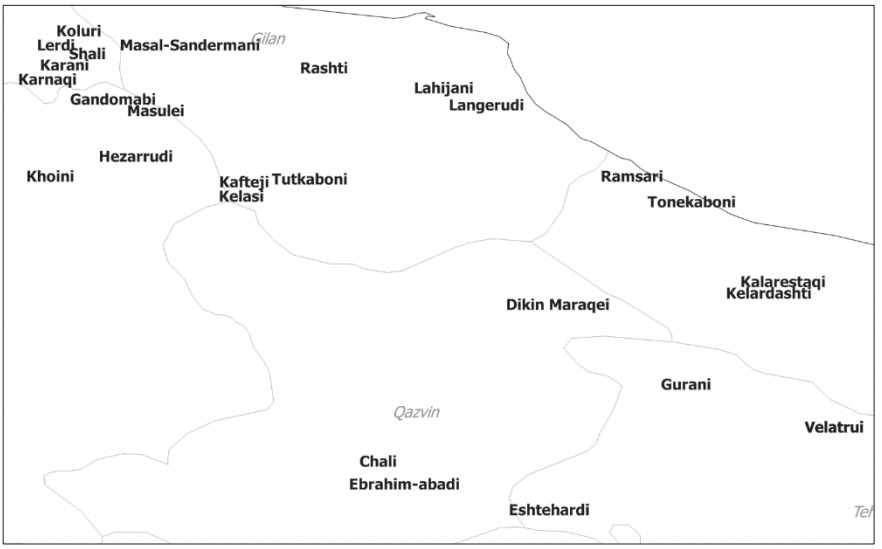
Sangesari, Semnanic	Semnan, Ziarat, Talebabad, dargazin	DAR	Azami & Windfuhr (1974)
Saravi [M]	Sari	DAR	Stilo (forthcoming a.)
Shali [T]	Shal	KAR	Yarshter (1959)
Shahmirzadi [M]	Shahmirzad	DAR	PROGQ
Shuvi [TAL]	not exact	DAR	Schulze (2000)
Tonekaboni, Central Caspian	Tonekabon	DAR	Stilo (in press), Stilo (forthcoming a.)
Turkish (Turkic)	Turkey	other	Göksel & Kerslake (2005) Kornfilt (1997)
Tutkaboni [Tatoid]	Tutkabon	DAR	Stilo (in press)
Udi, Nakh-Daghestanian	Northern Azerbaijan	other	Schulze-Führhoff (1994), Noorlander & Stilo (2015)
Urmi (Christian), Aramaic	Urmia	other	Khan (2008), Noorlander & Stilo (2015)
Urmi (Jewish), Aramaic	Urmia	other	Khan (2008), Noorlander & Stilo (2015)
Velatrui [M]	Velayatrud	DAR	Stilo (forthcoming a.)
Viznei [TAL]	Viznei	DAR	Paul (2011)
Xoynarudi [T/TAL]	Xoynarud	other	Yarshater (2005)
Yushi [M]	Yush	DAR	Stilo (forthcoming a.)
Ziarati [M]	Ziarat	DAR	Shokri, Jahani & Barani (2013)

F. Location of varieties in Chapter 5

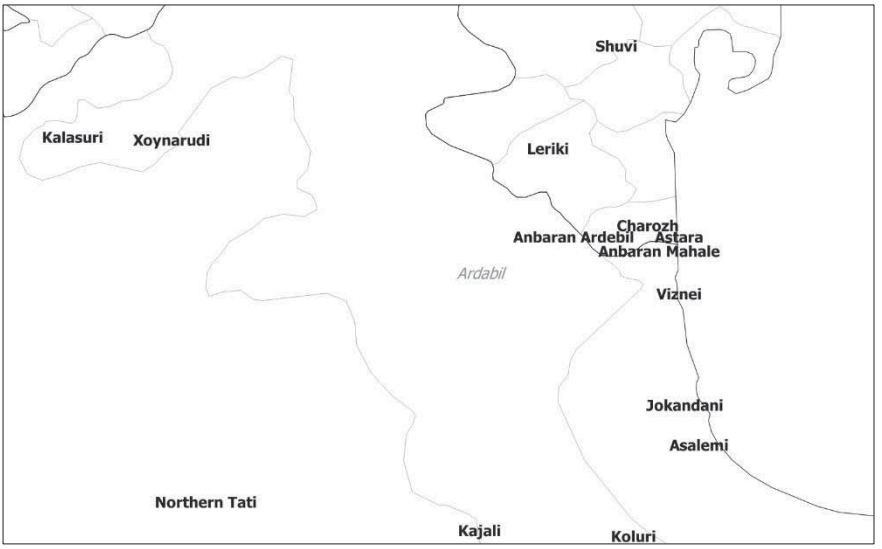
Location of the varieties, discussed in Chapter 5 starting in the east and moving to the west and north.



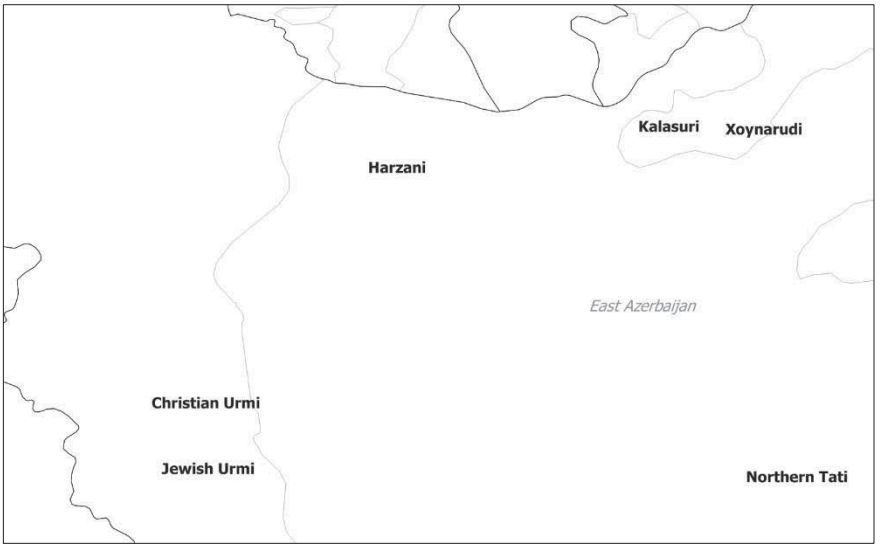
Provinces of Iran: Golestan, Semnan, Mazandaran, Tehran.



