Bild på omslaget: "Ark of Bukhara". 1929. Oil on canvas. Lev Leonardovich Bure (1887–1943)

Kph Trycksaksbolaget AB, Uppsala 2022
Evidentiality in Tajik

Alexander Nilsson
The focus of this study is evidentiality—the grammatical marking of information source—in written and spoken Tajik. For marking evidential statements, Tajik employs a secondary register of verb forms (called register II forms) that all end in -a: the Perfect and the Perfectoids (the Perfectoid Imperfect and Perfectoid Pluperfect). Whereas the Perfect is ambiguous as regards its evidential status, the Perfectoids are unambiguously non-first hand in nature. The register II forms can signal that propositions are reportative or inferential. Moreover, they are used for mirative purposes, i.e. to signal new information or surprise. In addition to the register II forms, Tajik also employs a special verb form called the Presumptive, which is also included in this study.

The analysis of written Tajik is based on a close reading of 424 pages of prose fiction texts that can be divided into three groups in accordance with the time periods when they were written: the 1920s–1940s, 1980s and 2010s. The analysis of spoken Tajik is based on ca 9.5 hours of recorded material gathered in Dushanbe, Tajikistan, in the years 2016 and 2017. While the 2016 recordings were in the form of free conversations, the 2017 recordings involved a collaborative exercise developed as a method for investigating grammatical categories related to social cognition.

The results of the analyses show that the evidential uses of the register II forms and the Presumptive are more prevalent in the older texts than in the newer texts. As for the spoken language, the register II forms are rarely used for reportative statements. Instead, evidential lexical markers are used with register I forms. However, speakers tend to use register II forms in cases where there is greater temporal distance to the subject matter. Inferences are also usually made with register I forms together with evidential lexical markers. However, the Perfect is employed in the creation of coherent hypothetical narratives.

In conclusion, it is argued that the Tajik “mirative” is in fact evidential, since it marks direct experience of two types: new information (mirative) and personal impression (evaluative). It is also argued that the Presumptive is evidential, since it marks conclusions based on either past experiences or previous knowledge.

Keywords: Evidentiality, Mirativity, Tajik, Persian, Epistemic Modality

Alexander Nilsson, Department of Linguistics and Philology, Box 635, Uppsala University, SE-75126 Uppsala, Sweden.

© Alexander Nilsson 2022

ISSN 1100-326X
ISBN 978-91-513-1417-4
URN urn:nbn:se:uu:diva-466358 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-466358)
To the Tajiks
# Contents

List of Tables xi
List of Figures xii
Acknowledgements 13
Abbreviations 15
Notes on transliteration and glossing 17

## Part I: Introduction 21

1 Aim, sources, and outline 23
  1.1 Aim, method, and material 25
  1.2 Primary sources 26
  1.3 Outline of thesis 28

## Part II: Historical and linguistic background 29

2 Historical background 31
  2.1 Transoxania 31
  2.2 Alexander the Great and the Arab conquest 32
  2.3 The Samanids 33
  2.4 Mongol expansions 35
  2.5 The Timurids and the Shaybanids 35
  2.6 The Bukhara Emirate and the Russian Conquest 36
  2.7 Educational systems and Jadidism 38
  2.8 National-territorial delimitations 40
  2.9 Tajik-Turkic identities and relations 41
  2.10 Language policy 42

3 Evidentiality – definitions and classifications 47
  3.1 Defining evidentiality 47
    3.1.1 The notional definition of evidentiality 48
10.1 Reportative ...................................................... 176
  10.1.1 Non-uses ............................................. 177
10.2 Inference ...................................................... 178
  10.2.1 Non-uses ............................................. 178
10.3 Evaluative ...................................................... 179
  10.3.1 Non-uses ............................................. 179
10.4 Epistemic verbs and phrases ................................. 180
  10.4.1 Non-uses ............................................. 180
10.5 Summary ...................................................... 180

Part IV: Presentation and analysis of the oral sources 181

11 Recordings from 2016 ........................................... 183
  11.1 Contemporary events and people ................. 183
  11.2 Past events and people ............................... 185
    11.2.1 Soviet times and civil war .............. 185
    11.2.2 Notable people of the 20th century .... 186
    11.2.3 Ancient history ................................ 189
  11.3 Narrative jokes and mirative expressions ............ 192
  11.4 Summary .................................................. 193

12 Recordings from 2017 ........................................... 194
  12.1 General findings ....................................... 194
  12.2 The resultative Perfect .............................. 196
  12.3 The inferential Perfect .............................. 198
  12.4 The inferential Perfectoid Imperfect ............. 201
  12.5 Hypothetical narratives in the Perfect ........... 201
  12.6 Hypothetical questions ............................. 206
  12.7 Gnomic statements .................................... 208
  12.8 Perfect after epistemic verbs ..................... 208
  12.9 Indirectives in the first person version of the story 208
  12.10 Summary ................................................ 210

Part V: Concluding discussion 213

13 Literary sources .............................................. 215
14 Oral sources .................................................. 218
15 General conclusions—revising the Tajik verb system 220
  15.1 The Presumptive ........................................ 222
  15.2 The evaluative and the mirative .................... 223
  15.3 Evidential categories in Tajik ....................... 223
List of Tables

2 Cyrillic transliteration 18
3 Transliteration of Arabic and Persian 19
3.1 Aikhenvald’s three categories of evidential status 51
4.1 The two registers of the past tense in Persian 66
4.2 Kasravi’s analysis of the “unseen” forms 75
4.3 The two registers of Persian past tenses 79
4.4 Indirective forms in Tajik grammars and course books 85
4.5 Indirective markers in Turkic 101
4.6 Features of Past Tense Forms in Uzbek 103
4.7 Tajik and Uzbek verb forms 106
4.8 Texts used in part II 113
10.1 Recording sessions 181
13.1 Statistical summary of the texts in part II 215
List of Figures

2.1 Tajik dialects ................................................. 46
3.1 Epistemicity .................................................. 54
3.2 The two dimensions of modality ............................. 55
3.3 Epistemic modality and evidentiality ....................... 56
3.4 Evidential categories ........................................ 60
4.1 The evolution of New Persian ............................... 70
10.1 Picture 1, titled Homecoming ............................... 182
12.1 Picture 5, titled Court room ............................... 196
12.2 Picture 9, titled Arrest ..................................... 197
12.3 Picture 10, titled Thinking of jail ......................... 198
12.4 Picture 16, titled Thinking of home ....................... 200
12.5 Picture 2, titled Receiving clothes ....................... 202
12.6 Picture 4, titled Drunken gossip .......................... 203
12.7 Picture 12, titled Hitting ................................... 205
15.1 Evidentiality in Tajik ..........................
Acknowledgements

First of all, let me express my deepest gratitude to my main supervisor, Carina Jahani, for her encouragement, patience, knowledge, and wisdom. I feel immensely blessed to have had the privilege of receiving your guidance and being your student. Thank you for having kept my best interest at heart, and for taking time from your busy schedule to read and comment on my manuscript and encourage me when I needed it the most. I am also infinitely grateful to my assistant supervisor, Kasper Boye, without whom I never would have made sense of much of the theory surrounding evidentiality. Thank you for always lifting my spirits, encouraging me, and explaining complex matters with such clarity and ease.