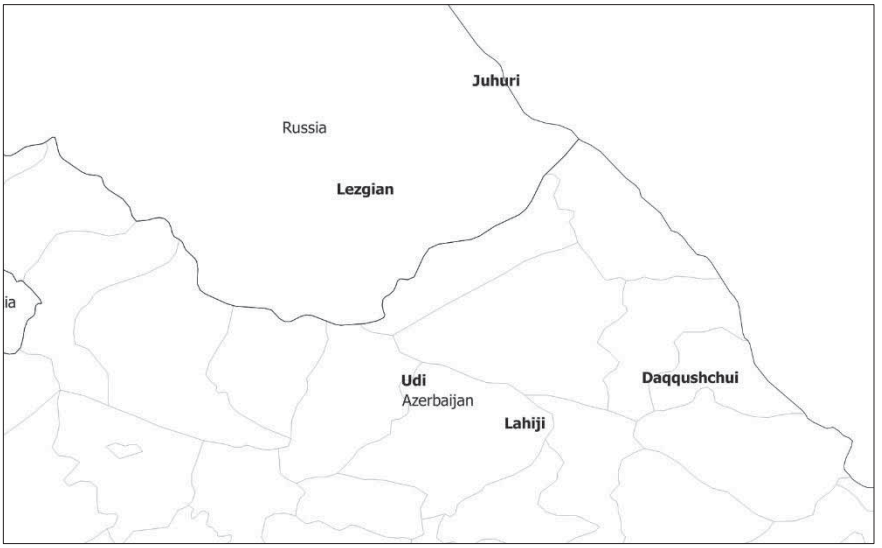
Provinces of Iran: Mazandaran, Alborz, Qazvin, Zanjan, Gilan.



Provinces of Iran: Gilan, Ardabil, East Azerbaijan. Southeastern Azerbaijan.



Provinces of Iran: East Azerbaijan, West Azerbaijan.



Northeastern Azerbaijan and Dagestan, Russia.

G. Constructional schemas in Chapter 5

Constructional schemas for the varieties discussed in section 5.4 with regard to present, past imperfective, present progressive and past progressive functions, in alphabetical order.

PRESENT TENSE	Amoli [M]	V.PRS-(n)-PN
	Anbaran Ardebil[TAL]	VI=DAR(na)=PN1b
	Anbaran Mahalle [TAL]	VI-INF=DAR=PN1b
	Asalemi [TAL]	ba-VII=PN1b
	Babolsari [M]	V.PRS-(n)-PN
	Chali [T]	me/mi-V.PRS-PN
	Eshtehardi [T]	me/mi-V.PRS-PN
	Hezarrudi [T]	m(e)-V.PRS-PN
	Jokandani [TAL]	VI-INF=DAR=PN1b
	Kajali [T]	m(e)-V.PRS-PN
	Kalasuri [T/TAL]	be/ba-VII-INF-PN
	Khatirabadi [M]	V.PRS-(n)-PN
	Lahijani [G]	V.PRS-PN
	Masal-Sandermani [TAL]	VI-PN
	Masulei [TAL]	VI-PN1
	Rashti [G]	V.PRS-PN
	Saravi [M]	V.PRS-(n)-PN
	Shali [T]	V.PRS-PN.PRS
	Viznei [TAL]	VI=DAR=PN1b
	Xoynarudi [T/TAL]	be/ba-VII-INF-PN
	Ziarati [M]	me-V.PRS-PN
PAST IMPERFECTIVE	Amoli [M]	V.PST-PN
	Anbaran Ardebil[TAL]	a-VI-PN1a (?)
	Anbaran Mahalle [TAL]	a-V-PN
	Asalemi [TAL]	a-VI-PN1a or a-VI-i-PN1a
	Babolsari [M]	V.PST-PN
	Chali [T]	me-/mi-V-PST-PN
	Eshtehardi [T]	me-/mi-V.PST-PN
	Hezarrudi [T]	m(e)-V.PST-PN
	Jokandani [TAL]	a-V-PN
	Kajali [T]	m(e)-V.PST-PN
	Kalasuri [T/TAL]	be/ba-V.II-INF=be.PST.PN

	Khatirabadi [M]	V.PST-PN
	Lahijani [G]	V.PST-PN
	Masal-Sandermani [TAL]	VI-PN.IPFV
	Masulei [TAL]	VI-i/y-PN2
	Rashti [G]	V.PST-i-PN
	Saravi [M]	V.PST-PN
	Shali [T]	V.PRS-PN.IPFV
	Viznei [TAL]	a-V-PN
	Xoynarudi [T/TAL]	be/ba-VII-INF-COP.PST.PN
	Ziarati [M]	(m(e))-V.PST-PN
PRESENT PROGGRESSIVE	Amoli [M]	HAVE.PRS-PN V.PRS-(n)-PN
	Anbaran Ardebil[TAL]	-
	Anbaran Mahalle [TAL]	-
	Asalemi [TAL]	KAR=PN VII-INF
	Babolsari [M]	DAR.PRS-3SG V.PRS-(n)-PN
	Chali [T]	V.PST-INF-DAR(u)-COP.PRS.PN
	Eshtehardi [T]	mi/me-V.PRS-PN KAR-DAR-PN
	Hezarrudi [T]	m(e)-V.PRS-PN KAR
	Jokandani [TAL]	-
	Kajali [T]	KAR m(e)-V.PRS-PN
	Kalasuri [T/TAL]	be.PRS.PN VII-INF
	Khatirabadi [M]	DAR.PRS-3SG V.PRS-(n)-PN
	Lahijani [G]	V-INF-DAR.PRS.PN
	Lahijani [G]	V-INF-KAR-DAR.PRS.PN
	Masal-Sandermani [TAL]	KAR V.PRS/1-PN
	Masulei [TAL]	KAR VI-PN1
	Rashti [G]	KAR V-INF-DAR.PRS-PN
	Rashti [G]	V-INF-DAR.PRS-PN
	Rashti [G]	KAR V.PRS-PN
	Saravi [M]	DAR.PRS-3SG V.PRS-(n)-PN
	Saravi [M]	DAR.PRS-PN V.PRS-(n)-PN
	Shali [T]	KAR V.PRS-PN.PRS
	Viznei [TAL]	-
	Xoynarudi [T/TAL]	be.PRS.PN VII-INF
	Ziarati [M]	DAR.PRS-PN (m(e))-V.PRS-PN
	Ziarati [M]	HAVE.PRS-PN me-V.PRS-PN
	Ziarati [M]	HAVE.PRS-3SG me-V.PRS-PN
PAST PROGRESSIVE	Amoli [M]	HAVE.PST-PN V.PST-PN
	Anbaran Ardebil[TAL]	=COP-PN1 V=DAR(na)
	Anbaran Mahalle [TAL]	V-INF=DAR(da)=COP-PN
	Asalemi [TAL]	KAR COP-PN VII-INF

	Asalemi [TAL]	damand=PN VII-INF
	Babolsari [M]	DAR.PST-3SG V.PST-PN
	Chali [T]	?
	Eshtehardi [T]	?
	Hezarrudi [T]	?
	Jokandani [TAL]	V-INF=DAR(da)=COP.PST-PN
	Kajali [T]	?
	Kalasuri [T/TAL]	be.PST.PN VII-INF
	Khatirabadi [M]	DAR.PST-3SG V.PST-PN
	Lahijani [G]	V-INF-DAR.PST.PN
	Lahijani [G]	V-INF-KAR-DAR.PST.PN
	Masal-Sandermani [TAL]	KAR V.PRS/1-PN.IPFV
	Masulei [TAL]	KAR VI-i/y-PN2
	Rashti [G]	KAR V-INF-DAR.PST-PN
	Rashti [G]	V.PST-INF KAR DAR-PRS-PN COP.PST.PN
	Rashti [G]	V-INF-DAR.PST-PN
	Rashti [G]	KAR V.PST-PN
	Saravi [M]	DAR.PST-3SG V.PST-PN
	Saravi [M]	DAR.PST-PN V.PST-PN
	Shali [T]	KAR V.PRS-PN.IPFV
	Viznei [TAL]	V-INF=DAR(da)=COP-PN
	Xoynarudi [T/TAL]	be.PST.PN VII-INF
	Ziarati [M]	DAR.PRS-PN me-V.PST-PN
	Ziarati [M]	?
	Ziarati [M]	?

Sammanfattning på svenska

Denna avhandling undersöker progressiva konstruktioner i flera språk från två huvudperspektiv: dels undersöks progressivens olika användningar och dels undersöks dess struktur såväl som användning i kontaktsituationer. Med *progressiv konstruktion* menas en grammatisk struktur som används för att beteckna handlingar som är pågående vid en specifik tidpunkt. På t.ex. svenska finns konstruktionen *hålla på att/och* följt av ett verb. På engelska finns en konstruktion där verbet *be* 'vara' kombineras med ett verb med ändelsen *-ing*.

1. a) Han håller på att röka.

   b) He is smoking.

En sats med enkelt presens i svenskan, såsom *Jag röker*, kan användas för att referera till något som är generellt giltigt (d.v.s. 'jag är en person som är rökare') men också för en händelse som är pågående vid en specifik tidpunkt (d.v.s. 'jag röker just nu'). Den progressiva satsen däremot används alltid för att referera till en händelse vid en specifik tidpunkt. Tidigare forskning har visat att progressiver kan utvecklas, alltså "grammatikaliseras", till att bli *imperfektiva konstruktioner* såsom markörer för enkelt presens och preteritum imperfektiv.

I denna avhandling undersöker jag progressiver i kontexter med nutids- och dåtidsreferens. Typiska användningar av progressiver såväl som perifera användningar av progressiver med nutids- och dåtidsreferens presenteras, liksom även grammatikaliseringen från progressiv till imperfektiv som den manifesteras i presens och preteritum.

Kapitel 1 innehåller en introduktion till viktiga begrepp och hur de används i avhandlingen samt ger en teoretisk bakgrund.

Kapitel 2 är en tvärspråklig studie där användningen av progressiver i 88 språk undersöks i parallella texter. Detta kapitel visar att i kontexter med nutidsreferens används progressiver oftast för att referera till händelser som är pågående vid en specifik tidpunkt. Däribland används progressiver i kontexter där mottagaren fordras vända sin uppmärksamhet mot den pågående handlingen, exempelvis genom att andra element såsom *titta!* förekommer i yttrandet. I kontexter med dåtidsreferens används progressiver typiskt i narrationer där de utgör en bakgrund till en annan telisk händelse (d.v.s. en händelse som har en naturlig slutpunkt), det är denna teliska händelse som för berättelsen framåt. Nedan ges engelska exempel

ur två olika parallellkorpora där satsen som innehåller progressiv är kursiverad. Exempel 2a) illustrerar en typisk nutidsanvändning och exemplet i 2b) illustrerar en typisk bakgrundsanvändning med dåtidsreferens.

2. Typiska användningar av progressiver med nutids- och dåtidsreferens

   a) The revolution for human equality can happen. *It is happening*. It will happen.

   b) Now it happened that when all the people were baptized, Jesus also was baptized, and *while he was praying*, heaven was opened,

I kapitel 2 visas också att progressiver oftare förekommer med nutidsreferens än med dåtidsreferens. Flera progressiver som endast har nutidsanvändning noteras också. Vidare presenteras progressiver med perifera användningar, såsom förekomster i habituella och performativ-liknande kontexter, användningar med stativa verb och användningar med framtidsreferens presenteras också. Några av dessa konstruktioner föreslås vara progressiver som är på väg att grammatikaliseras mot imperfektiv funktion.

I **Kapitel 3** undersöks användningarna av persiskans *dāštan*-progressiv (*dāštan* har betydelsen 'att ha' som huvudverb). Här fastställer jag att dess huvudsakliga användning betecknar händelser som är pågående under en specifik tidpunkt, men att den även har perifera användningar såsom proximativ ('vara på vippen att hända'), händelser med framtidsreferens, upprepade händelser samt att den kan förekomma i kontexter med durativ referenspunkt ('mellan klockan 2 och 3', 'hela dagen'). De kontexter där konstruktionen förekommer har ofta en *emotiv komponent*, såsom emfas, dramatiskt uttryck, irritation eller överdrift.

Kapitel 4 presenterar tvärspråkliga förklaringar till de användningar som progressiver har i ljuset av några av de resultat som presenterats i Kapitel 2 och 3. Progressivernas användningar knyts samman med typen av händelse som den progressiva konstruktionen refererar till samt den kontext i vilket den förekommer. En skillnad mellan punktuell (ex. 'just nu') och durativ (ex. 'mellan klockan 2 och 3') referenspunkt noteras också: medan den punktuella referenspunkten kan vara implicit, är den durativa referenspunkten alltid explicit given i kontexten.

Kapitel 5 undersöker progressiver i 50 städer och byar i den kaspiska regionen där varieteter av de iranska språken mazandarani, gilaki, taleshi och tati talas. Dessa progressiver antas ha uppkommit som en konsekvens av kontakt mellan dessa varieteter. Konstruktionernas skiftning från presens progressiv till generell presens, och i viss mån även från preteritum progressiv till preteritum imperfektiv, presenteras och diskuteras. Detta kapitel visar även att det finns fler progressiva *konstruktionsscheman* (d.v.s. en strukturell generalisering såsom ex. COP V-INF) än imperfektiva konstruktionsscheman i dessa språk.

Kapitel 6 återvänder till den persiska *dāštan*-progressiven och diskuterar dess ursprung. Närmare bestämt diskuteras huruvida denna konstruktion har uppkommit i kontakt med de konstruktioner som behandlats i Kapitel 5, eller om den uppkommit som en självständig grammatikalisering i persiska. På grund av de slående likheterna mellan de progressiva konstruktionerna i persiska och mazandarani dras slutsatsen att det är troligt att dessa har uppkommit som avbildningar av varandra, men att det inte kan fastställas om det är progressiven i persiska eller mazandarani som varit modell för denna avbildning.

Kapitel 7 innehåller en sammanfattning av resultaten samt en slutdiskussion. En av de slutsatser som denna avhandling kommer fram till är att det finns faktorer som ökar sannolikheten för användningen av en progressiv konstruktion. Dessa återges nedan (3). I progressivens grammatikaliseringsprocess förväntas dessa faktorer bli allt mindre relevanta.

3. Faktorer som ökar sannolikheten för användning av progressiv

- En punktuell referenspunkt.

- Det agentiva subjektets engagemang och upptagenhet med händelsen.

- En emotiv komponent, exempelvis att eventet som refereras till är dramatiskt eller brådskande, eller yttras med en ironisk ton, uttrycker överraskning osv.

- En önskan om att vända mottagarens uppmärksamhet mot den pågående handlingen.

Skillnaden mellan progressivernas frekvens i nutid och dåtid, där nutidsanvändningar visats vara betydligt mer frekventa än dåtidsanvändningar, förklaras med att imperfektiva (oavslutade) yttranden är mer vanliga med nutids- än med dåtidsreferens eftersom saker som sägs vid talögonblicket ofta är oavslutade medan saker som sägs med dåtidsreferens ofta är avslutade. Detta kan ha som konsekvens att progressiva konstruktioner i dåtid, men inte i nutid, inger en känsla av att sakna ett avslut, de är oförmögna att föra narrationen framåt och skapar istället en paus eller platå i berättelser.