At an early stage of my PhD, I benefited greatly from the guidance and advice of Henrik Bergqvist, who introduced me to the Family Problems Picture Task, a tool for eliciting evidential statements. Thank you for setting me on my path. I also wish to thank Bernhard Wälchli for conscientiously reading and commenting on my dissertation in conjunction with my mock defence. I am very grateful for your valuable comments and constructive criticism.

Naturally, this study would have been impossible without the assistance of many helpful Tajiks who generously gave of their time and energy to help me conduct my research on their language. First of all, I am deeply indebted to my muallima Safargul Halimiyon. Thank you for teaching me the ins and outs of Tajik and for opening my eyes to Ayni’s greatness. Thank you also for pulling strings with the university so that I could meet the students who participated in this study. Needless to say, being able to record native speakers has been crucial for my work, and I warmly thank everyone who willingly gave of their time to enable me to do my research.

I have been very fortunate to be a part of the Department of Linguistics and Philology, with inspiring colleagues who have enriched my personal and academic life. Thank you Forogh Hashabeiky for your insight, support, guidance and discussions. Éva Á. Csató Johanson, thank you for showing interest in my work and sending me academic articles of interest. Bircel Karakoç, thank you for your support and valuable comments on an early draft of my chapter on evidentiality in Turkic languages. László Károly, thank you for helping me out whenever I had problems typesetting the dissertation with \LaTeX. This dissertation would never have seen the light of day if it was not for Jaroslava Obrtelová, who not only supported me and cheered me on, but also went out of her way to supply me with contemporary Tajik fiction texts that would prove to be indispensable for the completion of my study.
Mohsen Emami, many thanks for your laughs, support and being a great friend through thick and thin. The last two years of my PhD would have been a great deal bleaker and more dull if you had not become my next-door colleague and partner in crime. Padideh Pakpour, thank you for ensuring that I inherited your office after you finished your PhD. Your office has served me well, providing me with a safe haven without which my writing would have been severely hampered. Thank you also for your constant support and friendship. Maryam Nourzaei, thank you for your encouraging words. Esmat Esmailie, there are still so many things I wish you could have taught me; I thought we had more time. Thank you for being such a kind friend. I will forever cherish the times we read the Masnavi together. I dedicate this work to your memory.

Many of the sources referenced in this work are in Russian and French. I owe many thanks to Neige Rochant for helping me find the best English equivalents to many key grammatical terms mentioned in these languages. The text of this dissertation has benefited greatly from Everett Thiele’s many corrective suggestions. Thank you for reading my text with such diligence and also answering my many questions regarding English. My knowledge of Turkish and Turkic languages has been greatly enriched by a small group of friends who all share a great interest in everything Turkic. Thank you dear Renata, Memduh, Oğuzhan, and Saimo.

Finally, I’d like to extend my gratitude to my dear parents Mazhdieh and Tommy, for all the sacrifices that you have made for me throughout the years, and to my brothers Sebastian and Adrian for all your support.

Uppsala, February 2022
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( )</td>
<td>implied information not explicitly stated in the original text</td>
</tr>
<tr>
<td>[ ]</td>
<td>obligatory additions to make sense of the translation</td>
</tr>
<tr>
<td>[...]</td>
<td>intentionally omitted part</td>
</tr>
<tr>
<td>-</td>
<td>affix boundary</td>
</tr>
<tr>
<td>1</td>
<td>first person</td>
</tr>
<tr>
<td>2</td>
<td>second person</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
</tr>
<tr>
<td>=</td>
<td>clitic boundary</td>
</tr>
<tr>
<td>Ø</td>
<td>zero morpheme</td>
</tr>
<tr>
<td>ABL</td>
<td>ablative</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>AC</td>
<td>auxiliary clitic</td>
</tr>
<tr>
<td>AOR</td>
<td>aorist</td>
</tr>
<tr>
<td>CLM</td>
<td>clause linkage marker</td>
</tr>
<tr>
<td>COMP</td>
<td>comparative degree</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>CPST</td>
<td>converbial past</td>
</tr>
<tr>
<td>CVB</td>
<td>converb</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DEM.DIST</td>
<td>distal demonstrative (that)</td>
</tr>
<tr>
<td>DEM.PROX</td>
<td>proximal demonstrative (this)</td>
</tr>
<tr>
<td>DEM.PRT</td>
<td>demonstrative particle</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>DO</td>
<td>direct object marker</td>
</tr>
<tr>
<td>EMPH</td>
<td>emphatic particle</td>
</tr>
<tr>
<td>EZ</td>
<td>ezafe (linking particle)</td>
</tr>
<tr>
<td>FP</td>
<td>future participle</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>GGT</td>
<td>General-Governorship of Turkestan</td>
</tr>
<tr>
<td>IC</td>
<td>indirective copula</td>
</tr>
<tr>
<td>IMPF</td>
<td>imperfective</td>
</tr>
<tr>
<td>IMPV</td>
<td>imperative</td>
</tr>
<tr>
<td>IND</td>
<td>indefinite</td>
</tr>
<tr>
<td>INFR</td>
<td>inferred</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INTRA</td>
<td>intraterminal aspect</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>NF</td>
<td>non-feminine</td>
</tr>
<tr>
<td>NONVIS</td>
<td>= non-visual</td>
</tr>
<tr>
<td>NPST</td>
<td>= non-past</td>
</tr>
<tr>
<td>OPT</td>
<td>= optative</td>
</tr>
<tr>
<td>PC</td>
<td>= pronominal clitic</td>
</tr>
<tr>
<td>PL</td>
<td>= plural</td>
</tr>
<tr>
<td>PP</td>
<td>= perfect participle</td>
</tr>
<tr>
<td>PP₂</td>
<td>= perfect participle II (colloquial)</td>
</tr>
<tr>
<td>PRESU</td>
<td>= presumptive</td>
</tr>
<tr>
<td>PREV</td>
<td>= preverb</td>
</tr>
<tr>
<td>PRF</td>
<td>= perfect</td>
</tr>
<tr>
<td>PROG</td>
<td>= progressive</td>
</tr>
<tr>
<td>PRSP</td>
<td>= present participle</td>
</tr>
<tr>
<td>PST</td>
<td>= past</td>
</tr>
<tr>
<td>Q</td>
<td>= question particle</td>
</tr>
<tr>
<td>RA</td>
<td>= Rawżat al-aḥbāb by Daštaki Širāzi (1484).</td>
</tr>
<tr>
<td>REC.P</td>
<td>= recent past</td>
</tr>
<tr>
<td>REP</td>
<td>= reported</td>
</tr>
<tr>
<td>SG</td>
<td>= singular</td>
</tr>
<tr>
<td>SUBJ</td>
<td>= subjunctive</td>
</tr>
<tr>
<td>SUP</td>
<td>= superlative degree</td>
</tr>
<tr>
<td>SX</td>
<td>= Selsela-nāma-ye x’ajagān-e naqšband by Nur al-Din Qazvini (1570).</td>
</tr>
<tr>
<td>TJG</td>
<td>= Tārix-e jahān-gošā by ʿAlāʾ al-Din ʿAtā-Malek Jovayni (1260).</td>
</tr>
<tr>
<td>TOP.NON.A/S</td>
<td>= topical non-subject case</td>
</tr>
<tr>
<td>TT</td>
<td>= Taḏkirat al-tavārix by ʿAbdallah Kābuli (1582–1601).</td>
</tr>
<tr>
<td>VOC</td>
<td>= vocative</td>
</tr>
</tbody>
</table>
Notes on transliteration and glossing