4. a) Han läser.

   b) Han höll på å läste… (när plötsligt dörren öppnades)

Denna skillnad mellan nutid och dåtid har den något märkliga konsekvensen att medan progressiver i nutid används för att ge ny, dramatisk och brådskande

information, används progressiver i dåtid, åtminstone narrativ sådan, som bakgrundsinformation till en annan, telisk, händelse som för berättelsen framåt.

Kapitlet ger även en reviderad beskrivning av progressiva konstruktioner:

5. Omarbetad beskrivning av progressiver

> Progressiver är morfosyntaktiska konstruktioner som huvudsakligen används om händelser som är pågående vid referenspunkten. De är inte det primära valet i habituella kontexter eller med stativa predikat och de är inte tillgängliga i prototypiskt generiska kontexter såsom *Kor äter gräs*.

Denna avhandling har med empiriska data visat typiska såväl som mindre typiska användningar av progressiver. Avhandlingen har även visat på att denna konstruktion ofta lånas och omformas i kontaktsituationer. På så sätt har avhandlingen bidragit med nya insikter för förståelsen av denna konstruktion vad gäller dess användningar synkront och har även bidragit till ökad förståelse för dess diakrona utveckling.

Bibliography

Abylkasymova, Mairam & Gulaim Jumabaeva. 1997. *Kyrgyz Language Manual*. Peace Corps, Bishkek (Kyrgyzstan).

Ahangar, Abbas Ali. 2007. A Description of the Verbal System in Sarhaddi Balochi of Granchin. *Orientalia Suecana* 56. 5–24.

Ahangar, Abbas Ali. 2010. The Verb System in the Sistani Dialect of Persian. *Orientalia Suecana* 59. 5–44.

Aikhenvald, Alexandra Y. 2006. A Cross-Linguistic Perspective. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Grammars in Contact: A Cross-Linguistic Typology*, 1–66. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2013. Areal diffusion and parallelism in drift. In Martine Irma Robbeets & Hubert Cuyckens (eds.), *Shared Grammaticalization: With special focus on Transeurasian languages*, 23–42. (Studies in Language Companion Series 132). Amsterdam: John Benjamins Publishing Company.

Alekseev, Mikhail E. 1994a. Rutul. In Rieks Smeets (ed.), *The North East Caucasian Languages, Part 2. Presenting Three Nakh Languages and Six Minor Lezgian Languages*, vol. 4, 213–258. New York: Caravan Books.

Alekseev, Mikhail E. 1994b. Budukh. In Rieks Smeets (ed.), *The North East Caucasian Languages, Part 2. Presenting Three Nakh Languages and Six Minor Lezgian Languages*, vol. 4, 259–296. New York: Caravan Books.

Ameka, F. K. 2006. Grammars in Contact in the Volta Basin (West Africa): On Contact-Induced Grammatical Change in Likpe. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *Grammars in Contact: a Cross-Linguistic Typology*, 114–142. Oxford: Oxford University Press.

Anisimov, N.A. 1932. *Grammatika zuhun tati*. Moscow.

Anonymous. The Uzbek tense/aspect/modality system. http://www.uzbek-glossary.com/TAM/The%20Uzbek%20tense%20aspect%20modality%20system.pdf.

Anthonissen, Lynn, Astrid De Wit & Tanja Mortelmans. 2016. Aspect Meets Modality: A Semantic Analysis of the German am-Progressive. *Journal of Germanic Linguistics* 28(1). 1–30.

Asyik, Abdul Gani. 1987. A Contextual Grammar of Acehnese Sentences. University of Michigan, PhD.

Austin, Peter K. 2012. Tense, aspect and mood in Sasak, eastern Indonesian. In Peter K. Austin & Stuart McGill (eds.), *Language Documentation and*

Description, vol. 11, 231–251. (Working Papers in Sasak). London: SOAS.
Authier, Gilles. 2012. *Grammaire juhuri, ou judéo-tat, langue iranienne des Juifs du Caucase de l'est*. Wiesbaden: Reichert.
Azami, Cheragh Ali & Gernot L Windfuhr. 1972. *A dictionary of Sangesari: with a grammatical outline*. Tehran: Sherkat-e Sahami-ye Katabhaye Jibi.
Bergman, Brita. 1983. Verbs and adjectives: Morphological processes in Swedish Sign Language. In James G. Kyle & B. Woll (eds.), *Language in sign: an international perspective on sign language*, 3–9. London: Croom Helm.
Bertinetto, Pier Marco. 2000. The progressive in Romance, as compared with English. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, vol. 6, 559–664. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Bertinetto, Pier Marco & Denis Delfitto. 2000. Aspect vs. Actionality: Why they should be kept apart. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, vol. 6, 203–226. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Bertinetto, Pier Marco, Karen Ebert & Casper de Groot. 2000a. The progressive in Europe. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, vol. 6, 517–558. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Bertinetto, Pier Marco, Karen Ebert & Casper de Groot. 2000b. Questionnaire on the Progressive Aspect. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, vol. 6, 810–818. (Empirical Approaches to Language Typology 20). Berlin: De Gruyter.
Bhatia, Tej K. 1993. *Punjabi: A Cognitive-Descriptive Grammar*. (Descriptive Grammars Series). London, New York: Routledge.
Blansitt, Edward L. Jr. 1975. Progressive aspect. *Working papers on Language Universals* 18. 1–34.
Blensenius, Kristian. 2015. Progressive constructions in Swedish. Göteborg: Department of Swedish, University of Gothenburg, PhD.
Borjian, Habib. 2004. Māzandarān: Language and the people (The state of research). *Iran and the Caucasus* 8(2). 289–328.
Borjian, Habib. 2008. Two Mazandarani texts from the nineteenth century. *Studia Iranica* 37. 7–49.
Borjian, Habib. 2010. Kalārestāq ii. The Dialect. *Encyclopædia Iranica*, vol. XV, Fasc. 4, 370–373.
Borjian, Habib. 2013a. *Is there continuity between Persian and Caspian? Linguistic relationships in the south-central Alborz*. (American Oriental Series Essay 13). American Oriental Series.
Borjian, Habib. 2013b. Perso-tabaric dialects in the language transition zone bordering mazanderan. *Studia Iranica* 42. 195–225.
Borjian, Habib. 2013c. The Tabaroid dialects of the Central Alborz: Language convergence between Tabari and Persian. *Acta Orientalia* 66. 427–441.
Borjian, Habib. 2016. Kojour ii. Language. *Encyclopædia Iranica*.

Bybee, Joan. 1985. *Morphology*. Amsterdam: John Benjamin's Publishing Company.
Bybee, Joan & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13. 51–103.
Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar. Tense Aspect, and Modality in the Languages of the World*. London: The University of Chicago Press.
Center for International Earth Science Information Network - CIESIN - Columbia University & Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid. NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4XK8CG2.
Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb*. Leiden: Brill.
Childs, G. Tucker. 1995. *A Grammar of Kisi: A Southern Atlantic Language*. (Mouton Grammar Library 16). Berlin: Mouton de Gruyter.
Clifton, John M., Calvin Tiessen, Gabriela Deckinga & Laura Lucht. 2005. The sociolinguistic situation of the Talysh in Azerbaijan. In John M. Clifton (ed.), *Studies in languages of Azerbaijan*, vol. 2. Baku: SIL International.
Comrie, Bernard. 1976. *Aspect. An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
Craig, Colette G. 1977. *The Structure of Jacaltec*. reprint ed. Austin: University of Texas Press.
Croft, William. 2013. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press.
Csató, Éva Ágnes. 2001. Present in Kashkay. *Turkic Languages* 5. 104–119.
Csató, Éva Ágnes. 2005. On copying in Kashkay. In Éva Ágnes Csató, Bo Isaksson & Carina Jahani (eds.), *Linguistic Convergence and Areal Diffusion. Case studies from Iranian, Semitic and Turkic*, 271–284. London: RoutledgeCurzon.
Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: using translational equivalents in linguistic typology. *STUF – Sprachtypologie und Universalienforschung* 60(2). 95–99.
Dahl, Östen. forthcoming a. Indo-European perfects in typological perspective.
Dahl, Östen. forthcoming b. "Universal" readings of perfects in typological perspective.
Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
Dahl, Östen. 2000a. The tense-aspect systems of European languages in a typological perspective. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, vol. 6, 3–25. (Empirical Approaches to Language Typology 20). Berlin: de Gruyter.
Dahl, Östen. 2000b. The grammar of future time reference in European languages. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, vol. 6, 309–328. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Dahl, Östen. 2001. Principles of Areal Typology. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.),

Language typology and language universals: An international handbook, vol. 2, 1456–1470. Berlin: Mouton de Gruyter.
Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
Dahl, Östen. 2013. How Telicity Creates Time. *Journal of Slavic Linguistics* 21(1). 45–76.
Dahl, Östen. 2014. The perfect map: Investigating the cross-linguistic distribution of TAME categories in a parallel corpus. In Bemedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech*, 268–289. Berlin: Walter de Gruyter.
Dahl, Östen & Bernhard Wälchli. 2016. Perfects and iamitives: two gram types in one grammatical space. *Letras de Hoje* 51(3). 325–348.
Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: A corpus-based study of derivation*. Berlin: Mouton de Gruyter.
Davari, Shadi & Mehrdad Naghzguy-Kohan. 2017. The grammaticalization of progressive aspect in Persian. *The Grammaticalization of Tense, Aspect, Modality and Evidentiality: A Functional Perspective*, vol. 311. Berlin: De Gruyter Mouton.
Davidsen-Nielsen, Niels. 1990. *Tense and mood in English: A comparison with Danish*. Berlin: Mouton de Gruyter.
Day, Christopher. 1973. *The Jacaltec Language*. (Indiana University Publications: Language Science Monographs 12). Bloomington: Indian University.
de Groot, Casper. 2000. The absentive. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, vol. 6, 693–722. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
de Vries, Lourens. 2009. Some remarks on the use of Bible translations as parallel texts in linguistic research. *STUF - Sprachtypologie und Universalienforschung* 60(2). 148–157.
De Wit, Astrid & Frank Brisard. 2014. A Cognitive Grammar account of the semantics of the English present progressive. *Journal of Linguistics* 50(1). 49–90.
De Wit, Astrid, Adeline Patard & Frank Brisard. 2013. A contrastive analysis of the present progressive in French and English. *Studies in Language* 37(4). 846–879.
Dehghan, I. 1972. Dāštan as an auxiliary in contemporary Persian. *Archiv Orientalni* 40. 198–205.
Dehkhoda, Ali Akbar. 1907. Čarand Parand. *Shur-e Esrafil*. Tehran.
Dem, Sierra. 1995. *Wollof grammar manual*. Banjul: Peace Corps The Gambia.
Diouf, Jean-Léopold. 2009. *Grammaire du wolof contemporain*. Revue et complétée. Paris: L'Harmattan.
Doerfer, Gerhard. 1988. Azerbaijan viii. Azeri Turkish. *Encyclopædia Iranica*, vol. III, Fasc. 3, 245–248.

Doerfer, Gerhard, Wolfram Hesche & Jamshid Ravanyar. 1990. *Oghusica aus Iran*. Wiesbaden: Harrassowitz.
Dowty, D. 1977. Towards a semantic analysis of verb aspect and the English 'imperfective' progressive. *Linguistics and Philosophy* 1. 45–77.
Dryer, Matthew S. 2013. *A grammatical description of Kara-Lemakot*. (Asia-Pacific Linguistics 009). Canberra: Asia-Pacific Linguistics.
Durkin-Meisterernst, Desmond. 2009. Khwarezmian. In Gernot Windfuhr (ed.), *The Iranian languages*, 336–376. London: Routledge.
Dzablu-Kumah, Simon Wellington. 2006. *Basic Ewe: for foreign students*. (Ed.) Ulrike Claudi & Johannes Ayao Ossey. Köln: Institut für Afrikanistik.
Eades, Yusuf. 1998. Relativization. In Peter K. Austin (ed.), *Sasak*, vol. 1, 119–129. (Working Papers in Sasak). Melbourne: University of Melbourne, Lombok and Sumbawa Research Project.
Ebert, Karen. 2000. Progressive Markers in Germanic Languages. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, vol. 6, 605–653. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Elbert, Samuel H. & Mary Kawena Pukui. 1979. *Hawaiian Grammar*. Honolulu: University of Hawaii Press.
Əfəndiyeva, Türkan. 2005. *Müasir Azərbaycan dilində felin keçmiş zaman formaları*. Baku: Nurlan.
Field, Fredric W. 2002. *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins.
Fortson IV, Benjamin W. 2014. *Indo-European Language and Culture: An Introduction*. Oxford: Blackwell Publishing.
Garbell, Irene. 1965. *The Jewish Neo-Aramaic Dialect of Persian Azerbaijan*. The Hauge: Mouton de Gruyter.
Gast, Volker & Johan van der Auwera. 2012. What is 'contact-induced grammaticalization'? Examples from Mayan and Mixe-Zoquean languages. In Björn Wiemer, Bernhard Wälchli & Björn Hansen (eds.), *Grammatical Replication and Borrowability in Language Contact*, 381–426. Berlin: De Gruyter.
Göksel, Asli & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.
Greenberg, Joseph Harold. 1978. Diachrony, synchrony, and language universals, I: Method and theory. In Joseph Harold Greenberg (ed.), *Universals of human language*. Stanford: Stanford University Press.
Grjunberg, A. L. 1963. *Jazyk Severoazerbaidzjanskikh Tatov*. Moscow: Akademia Nauk SSSR.
Haiman, John. 2011. *Cambodian (Khmer)*. (London Oriental and African Language Library 16). Amsterdam: Benjamins.
Haspelmath, Martin. 1993. *A Grammar of Lezgian*. (Mouton Grammar Library 9). Berlin: De Gruyter Mouton.
Haspelmath, Martin. 1998. The Semantic Development of Old Presents: New futures and subjunctives without grammaticalization. *Diachronica* 15(1). 29–62.