To facilitate the reading of the examples, the Perfect, Perfectoids, Presumptive, and any perfect participles used in serial verb constructions with these verb forms are in **boldface**. In cases where there is a lexical marker of evidentiality, such as gūyo ‘it is said’, the lexical marker is in *italics*. Brackets enclosing an ellipsis ([...]) mark that the quote has been truncated. The literal translation of the Perfect is added within parentheses to further emphasise the Tajik use of the Perfect when English favours the Past Tense.

Table 2 describes how Tajik, Uzbek, and Russian are transliterated. The system for transcribing Uzbek text in Cyrillic essentially converts it to the official Latin-based Uzbek writing system (cf. Bodrogligeti 2003: 14–15).

Table 3 describes the system for transcribing Arabic and Iranian, Afghan, and Classical Persian. Essentially, one system is used for Iranian Persian, while Arabic and Afghan and Classical Persian share the same system by virtue of having similar pronunciations of certain phonemes, such as /a/ for the word-final /e/ in Iranian Persian (xāna vs xāne ‘house’). Note, however, that this is a simplification, as some words in Afghan Persian also end in /e/, for instance *ke* (‘that, which’). The letter vāv occurring after alef is raised in transliteration: xᵛāhar. Spoken Tajik is transliterated as it is spoken. By this follows that the sound /ū/ in rūz ‘day’ is only transcribed with the letter *ū* if the speakers actually pronounces it like that. In the Dushanbe dialect, the most common pronunciation is /u/.

When other languages, or dialects of the above-mentioned languages, are cited, the original transcription system is used. Foreign names and terms that are already familiar to a Western audience, such as Djingiz Khan and *ulama*, are not transcribed according to table 3, but retain their common spellings. The names of authors and works from the Classical and Post-Classical eras will be transliterated according to the standards for Classical Persian.
### Table 2. Cyrillic transiliteration.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Аа</td>
<td>A a</td>
<td>Аа</td>
<td>A a</td>
<td>Аа</td>
<td>A a</td>
</tr>
<tr>
<td>Бб</td>
<td>B b</td>
<td>Бб</td>
<td>B b</td>
<td>Бб</td>
<td>B b</td>
</tr>
<tr>
<td>Вв</td>
<td>В в</td>
<td>В v</td>
<td>В v</td>
<td>В в</td>
<td>В v</td>
</tr>
<tr>
<td>Гг</td>
<td>Г g</td>
<td>Г г</td>
<td>G g</td>
<td>Г г</td>
<td>G g</td>
</tr>
<tr>
<td>Фф</td>
<td>Ф f</td>
<td>Ф f</td>
<td>F f</td>
<td>Ф ф</td>
<td>F f</td>
</tr>
<tr>
<td>Дд</td>
<td>Д д</td>
<td>Д d</td>
<td>Д d</td>
<td>Д d</td>
<td>D d</td>
</tr>
<tr>
<td>Ее</td>
<td>Е е</td>
<td>Е e</td>
<td>Е e</td>
<td>Е е</td>
<td>E e</td>
</tr>
<tr>
<td>Ёё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
</tr>
<tr>
<td>Жж</td>
<td>Ж ж</td>
<td>Ж j</td>
<td>Ж ж</td>
<td>Ж ж</td>
<td>Ж ж</td>
</tr>
<tr>
<td>Зз</td>
<td>З з</td>
<td>З z</td>
<td>З z</td>
<td>З з</td>
<td>Z z</td>
</tr>
<tr>
<td>Ии</td>
<td>И и</td>
<td>И i</td>
<td>И i</td>
<td>И и</td>
<td>I i</td>
</tr>
<tr>
<td>Йй</td>
<td>Й й</td>
<td>Й й</td>
<td>Й й</td>
<td>Й й</td>
<td>Й й</td>
</tr>
<tr>
<td>Кк</td>
<td>К к</td>
<td>К k</td>
<td>К k</td>
<td>К к</td>
<td>K k</td>
</tr>
<tr>
<td>Кк</td>
<td>К к</td>
<td>К k</td>
<td>К k</td>
<td>К к</td>
<td>K k</td>
</tr>
<tr>
<td>Лл</td>
<td>Л л</td>
<td>Л l</td>
<td>Л l</td>
<td>Л л</td>
<td>L l</td>
</tr>
<tr>
<td>Мм</td>
<td>М м</td>
<td>М m</td>
<td>М m</td>
<td>М м</td>
<td>M m</td>
</tr>
<tr>
<td>Нн</td>
<td>Н н</td>
<td>Н n</td>
<td>Н н</td>
<td>Н н</td>
<td>N n</td>
</tr>
<tr>
<td>Оо</td>
<td>О о</td>
<td>О o</td>
<td>О о</td>
<td>О о</td>
<td>O o</td>
</tr>
<tr>
<td>Пп</td>
<td>П п</td>
<td>П p</td>
<td>П p</td>
<td>П п</td>
<td>P p</td>
</tr>
<tr>
<td>Рр</td>
<td>Р р</td>
<td>Р r</td>
<td>Р р</td>
<td>Р р</td>
<td>R r</td>
</tr>
<tr>
<td>Сс</td>
<td>С с</td>
<td>С s</td>
<td>С с</td>
<td>С с</td>
<td>S s</td>
</tr>
<tr>
<td>Тт</td>
<td>Т т</td>
<td>Т t</td>
<td>Т т</td>
<td>Т т</td>
<td>T t</td>
</tr>
<tr>
<td>Уу</td>
<td>У у</td>
<td>У u</td>
<td>У у</td>
<td>У у</td>
<td>U u</td>
</tr>
<tr>
<td>Ыы</td>
<td>Ы Ы</td>
<td>Ы Ы</td>
<td>Ы Ы</td>
<td>Ы Ы</td>
<td>Ы Ы</td>
</tr>
<tr>
<td>Фф</td>
<td>Ф ф</td>
<td>Ф f</td>
<td>Ф ф</td>
<td>Ф ф</td>
<td>F f</td>
</tr>
<tr>
<td>Хх</td>
<td>Х х</td>
<td>Х x</td>
<td>Х х</td>
<td>Х х</td>
<td>X x</td>
</tr>
<tr>
<td>Хх</td>
<td>Х х</td>
<td>Х x</td>
<td>Х х</td>
<td>Х х</td>
<td>X x</td>
</tr>
<tr>
<td>Чч</td>
<td>Ч ч</td>
<td>Ч ch</td>
<td>Ч ch</td>
<td>Ч ч</td>
<td>Ч ч</td>
</tr>
<tr>
<td>Щщ</td>
<td>Щ щ</td>
<td>Щ ш</td>
<td>Щ ш</td>
<td>Щ щ</td>
<td>Щ щ</td>
</tr>
<tr>
<td>Ёё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
<td>Ё Ё</td>
</tr>
<tr>
<td>Ээ</td>
<td>Э э</td>
<td>Э e</td>
<td>Э э</td>
<td>Э э</td>
<td>E э</td>
</tr>
</tbody>
</table>
Table 3. Transliteration of Arabic and Persian (Classical, Afghan, and Iranian). Letters appearing after the slash (/) are possible Classical and Afghan Persian variants (the so-called majhul vowels), whereas those appearing after the vertical bar (|) are equivalents.

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Transliteration</th>
<th>Vowels</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>ء</td>
<td>'</td>
<td>آ،ـَا</td>
<td>ā</td>
</tr>
<tr>
<td>ب</td>
<td>b</td>
<td>؛ُو</td>
<td>u / ō</td>
</tr>
<tr>
<td>پ</td>
<td>p</td>
<td>یِ</td>
<td>i / ē</td>
</tr>
<tr>
<td>ت</td>
<td>t</td>
<td>-َ</td>
<td>a</td>
</tr>
<tr>
<td>ت</td>
<td>t</td>
<td>-ُ</td>
<td>o</td>
</tr>
<tr>
<td>ج</td>
<td>j</td>
<td>-ِ</td>
<td>e</td>
</tr>
<tr>
<td>چ</td>
<td>ğ</td>
<td>؛،ه</td>
<td>e</td>
</tr>
<tr>
<td>خ</td>
<td>h</td>
<td>؛و</td>
<td>ow</td>
</tr>
<tr>
<td>د</td>
<td>d</td>
<td>؛ی</td>
<td>ey</td>
</tr>
<tr>
<td>ذ</td>
<td>ḍ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ر</td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ز</td>
<td>ž</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ز</td>
<td>ž</td>
<td></td>
<td></td>
</tr>
<tr>
<td>س</td>
<td>s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ژ</td>
<td>ḍ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ژ</td>
<td>ž</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ژ</td>
<td>ž</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ژ</td>
<td>ž</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td></td>
<td></td>
</tr>
<tr>
<td>q</td>
<td>q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ن</td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ی</td>
<td>y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Part I: Introduction
1. Aim, sources, and outline

Annem beni ilk gördüğünde benden iki yaş büyük ağabeyime göre daha zayıf, daha kırılgan ve daha ince olduğumu düşündü. Aşında “düşünmüş” demeliydim. Türkçede rüyaları, masalları ve doğrudan yaşamadığımız şeyler anlatırken kullandığımız ve çok sevdiğimmiş’li geçmiş zaman beşikteyken, tekerlekli çocuk arabasındayken ya da ilk defa yürürken yaşadıklarımızı anlatmak için daha uygundur.

When my mother set eyes on me for the first time, she thought (düşündü) I looked weaker, more frail, and thinner than my brother, who is two years older than me.

In fact, I should have said düşünmüş (‘I have been told that she thought’). In Turkish, when conveying dreams, fairy tales and things that we have not experienced directly, it is more appropriate to employ the frequently used, and much loved by me, past tense withmiş when talking about the time when we were in the cradle, the baby carriage, or taking our first steps.

Orhan Pamuk (my translation), Resimli İstanbul – Hatıralar ve Şehir (2015: 48)

In many languages of the world, the source of a particular piece of information is grammatically marked on the verb. This phenomenon—the grammatical marking of information source—is commonly called evidentiality. It is known from typological studies that evidentiality is very easily diffused to other languages through intensive language contact (Aikhenvald 2004: 271, 288). This explains why we find large linguistic areas (Sprachbunds) displaying evidentiality despite the fact that the languages in these Sprachbunds do not share a close genetic relationship.

One such area, which is of particular importance for this dissertation, is the “Great Evidential Belt”, consisting of languages stretching from the Balkans to Central Asia, and further to Siberia (Plungian 2010: 19). What the languages in this linguistic area all have in common is that they are endowed with the ability to grammatically mark information which has not been experienced directly by the speaker, a phenomenon which Haarmann (1970) refers to as a Eurasian isogloss. Typically, the grammatical vehicle for expressing such indirectiveness in these languages is the perfect tense.

In this vast stretch of land, which straddles the continents of Europe and Asia, we find languages that, notwithstanding their genetic diversity, still display striking similarities in their grammatical display of information source.
Indo-European, Turkic, and Kartvelian languages in this area all display a type of evidentiality which traditionally has been described as simple or binary, due to its having one type of evidential marker for first-hand and another one for non-first hand information (Plungian 2010: 42).

Example (1) is from Tajik, a variety of Persian\(^1\) spoken mainly in Tajikistan and Uzbekistan. In (1a) the first-hand information is conveyed with the Imperfect, whereas the non-first hand information is conveyed with the Perfect-oid Imperfect in (1b). The marking of non-first hand information entails that the information was obtained either through reports from others or through inference.

(1)

1. **me-raft-∅.**  
   IMPF-GO.PST-3SG  
   He used to go.  
   \textit{first-hand information}

2. **me-raft-a=ast.**  
   IMPF-GO.PST-PP=COP.NPST.3SG  
   He used to go.  
   \textit{non-first hand information}

As we can see from the autobiographical quote by Pamuk which introduces this chapter, the use (and the non-use) of evidentiality can be vexing not only for the linguist but also for the everyday speaker of the language. As the author himself explains, there is in Turkish a special verb form for “dreams, fairy tales and non-first hand experiences”, which is contrasted with another verb form used for information that one has acquired first hand.\(^2\) When retelling how his mother perceived him on the day of his birth, the author, acutely aware of the prescriptive grammar of Turkish, feels compelled to correct his former utterance which wrongly had conveyed that the information was acquired first-hand. Having no recollection of his first years in life, he has to rely on information given to him by his parents which, when retold by himself, should be marked with the proper verbal inflectional suffix \(-\text{mIş}\), which in this context signals that the relayed information is not based on the author’s own experiences.