Heath, Jeffrey. 1978. *Linguistic diffusion in Arnhem Land*. (Australian Aboriginal Studies: Research and Regional Studies 13). Canberra: AIAS.
Heath, Jeffrey. 2014. A Grammar of Yorno-So.
Heinämäki, Orvokki. 1995. The progressive in Finnish: Pragmatic constraints. In Pier Marco Bertinetto, Valentina Bianchi, Östen Dahl & Mario Squartini (eds.), *Temporal reference, aspect and actionality.*, 143–153. (Typological Perspectives 2). Torino: Rosenberg & Sellier.
Heine, Bernd. 2002. On the role of context in grammaticalization. *Typological studies in language* 49. 83–102.
Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Heine, Bernd & Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529–572.
Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
Heine, Bernd & Tania Kuteva. 2006. *Changing Languages of Europe – Oxford Scholarship*. Oxford: Oxford University Press.
Heise, Jennifer. 2003. Participant reference and tracking in San Francisco Ozolotepec Zapotec. Graduate Institute of Applied Linguistics, Master thesis.
Hekking, Ewald & Dik Bakker. 2007. The case of Otomi: A contribution to grammatical borrowing in cross-linguistic perspective. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 435–464. New York: Mouton de Gruyter.
Henning, Walter B. 1934. Das Verbum des Mittelpersischen der Turfanfragmente. *Zeitschrift für Indologie und Iranistik* 9.
Hieda, Osamu. 2011. *Kumam Vocabulary with Grammatical Notes*. (Studies in Nilotic Linguistics 4). Tokyo: Research Institute for Languages and Cultures of Asia and Africa.
Hopkins, Bradley L. 1995. *Contribution a une etude de la syntaxe Diola-Fogny*. (Cahiers de Recherche Linguistique 4). Dakar: Société Internationale de Linguistique.
Hopper, Paul. 1979. Aspect and foregrounding in discourse. In Talmy Givón (ed.), *Discourse and Syntax*, 213–241. (Syntax and Semantics 12). New York: Academic Press.
Horn, Laurence R. 1985. Metalinguistic Negation and Pragmatic Ambiguity. *Language* 61(1). 121–174.
Hüseynova, Gülsüm. 2002. *Lahıc Tatlarının Dili*. Baku: Nurlan.
Jahani, Carina. 2008. Expressions of future in classical and modern new Persian. In Simin Karimi, Vida Samiian & Donald L. Stilo (eds.), *Aspects of Iranian linguistics*, 153–176. Newcastle upon Tyne: Cambridge Scholars Publishing.
Jahani, Carina. 2017. Prospectivity in Persian and Balochi and the preterite for non-past events. In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond*. Wiesbaden: Reichert Verlag.

Janda, Laura A. & Olga Lyashevskaya. 2011. Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cognitive Linguistics* 22(4). 719–762.
Jeremiás, Eva M. 1993. On the Genesis of the Periphrastic Progressive in Iranian Languages. *Medioiranica*, 99–116. Louvain-La-Neuve: Peeters.
Jespersen, Otto. 1924. *The Philosophy of Grammar*. London: George Allen and Unwin Ltd.
Johanson, Lars. 2000. Viewpoint operators in European languages. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, vol. 6, 27–188. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Johanson, Lars. 2002. *Structural Factors in Turkic Language Contacts*. Richmond: Curzon Press.
Johanson, Lars. 2017. Prospectives and proximatives. In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond*. Wiesbaden: Reichert Verlag.
Jondoh, Edina Ayaba Elemawussi. 1980. Some aspects of the predicate phrase in Gegbe. Indiana University, PhD.
Jügel, Thomas, Cemile Çelebi & Diako Nahid. 2017. Deixis and near future expressions in Kurdish. In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond*, 163–186. Wiesbaden: Reichert Verlag.
Kagaya, Ryohei. 1987. *A Classified Vocabulary of the Lenje Language*. (Bantu Vocabulary Series 4). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
Karlsson, Fred. 2015. *Finnish: An Essential Grammar*. 3rd ed. (Routledge Essential Grammars). London: Routledge.
Keshavarz, Karim. 1962. Mozāre' va māzi-ye malmus. *Rāhnemā-ye Ketāb*, 687–694. (5).
Khan, Geoffrey. 2008. *The Jewish neo-Aramaic dialect of Urmi*. Piscataway: Gorgias Press.
Kibrik, Aleksandr E. 1994. Khinalug. In Rieks Smeets (ed.), *The North East Caucasian Languages, Part 2. Presenting Three Nakh Languages and Six Minor Lezgian Languages*, vol. 4, 367–406. New York: Caravan Books.
Kim, Nam-Kil. 1986. The progressive in Korean. In Kim, Nam-Kil (ed.), *Studies in Korean language and linguistics*, 97–117. Los Angeles: East Asian Studies Center, University of Southern California.
Koptjevskaja-Tamm, Maria. 2010. Linguistic Typology and Language Contact. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*. (Oxford Handbook Online). Oxford University Press.
Korn, Agnes. 2016. A partial tree of Central Iranian. A new look at Iranian subphyla. *Indogermanische Forschungen* 121(1). 401–434.
Korn, Agnes. 2017. What to look out for: Morphology of prospectives and futures in Iranian. In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond*. Wiesbaden: Reichert Verlag.

Korn, Agnes & Irina Nevskaya (eds.). 2017. *Prospective and Proximative in Turkic, Iranian and beyond*. Wiesbaden: Reichert Verlag.
Kornfilt, Jaklin. 1997. *Turkish*. (Descriptive Grammars Series). London: Routledge.
Kuteva, Tania. 2001. *Auxiliation: An enquiry into the nature of grammaticalization*. Oxford: Oxford University Press.
Kvist Darnell, Ulrika. 2008. Pseudosamordningar i svenska: särskilt sådana med verben sitta, ligga och stå. Stockholm: Stockholm University.
Lambton, Ann K. S. 1957. *Persian grammar*. Reprinted with corrections. Cambridge: Cambridge University Press.
Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.
Lazard, Gilbert. 1978. Le Dialecte tâleši de Mâsule (Gilân). *Studia Iranica* 7. 251–268.
Lazard, Gilbert. 1989. Le Persan. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 263–293. Wiesbaden: Reichert Verlag.
Lazard, Gilbert. 1992. *A grammar of contemporary Persian*. (Trans.) Shirley A. Lyon. Original title: Grammaire du Persan contemporain (1957). Cosa Mesa California: Mazda Publishers in association with Bibliotheca Persica.
Lindstedt, Jouko. 2000. The perfect – aspectual, temporal and evidential. *Tense and aspect in the languages of Europe*, vol. 6, 365–383. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Ljung, Magnus. 1980. *Reflections on the English progressive*. (Gothenburg Studies in English 46). Göteborg: Acta Universitatis Gothoburgensis.
Mace, John. 2003. *Persian grammar: for reference and revision*. London: RoutledgeCurzon.
Mahootian, Shahrzad. 1997. *Persian*. London: Routledge.
Marchese, Lynell. 1986. *Tense/aspect and the development of auxiliaries in Kru languages*. (SIL Publications in Linguistics 78). Dallas: Summer Institute of Linguistics.
Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36. 281–331.
Matras, Yaron. 2000. How predictable is contact-induced change in grammar? In Colin Renfrew, April McMahon & Larry Trask (eds.), *Time depth in historical linguistics*, vol. 2, 585–620. Cambridge: The McDonald Institute for Archaeological Research.
Matras, Yaron. 2007. The borrowability of structural categories. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical Borrowing in Cross-Linguistic Perspective*, 31–74. Berlin: De Gruyter.
Matras, Yaron & Jeanette Sakel. 2004. *Database of Convergence and Borrowing*. Manchester: University of Manchester.
Matras, Yaron & Jeanette Sakel. 2007a. Investigating the mechanism of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.