Why then did Pamuk initially tell the story as if he had first-hand knowledge of his mother’s thoughts upon seeing him for the first time? One might

---

\(^1\) The other varieties of Persian, i.e. Iranian and Afghan Persian, also exhibit the same kind of evidential juxtaposition. See section 4.1 for a thorough discussion of the differences between the three as regards the marking of non-first hand information.

\(^2\) As we shall see in section 4.2.2, even though this is the way the contrastive nature of the opposing verb forms is usually understood among native speakers of Turkish, it is somewhat of a simplification of how the indirective verb forms actually operate in Turkish.

\(^3\) When transcribing Turkic morphemes, a capital letter denotes that it has different possible realisations according to the rules of vowel harmony and consonant alternation (Göksel and Kerslake 2005: xxxiii). In the case of the evidential / perfective suffix \(-\text{mIş}\), the capital letter \(I\) can be realised as either \(i, ü\), or \(u\) (ibid.: 22).
argue, and rightfully so, that the author’s self-correction is merely a rhetorical technique, a pretext to muse on the fascinating topic of evidentiality in Turkish. Be that as it may, the section also serves to question the dogma of prescriptive grammar, and hints at the fact that native speakers do not always conform to prescribed rules dictating what is “more appropriate”.

1.1 Aim, method, and material

The aim of this dissertation is to describe and account for the evidential uses of the Perfect, the Perfectoids and the Presumptive in Modern Tajik as they appear in the written language of the last 100 years and in the contemporary spoken language. These two fields of Tajik linguistics constitute a research lacuna, which hopefully this dissertation can help to fill.

The main philosophical underpinning of the present dissertation is that language is mainly a means of communication which first and foremost involves a subject, a recipient, and context. In other words, the approach to language employed in this dissertation is functional—language is regarded as being primarily concerned with establishing “communicative relations between speakers and addressees…” (Dik 1978: 1).

Given these qualities of language as a means of communication, it follows that the best way of studying evidentiality and related notions is to employ an empirically based method of investigation that pays careful attention to context. The context in which a specific evidential form is used is just as interesting as when it is not used. In other words, one cannot merely search for the desired verb forms in a corpus, note their occurrence and present them in a table. Rather, one must analyse them in their contexts, which is the method that will be employed in this dissertation.

As is explained in section 4.1.2, grammars of the Tajik language often provide no more than cursory remarks on the function of the above-mentioned verb forms, and there is to date no chronological study accounting for their usage and prevalence since the founding of the Tajik state. As re-

4. Since Tajik mirative statements (traditionally seen as expressing surprise or new information) also employ register II forms, and given the close connections between expressions of mirativity and evidentiality (see section 3.2.2), mirative statements will also be analysed in this study. Moreover, pragmatic uses of the register II forms and the Presumptive (such as sarcasm, questions, gnomic statements, etc.) will also be analysed and accounted for, since it is not always clear to what extent they have ceased being evidential.

5. Descriptive categories will be referred to with an initial capital letter (e.g. “the Tajik Pluperfect”), whereas comparative concepts will be written with an initial lowercase letter (e.g. “In these languages, the perfect is often used to express indirectivity”). See Haspelmath (2010: 674) for a discussion of this usage and its history within linguistics.

6. See, for example, the reasoning of Benveniste (1971: 225) on subjectivity: “Language is marked so deeply by the expression of subjectivity that one might ask if it could still function and be called language if it were constructed otherwise.”

7. See, for example, how Jakobson (1960: 353) describes communication in language.
gards the spoken language, there has never been any study on their usage and prevalence in the contemporary spoken language of Dushanbe.

Among these verb forms, the Perfect is the most elusive as regards its evidential status. It was originally a resultative\(^8\) and still functions mainly to this effect. In other words, it is very hard to tease out the non-resultative functions of the Perfect from its resultative ones. Only in certain cases do we find a stronger prevalence of non-resultative meanings, for instance when we encounter the Perfect of *budan* ‘to do’.

The Perfectoids are two verb forms that are morphologically similar to the Perfect. In essence they are modelled on the Imperfect and the Pluperfect, but with an added Perfect-like (hence Perfectoid) ending in `-a`. Unlike the Perfect, however, these forms are unambiguously non-first hand in nature. In other words, by using them the speaker is actively signalling that the information has been retrieved either through report or inference. Verb forms which function in this way have been called indirective (Johanson 2000), mediative (Lazard 1956), and non-witnessed (Perry 2005).

The Presumptive is morphologically derived from the Perfect and is peculiar to the Tajik language in that it does not exist in the varieties of Persian spoken in Iran and Afghanistan. Given that its function is described sparingly and in conflicting ways (cf. Rzehak 1999, Ido 2005, Perry 2005), and in view of the close connections between evidentiality and epistemic modality, the Presumptive will also be included in this study in order to shed more light on the function of this verb form.

The analytical chapters found in parts III and IV of this dissertation mainly focus on the Perfect, the Perfectoids and the Presumptive. Their non-resultative uses are traditionally broken down into three distinct categories: reportatives, inferences and miratives. Apart from these more prominent categories, there are certain cases where these verb forms appear in pragmatic extensions: after epistemic verbs and phrases, and in evidential questions, gnomic statements, hypothetical questions, and sarcastic utterances.

1.2 Primary sources

In order to thoroughly examine the use of the above-mentioned verb forms, two different types of primary sources have been used. The first is literary prose in the form of Tajik novels and short stories while the second consists of recordings of university students in Dushanbe, Tajikistan. The Tajik literature considered in part III consists of six pieces of fictional prose written in three different time periods: an early (1920s–1940s), a middle (1980s), and a more recent period (2010s). The reason for choosing novels from these periods is that they represent three different epochs in the development of Mod-

\(^8\) Likewise, the perfect is often described as conveying “current relevance”; see 3.2.3 for an explanation of this term.
ern Literary Tajik. The literature from the first period represents a time in Tajik history which was focused on consolidating and nurturing a new type of Modern Tajik written in an adapted Cyrillic script and exhibiting several grammatical idiosyncrasies adopted from the Northern dialects. The second period is a sort of middle ground between the first and third periods, when society was beginning to open up. The authors of this period were all raised with literature from the first period, and the ongoing Soviet-Afghan war accentuated the fact that the Tajik variety of Persian was but one piece of a mosaic spanning multiple countries where varieties of Persian were spoken. At the same time, the opening up of society due to perestroika allowed for new cultural influences from Afghanistan and Iran (Foltz 2019:140). The authors in the third period are writing in a society that is vastly different from that of their predecessors. The emergence of Tajikistan as an independent country in the 1990s led to a diminishing influence of the Russian language, the introduction of more words from Iran and an increased production and importation of materials written in the Perso-Arabic script (Perry 2005:489; 491).

The recordings were conducted in Dushanbe, Tajikistan, in the years 2016 and 2017. The 2016 recordings basically took the form of free conversations regarding different topics. The 2017 recordings were of a more structured kind, consisting of conversations based on the Family Problems Picture Task, a type of collaborative storytelling exercise where the two participants are tasked with constructing possible storylines from a set of pictures that are given to them one at a time. The exercise was developed as a method for investigating grammatical categories related to social cognition (San Roque et al. 2012:135).

In their entirety, the pictures can make up multiple storylines, but they are shown in a non-chronological sequence in order to elicit expressions of supposition, doubt and inference. At the end of the exercise, the participants are asked to settle on a final storyline by rearranging the 16 picture cards. They are then asked to retell it in the third person, through the eyes of an objective onlooker and again in the first person, through the eyes of one of the characters in the story (ibid.:135).

Ideally, the session should be video-recorded, and in the last part of the exercise, when the participants narrate their stories, a new person who has not participated in the previous part of the exercise should be introduced in order to partake of their story.

For practical reasons, I did not fully adhere to these recommendations, even though they may have been conducive to the eliciting of further evidential statements. Video-recording the participants adds an extra element of deterrence for people who might already be hesitant about participating in such an exercise. Moreover, I did not intend to analyse their body language, which is why I settled for only recording their audio. Introducing a new per-
son for the last part would have been ideal, but for practical reasons and due to the scarcity of volunteers, this was not possible.

The participants in these recordings were all university students between the ages of 18 and 23. All of them were from Dushanbe or its surrounding villages.

1.3 Outline of thesis

The thesis will be structured as follows: Part II, consisting of chapters 2, 3 and 4, will present the reader with the necessary background information on the history of Tajikistan, Tajik and the academic study of evidentiality.

Parts III and IV display the different ways that evidentiality is manifested in written and spoken Tajik, respectively. Part III consists of six chapters, each presenting an analysis of a fictional prose text as regards the use of evidential forms. Ample glossed examples are given in order to further illustrate the interplay of register I and II forms. Part III concludes with a summary of the findings from the six texts. Part IV consists of two chapters, each presenting findings of evidentiality from sessions with native speakers recorded in 2016 and 2017. A summary of the main results of the summary is given at the end of each chapter.

In part V, a summary of the findings of this thesis is presented. A revised, more accurate presentation of the Tajik verbal system is also proposed, in which the traditional description of the function of the Presumptive, and the mirative functions of the Perfect and the Perfectoids, are cast in a new light.
Part II: 
Historical and linguistic background

This part of the thesis consists of three chapters which aim to give some historical background regarding Transoxania and linguistic background dealing with the academic study of evidentiality, the Tajik language and Tajik’s relationship to the other national varieties of Persian and Turkic languages.

Chapter 2 will give a brief introduction to the historical background of Transoxania, i.e. the part of Central Asia that roughly corresponds to the modern states of Uzbekistan and Tajikistan. Of special importance are the political developments in the last century which have had a decisive influence in shaping Modern Literary Tajik as a national variety of Persian.

Chapter 3 will present the reader with an overview of the linguistic field of evidentiality. Its main purpose is to describe evidentiality notionally, and to distinguish it from related concepts such as epistemic modality and mirativity in order to have a secure basis from which to analyse evidentiality in Tajik.

Chapter 4 will give a historical survey of evidentiality in New Persian, starting from the earliest texts and continuing until the modern-day national varieties of Persian, with the main focus being on Tajik. Since there is a longstanding history of Persian-Turkic language contact, the main characteristics of evidentiality in Turkic languages will be presented, and common traits between Tajik and Uzbek will be explored.
2. Historical background

This chapter has two aims. The first is to provide the reader with a brief history of the stretch of land that historically was known as Transoxania and nowadays roughly constitutes the sovereign nations of Uzbekistan and Tajikistan. The second aim is to highlight the effects of these historical events, especially the more recent ones, on the type of Persian spoken in this region, ultimately resulting in the contemporary national variety of Persian known as Tajik.

The latter part of this chapter, from section 2.6 onwards, will be of special importance for understanding the emergence of Tajik as a national language. It covers political developments which have had far-reaching consequences for Tajik, causing changes to the language that have set it apart from the other national varieties of Persian, i.e. Iranian and Afghan Persian.

2.1 Transoxania

Present day Tajikistan is situated in a region of Central Asia which is demarcated by two great rivers. The southern river, known in Greek and Latin as Oxus, is the reason for the name Transoxania ‘That which lies beyond the river Oxus’. In the same vein, the Arabs called this region Mā warā’ al-nahr ‘That which is beyond the river’. In Middle Persian, the Oxus was named Wehrōd ‘good river’, and was seen as the boundary between the ancient homelands of the Iranians, Īrān, and that of the Turks, Tūrān (Spuler 1989).

Nowadays, the Oxus goes by the name of Amu Darya ‘The Āmu River’, the main tributary of which is the Panj River which has its source in the mountainous region of Badakhshan in the eastern part of the country. From there it flows westward forming a natural southern border to present-day Afghanistan. The northern demarcation of Transoxania is the river Jaxartes (nowadays known as Syr Darya), which has its source in Tianshan in present-day Kyrgyzstan, and flows past Andijan and Khujand on its way to the Aral Sea.

The earliest source for a naming of Transoxania is found in the opening paragraphs of the Vēndīdād of the Zoroastrian holy text Avesta, where it is called Gava and is described as a land inhabited by suγδa ‘the Sogdians’, who were an Iranian people. According to this text, Ahura Mazda created Gava
as the second of sixteen good lands, the first one being Aryana Vaēja, the semi-mythical homeland of the Aryans\(^1\) (Grenet 2005:30).

The Sogdians had already inhabited Sogdiana, a part of Transoxania centred around Samarkand, for centuries when in about 540 BCE Cyrus the Great conquered it and made it a part of the Persian Empire as a “distant frontier province” (La Vaissière 2011). The Sogdians, from which Sogdiana gets its name, have been attested in many languages of Antiquity and the Middle Ages. Among these languages we find Avestan, Old Persian, Akkadian, and Latin. To this day, the northernmost province of Tajikistan is still referred to as Suğd ‘Sogdiana’ (Lurje 2017).

2.2 Alexander the Great and the Arab conquest

After waging war against the Achaemenid Empire and defeating King Darius III in 331 BCE, Alexander the Great conquered Transoxania in 329 BCE. Following Alexander’s death in 323 BCE, his generals governed the region until 247 BCE after which the Greco-Bactrian Kingdom came into existence. According to some estimates, the Greco-Bactrian Kingdom held sway over the region until approximately 140–130 BCE (La Vaissière 2011).