Matras, Yaron & Jeanette Sakel. 2007b. Introduction. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*. New York: Mouton de Gruyter.
Miestamo, Matti. 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. (Empirical Approaches to Language Typology 31). Berlin: Mouton de Gruyter.
Miestamo, Matti & Johan van der Auwera. 2011. Negation and perfective vs. imperfective aspect. In Jesse. Mortelmans, Tanja. Mortelmans & Walter. De Mulder (eds.), *From now to eternity*, 65–84. (Cahiers Chronos 22). Amsterdam: Rodopi.
Miller, Boris V. 1953. *Talyšskij jazyk*. Moscow: Akademija Nauk SSSR.
Mittwoch, Anita. 1988. Aspects of English aspect: On the interaction of perfect, progressive and durational phrases. *Linguistics and Philosophy* 11(2). 203–254.
Moravcsik, Edith A. 1975. Borrowed Verbs. *Wiener Linguistische Gazette* 8. 3–30.
Moravcsik, Edith A. 1978. Language contact. In Joseph Harold Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human languages*, vol. 1. Stanford: Stanford University Press.
Narumov, Boris Petrovich. 1987. *Formirovanie romanskij literaturnyj iazykov: sovremennyï galisiïskiï iazyk*. Moskva: Nauka.
Nematollahi, Narges. in press. HAVE-progressive in Persian: a case of pattern replication? *Diachronica*.
Nematollahi, Narges. 2014. [Proceeding] Development of the progressive construction in Modern Persian. In Öner Özçelik & Amber Kennedy Kent (eds.), *Proceedings of the 1st Conference on Central Asian Languages and Linguistics*, vol. 1, 102–112. Bloomington, Indiana.
Noorlander, Paul M. 2017. The Proximative and its correlatives in North Eastern Neo-Aramaic. In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond*, 187–212. Wiesbaden: Reichert Verlag.
Noorlander, Paul M. & Donald L. Stilo. 2015. On the Convergence of Verbal Systems of Aramaic and its Neighbours. Part I: Present-Based Paradigms. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its Linguistic Context*, 426–452. (Georgiaa Neo-Aramaic Studies 14). New York: Georgias Press LLC.
Nyberg, Henrik S. 1974. *A manual of Pahlavi 2: Ideograms, glossary, abbreviations, index, grammatical survey, corrigenda to Part 1*. Wiesbaden: Harrassowitz.
Östling, Robert. 2015. Bayesian models for multilingual word alignment. Stockholm: Department of Linguistics, Stockholm University, PhD.
Paul, Daniel. 2011. A Comparative Dialectal Description of Iranian Taleshi. Manchester: The University of Manchester.
Perry, John R. 2005. *A Tajik Persian Reference Grammar*. Boston: Brill.
Phan, Trang. 2013. Syntax of Vietnamese Aspect. Sheffield: University of Sheffield, PhD.

Phillott, D. C. 1919. *Higher Persian grammar for the use of the Calcutta University, showing differences between Afghan and modern Persian, with notes on rhetoric.* Calcutta: Calcutta University.
Pistoso, Maurizio. 1974. L'ausiliare Daštan in neo‑persiano: un costrutto linguistico nord‑iranico? *Oriente Moderno* 54(4). 157–162.
Pixabaj, Telma Can. 2006. *Gramática Descriptiva Uspanteka*. Ciudad de Guatemala: OKMA – Proyecto de documentación: idioma Uspanteko.
Po-Ching, Yip & Don Rimmington. 2004. *Chinese: A Comprehensive Grammar*. (Routledge Comprehensive Grammars). London & New York: Routledge.
Polian, Gilles. 2013. *Gramática del tseltal de Oxchuc*. Ciudad de México: Centro de Investigaciones y Estudios Superiores en Antropología Social.
Poulos, George. 1990. *A Linguistic Analysis of Venda*. Pretoria: Via Afrika.
Rafiei, Adel. 2017. Prospectivity in Persian and Azeri Turkish (Iran). In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond*, 253–259. Wiesbaden: Reichert Verlag.
Rastorgueva, V. S, Aza Alimovna Kerimova, Akhmed Kerimovich Mamedzade, L. A Pireĭko & D. I Ėdel′man. 2012. *The Gilaki language*. (Trans.) Ronald M Lockwood. Original title: Giljanskij jazyk (1971). (Acta Universitatis Upsaliensis, Studia Iranica Upsaliensia 19). Stockholm: Elanders Sverige AB.
Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: Macmillian & Co.
Reichle, Verena. 1981. *Bawm Language and Lore: Tibeto-Burman Area*. (Europäische Hochschulschriften: Reihe XXI: Linguistik 14). Frankfurt am Main: Peter Lang.
Ricca, Davis. 1998. Una perifrasi continua-iterative nei testi piemontesi dal Cinquecento all'Ottocento: TENERE + participio passato. In Paolo Ramat & Roma, Elisa (eds.). Roma: Bulzoni.
Robert, Stéphane. 1989. Système verbal et énonciation en Wolof. Université de Paris 7 – Dénis Diderot, PhD.
Rubongoya, L. T. 1999. *A Modern Runyoro-Rutooro Grammar*. (East African Languages and Dialects 9). Cologne: Rüdiger Köppe.
Saadiev, Sh. M. 1994. Kryts. In Rieks Smeets (ed.), *The North East Caucasian Languages, Part 2: Presenting Three Nakh Languages and Six Minor Lezgian Languages*, vol. 4, 407–446. New York: Caravan Books.
Saidi Sirjani, A-A. 1994. Dehkhoda, Mirza Ali-Akbar Qazvini. *Encyclopædia Iranica*, vol. VII, Fasc. 2, 216–220.
Sakel, Jeanette. 2007. Types of loan: Matter and pattern. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 15–29. New York: Mouton de Gruyter.
Sapir, David J. 1965. *A Grammar of Diola-Fogny: A Language Spoken in the Basse-Casamance Region of Senegal*. (West African Language Monograph Series 3). Cambridge: Cambridge University Press.
Schönig, Claus. 1998. Azerbaijanian. In Lars Johanson & Éva Ágnes Csató (eds.), *The Turkic languages*, 248–260. London: Routledge.