Alexander the Great’s conquest of this area played a crucial role in ushering in a new cultural paradigm which saw the introduction of Western (Hellenistic) culture and its merging with local cultural expressions. The result was a cultural synthesis and a radical transformation of the socio-political and economic landscape of the region (Jonson 2006:24). The five centuries following the end of Greek rule in the region are very obscure. In the 6th century, Chinese sources point to a people they call Tujue (T’u-Chüe), in Southern Siberia and Mongolia. This people, whom modern scholars have identified as Turks, would later come to play a pivotal role in the history of Transoxania (Findley 2005:21).

By the time of the Arab conquest of Sasanid Iran in the middle of the 7th century, Transoxania had been divided into 27 petty princedoms (Nourzhanov and Bleuer 2013:12). This mosaic of governing entities paved the way for a piecemeal conquest of Transoxania, beginning in 705 with the campaigns of the commander Qutayba ibn Muslim and culminating in 751 with the victory over the Chinese at the battle of Talas.

The Arab conquest of Sasanid Iran in the middle of the 7th century resulted in the adoption of a new religion, Islam, and also the arrival of Sasanian

\(^1\) The exact location of Aryana Vaēja was long a contested issue. By late Sasanian times, Aryana Vaēja was believed to be situated in the western part of Iran, but since the beginning of the 20th century, scholars have found compelling reasons to locate it to Chorasmia in the eastern part of Greater Iran (MacKenzie 1989). Gava, however, and six more of the sixteen lands, are more readily identified geographically since they have either kept the same names throughout the millennia or have been associated with ethnic groups which give us a clue about their location (Grenet 2005:30).
aristocrats who fled from the Iranian heartland to Transoxania during the Ummayad Caliphate (661–750). This led to an eastward spread of Persian language and culture, all the way to the frontiers of China (Frye 1975: 96). By the time the conquests reached Central Asia in the early 8ᵗʰ century, however, it is perhaps more correct to label them “Arabo-Islamic” rather than “Arab” since, by this time, converted Persians were part of the armies (Morony 2012).

These waves of Persians entering Transoxania, both in the form of refugees and as Muslim converts with ties to the military campaigns, initiated a language shift from the vernacular Sogdian to Persian. By virtue of being part of the Sasanian empire, and having a Sasanian presence in the form of administration and garrisons, the areas south of Transoxania, the historical province of Khorasan, including western Afghanistan, had already seen the abandonment of the local Iranian language Parthian in favour of Middle Persian (Frye 1975: 52; 100).

2.3 The Samanids

The greatest boost to New Persian’s status as a written and spoken language came with the formation of the Samanid dynasty (819–1005). After the setbacks of the Arab invasion and the fall of the Sasanid dynasty, Persian culture and literature experienced a renaissance at the Samanid court in Bukhara. The Samanid rulers favoured New Persian as a literary language, attracting poets and scholars alike. The ulama², who still adamantly stressed the importance of Arabic on religious grounds, soon also saw the benefits of using Persian for reaching the masses (ibid.: 203).

Historical and scientific texts in Arabic were translated into New Persian, elevating the status of the language by making it a means of conveying scientific and historical knowledge. Drawing on ancient Iranian mythology and legends, the poet Ferdawsi wrote the epic Šāhnāma in this new language. The Samanid rule in Transoxania and Khorasan thus played a pivotal role in forming and strengthening a new identity that was simultaneously Iranian and Muslim (Jonson 2006: 29). The Oxus river divided the Samanid state into two parts, the southern of which, Khurasan, had the city Nishapur as its provincial capital. But it was the northern part, Transoxania, with Bukhara as its capital, that was the de facto administrative centre for the dynasty (Frye 1975: 203).

Towards the end of the 10ᵗʰ century, the Samanid Empire fell into turmoil and confusion and became an easy prey for the Qarakhanid Turks and the Ghaznavids, who divided the spoils of the Samanid empire amongst themselves. The Qarakhanids reigned over what was once the northern part of

² In Islam, the term ‘olamā’ ‘men of knowledge’ refers to those who have received an education in the Qur’an, hadith (traditions regarding the sayings of the prophet Muhammad), fiqh (Islamic jurisprudence), and other religious sciences (Gilliot 2000: 801).
the Samanid empire, while the Ghaznavids ruled over the southern part. At its peak, the Ghaznavid Empire covered an area stretching from Central Asia and Khurasan all the way to India (Findley 2005: 67).

For their armies, the Samanids had relied heavily on Turkic slave-soldiers called ġelmān3 and had implemented a system of slave-training to this end. The slave trade was also a very lucrative business and enabled the Samanid state to amass great wealth, which was a contributing factor in financing the renaissance of Persian as a literary language (Golden 2006: 27).

Ironically, the institution of slavery4, which had filled the treasuries of the Samanids with gold and their armies with skilled soldiers, ultimately became their downfall, since it was rebellious slave soldiers who had received their training from the Samanids that founded the Ghaznavid Empire.

For the nine centuries that followed Samanid rule, and until the Russian encroachment in the middle of the 19th century, Turkic rulers reigned over Central Asia, and great waves of Turkic tribes entered the region. This influx of Turkic tribes had begun, on a smaller scale, several centuries before, when migrating Turkic tribes supplanted or absorbed Iranian peoples in Central Asia. However, the promotion of the Persian language by the Samanid rulers resulted in Persian being a prestige language in an area covering Iran, Central Asia and, until the British encroachments, Northern India (Findley 2005: 50).

Just as the Samanids had supported Persian and Iranian culture, which resulted in a Persian Islamic literature, the Qarakhanids, a Turkic dynasty which ruled Central Asia from the 9th to the early 13th centuries (Bosworth 1986: 1113), inspired the formation of a new Turkic literature, and thereby a new Turkic-Islamic civilisation (Lapidus 2012: 230). Before them, the Seljuks, hailing from a minor ruling clan in Transoxania, had converted to Islam and migrated to Iran where they established the Seljuk empire, thereby forming the first of many Turko-Islamic empires in the Middle East.

The Seljuks were, just like the Qarakhanids, highly Persianised, but unlike the Qarakhanids, they never established Turkic as a literary language. Maḥmud al-Kāšģari’s Divān Loğāt al-Tork from 1077 CE is probably the most important work from the Qarakhanid era. Though written in Arabic, it is a pioneering work of Turkic lexicography and is of special importance to the study of evidentiality, since it is the first work to describe the Turkic evidential system (Findley 2005: 77).

3. This is the Arabic plural of the word golām, literally ‘boy, servant’, but denotes ‘military slave’ in this context.
4. The use of slaves in the military was not a specifically Samanid, or Islamic, practice. A predecessor to this type of military organisation can be seen in the čākar system of the Sogdian kings. Before the Samanids, the Abbasids and the Taherids also relied on Turkic slaves for their militaries (Golden 2006: 21).
2.4 Mongol expansions

In the 13ᵗʰ century, the Mongol expansions under Djingiz Khan also led to an increased influx of Turkic tribes into Central Asia. Mongols and Turks had a long history of mutual contact even before the Mongol expansion. This mutual contact is also reflected in their languages, which share many common features. These similarities have led some scholars to believe that there exists a genetic relationship between the two language families (Knüppel 2010).