Schulze, Wolfgang. 2000. *Northern Talysh*. (Languages of the World/Materials 380). Lincom Europe.
Schulze-Führhoff, Wolfgang. 1994. Udi. In Rieks Smeets (ed.), *The North East Caucasian Languages, Part 2: Presenting Three Nakh Languages and Six Minor Lezgian Languages*, vol. 4, 447–514. New York: Caravan Books.
Serebrennikov, Boris & Ninel′ Gadžieva. 1979. *Sravnitel'no-istoričeskaja grammatika tjurkskix jazykov*. Baku: Maarif.
Shahidi, Minoo. 2008. A Sociolinguistic Study of Language Shift in Mazandarani. Uppsala: Uppsala University.
Shokri, Guiti, Carina Jahani & Hossein Barani. 2013. *When tradition meets modernity: Five life stories from the Galesh community in Ziarat, Golestan, Iran*. (Acta Universitatis Upsaliensis, Studia Iranica Upsaliensia 21). Uppsala: Uppsala Universitet.
Simpson, Cyril Gordon. 1957. *The Turkish language of Soviet Azerbaijan: Selections from the press and from modern authors with a grammatical introduction*. Oxford: Central Asian Research Centre.
Skjærvø, Prods O. 2009. Middle West Iranian. In Gernot Windfuhr (ed.), *The Iranian languages*, 196–278. London: Routledge.
Smith, Carlota. S. 1997. *The Parameter of Aspect*. 2nd ed. (Studies in Linguistics and Philosophy 43). Dordrecht: Kluwer Academic Publishers.
Smyth, David. 2002. *Thai: An essential grammar*. London: Routledge.
Sneddon, James Neil. 2006. *Colloquial Jakartan Indonesian*. (Pacific Linguistics 581). Canberra: Research School of Pacific and Asian Studies, Australian National University.
Sneddon, James Neil, Alexander Adelaar, Dwi Noverini Djenar & Michael C. Ewing. 2010. *Indonesian Reference Grammar*. 2nd ed. Sydney: Allen & Unwin.
Soga, Matsuo. 1983. *Tense and Aspect in Modern Colloquial Japanese*. Vancouver: University of British Columbia Press.
Stilo, Donald L. in press. Caspian and Tatic. In Geoffrey Khan & Geoffrey Haig (eds.), *Language contact and language change in west Asia*. Berlin: De Gruyter Mouton.
Stilo, Donald L. forthcoming a. Mazandarani (to appear in Encyclopædia Iranica).
Stilo, Donald L. forthcoming b. Loss vs. Expansion of Gender in Tatic Languages: Kafteji (Kabatei) and Kelasi.
Stilo, Donald L. 1981. The Tati Language Group in the Sociolinguistic Context of Northwestern Iran and Transcaucasia. *Iranian Studies* 14(3/4). 137–187.
Stilo, Donald L. 2001. Gilan x. Languages. *Encyclopædia Iranica*, vol. X, Fasc. 6, 660–668.
Stilo, Donald L. 2003. Hamadan ix. Jewish dialect. *Encyclopædia Iranica*, vol. XI, Fasc. 6, 623–627.
Stilo, Donald L. 2007a. Isfahan xxii. Gazi dialect. *Encyclopædia Iranica*, vol. XIV, Fasc. 1, Fasc. 2. 112-119.

Stilo, Donald L. 2007b. Isfahan xix. Jewish Dialect. *Encyclopædia Iranica*, vol. XIV, Fasc. 1, 77–84.
Stilo, Donald L. 2008. Person Agreement Markers in the Northern Talyshi Language. In Simin Karimi, Donald L. Stilo & Vida Samiian (eds.), *Aspects of Iranian linguistics*, 363–390. Newcastle upon Tyne: Cambridge Scholars Publishing.
Stilo, Donald L. & Paul M. Noorlander. 2015. On the Convergence of Verbal Systems of Aramaic and its Neighbours. Part II: Past Paradigms derived from Present Equivalents. In Geoffrey Khan & Lidia Napiorkowska (eds.), *Neo-Aramaic and its Linguistic Context*, 453–484. (Georgiaa Neo-Aramaic Studies 14). New York.
Suleymanov, Murad. 2015. Les catégories morphosémantiques du verbe en azéri et en tat: Un cas de métatypie? Paris: Institut national des langues et civilisations orientales, Master thesis.
Thompson, Laurence C. 1987. *A Vietnamese Reference Grammar.* (Mon Khmer Studies 13–14). Honolulu: University of Hawaii Press.
Timberlake, Alan. 2007. Aspect, tense, mood. In Timothy Shopen (ed.), *Language typology and syntactic description*. 2nd ed. Cambridge: Cambridge University Press.
Tissen, C. 2003. Positive Orientation Towards the Vernacular among the Talysh of Sumgayit. University of North Dakota, Master thesis.
Tommola, H. 2000. Progressive Aspect in Baltic Finnic. In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, vol. 6, 655–692. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.
Utas, Bo. 2013. The Grammatical Transition from Middle to New Persian. In Bo Utas, Carina Jahani & Mehrdad Fallahzadeh (eds.), *From Old to New Persian: Collected essays*, 251–259. (Beiträge Zur Iranistik 38). Weisbaden: Reichert.
van den Berg, René. 1989. A Grammar of the Muna Language. Rijksuniversiteit te Leiden, PhD.
van den Berg, René & Robert L. Busenitz. 2012. *A grammar of Balantak, a language of Eastern Sulawesi*. SIL International.
Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review* 66(2). 143–160.
Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.
Wälchli, Bernhard. 2000. Livonian in a genetic, areal and typological perspective, or is Finnish better Finnic than Livonian? In Johanna Laakso (ed.), *Facing Finnic: Some challenges to historical and contact liguistics*, 210–226. Tartu, Estonia.
Wälchli, Bernhard. 2012. Grammaticalization clines in space: Zooming in on synchronic traces of diffusion processes. In Björn Wiemer, Bernhard Wälchli & Björn Hansen (eds.), *Grammatical Replication and Borrowability in Language Contact*, 233–272. Berlin: De Gruyter Mouton.

Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. New York: Publications of the Linguistic Circle of New York.
Wiemer, Björn & Bernhard Wälchli. 2012. Contact-induced grammatical change: Diverse phenomena, diverse perspectives. In Björn Wiemer, Bernhard Wälchli & Björn Hansen (eds.), *Grammatical Replication and Borrowability in Language Contact*, 3–64. (Trends in Linguistics. Studies and Monographs [TiLSM] 242). Berlin: De Gruyter Mouton.
Windfuhr, Gernot. 1979. *Persian grammar: History and state of its study*. The Hague: Mouton.
Windfuhr, Gernot. 1989a. New West Iranian. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 251–262. Wiesbaden: Reichert Verlag.
Windfuhr, Gernot. 1989b. New Iranian Languages: Overview. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 246–250. Wiesbaden: Reichert Verlag.
Windfuhr, Gernot. 2009. *The Iranian languages*. London: Routledge.
Windfuhr, Gernot & John R. Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian languages*, 416–545. London: Routledge.
Yarshater, Ehsan. 1959. The Dialect of Shāhrud (Khalkhāl). *Bulletin of the School of Oriental and African Studies, University of London* 22(1/3). 52–68.
Yarshater, Ehsan. 1960. The Tāti Dialect of Kajal. *Bulletin of the School of Oriental and African Studies, University of London* 23(2). 275–286.
Yarshater, Ehsan. 1969. *A grammar of Southern Tati dialects*. The Hague: Mouton.
Yarshater, Ehsan. 1970. The Tati dialects of Ṭārom. In Mary Boyce & Ilya Gershevitch (eds.), *The Henning Memorial Volume*, 451–467. London: School of Oriental and African Studies, University of London.
Yarshater, Ehsan. 1996. The Taleshi of Āsalem. *Studia Iranica* 25. 83–113.
Yarshater, Ehsan. 2005. The Tati dialect of Kalasur. In Dieter Weber (ed.), *Languages of Iran: Past and present*, 269–284. Wiesbaden: Harrassowitz.
Žukovskij, V. 1888. *Osobennoe značenie glagola dāštan a persidskom razgovornom yazyke (Special meaning of the verb (dāštan) in spoken Persian)*. 3rd ed. St. Petersburg.