After Djingiz Khan’s death in 1227, the Mongol’s unprecedented expansion was continued by his sons and grandsons. By the beginning of the 1230s, Transoxania and Iran were entirely under Mongolian rule. Reaching Baghdad in 1258, the Mongols ultimately dealt the death blow to the Abbasid caliphate, whose power and influence had already been waning. In 1279, when the Eastward expansion finally enveloped the whole of China, the Mongol Empire became the greatest empire the world had ever seen (Nordberg 1988: 43).

However, the Mongol Empire was not a fully centralised and unified enterprise. According to Turko-Mongolic customs, the conquered domains were seen as belonging to all of the members of the ruling family. The empire was thus divided among them. Since 1256, the grandson of Djingiz Khan, Qubilai Khan, was the *khāqān* ‘supreme ruler’ of Mongolia and China. His younger brother ruled over Iran and Iraq, an area that was called the Il-khanate, signifying that the ruler of this domain was only the *il-khān* ‘Khān of the people’; i.e. he was subordinated to the *khāqān*. (Findley 2005: 79).

2.5 The Timurids and the Shaybanids

The Turkic influence on Transoxania reached its apex with the creation of the Timurid Empire (1370–1507), which had Samarkand as its centre. The Timurids encouraged the Persian cultural heritage, but the Turkic language Chaghatay was gradually elevated as a literary language alongside Persian, even though the latter continued to hold sway. Chaghatay is hard to define as its meaning has varied throughout the centuries. One way of viewing it is as a high literary Turkic language which reached its classical form in the 15ᵗʰ century in the works of the Timurid poet ṢAli-Šēr Navā’ī (Boeschoten and Vandamme 1998: 166–167).

The 16ᵗʰ century saw the end of the Timurids. The Shaybanids (1500–1599), nomadic Uzbeks from the Kipchak steppe who were the descendants of Shayban, a grandson of Djingiz Khan, now controlled the greater part of Central Asia. From the beginning of Shaybanid rule until the middle of the 18ᵗʰ century, Central Asia experienced an economic, cultural and political decline (Becker 2004: 4).

---

5. I.e. Turkic-speaking peoples.
One important factor in the gradual decline of Central Asia was the discontinuation of the trade routes across Transoxania and Xinjiang in favour of a new sea route which was opened by the Europeans (Jonson 2006: 31). Another important reason for the decline of Central Asia is the emergence of the Safavid Empire (1501–1736) in Iran. Their continued efforts to spread Shi’ism throughout their domains created a rift between the Sunni Muslims of Anatolia and of Central Asia, since the Safavids blocked the flow of merchants, pilgrims and scholars between the eastern and western part of the Sunni Muslim world. Moreover, the decline due to the new Shi’a state was not only economical—the religious rift between the Persian speakers of Iran and Central Asia led to a cultural decline for Persian as a prestige language in Central Asia. In its stead, Chagatay made significant advances into territories which had previously been the domains of Persian (Spuler 1977: 468).

2.6 The Bukhara Emirate and the Russian Conquest

In the 17ᵗʰ and 18ᵗʰ centuries, three local dynasties emerged in Transoxania: the Khiva Khanate, the Bukhara Emirate, and the Kokand Khanate. The greatest of them, covering most of present-day Tajikistan, was the Bukhara Emirate, ruled by the Manghit dynasty from 1785 onward. This dynasty was the first post-Timurid dynasty not to use a genealogy tracing back to Djingiz Khan to legitimise their rule. This shift away from a Turco-Mongol tribal way of legitimising power is also seen in the change of the title of the ruler. Instead of calling themselves khans, they opted for emir, an Islamic title.

At the end of the 19ᵗʰ century, the Bukhara Emirate covered a stretch of land the size of the present day United Kingdom, but it had little geographic unity. In the western part of the Emirate, on a plain that was otherwise desert, three oases formed the basis for the demographic, cultural and political heart of the Emirate, which had little influence over its eastern mountainous territories for reasons of distance and accessibility (Becker 2004: 6). The remoteness of the eastern region gave its inhabitants a high level of autonomy (Nourzhanov and Bleuer 2013: 22).

The Khiva Khanate, on the other hand, was more compact and geographically unified. It was centred around only one oasis and its surrounding desert, making it easier for the central government to rule over the area without having to split the power with provincial rulers, as in Bukhara. Khiva also had begliks (provinces), but with rulers who enjoyed less power and authority than their Bukharan counterparts (Becker 2004: 8–11).

The Bukhara Emirate and the Khiva Khanate were similar in many ways—both were autocratic and had a heterogeneous population consisting of a majority of Sunni Uzbeks, who according to Becker (2004: 6) also “constituted
the social and political elite”.

Khiva, however, was smaller, less populous, poorer, and less urbanised than her northern neighbour Bukhara (ibid.: 6).

The despotic and conservative nature of the Bukhara Emirate made it incapable of addressing the Russian encroachments in Central Asia during the second half of the 19th century. Bukhara and Khiva are classic examples of pre-modern societies whose economic, social and political systems had not seen much change since the 10th century. Cultural exchange with other nations was almost non-existent, except for infrequent contacts with Constantinople, and even less with Iran and China. Such was the conservative nature and intellectual insularity of the rulers of Bukhara and Khiva that even the printing press was unknown in these lands before the Russian conquest (ibid.: 11).

In 1864, Russia gained Tashkent, which served both as a strategic vantage point for further encroachments in Central Asia and as a buffer against the Khiva Khanate and the Bukhara Emirate (Jonson 2006: 32). In 1866, the Bukhara Emirate was invaded by Russian troops. A year later, the General-Governorship of Turkestan (GGT) was established with Tashkent as its centre. The GGT comprised all of the territories of Bukhara and Kokand which the Russian Army had gained by that time. Throughout the 1860s and 1870s, Russia had the emir of Bukhara and the khan of Khiva sign treaties that made their territories into Russian protectorates.

The Russian conquest of Central Asia was swift. The Khanate of Kokand was formally abolished in 1876 by Tsar Alexander II. In 1884, Russia gained the whole of Turkestan when the Turkmen city of Merv finally surrendered (Nourzhanov and Bleuer 2013: 18).

Tajiks made up about nine per cent of the population of Turkestan in 1897 and were concentrated in the oblasts (administrative regions) of Samarkand and Ferghana (ibid.: 22). The Russian presence in the GGT brought about significant social and demographic changes. When the Turkologist Radlov visited Samarkand in 1868, he remarked that the people spoke Tajik almost exclusively, but by 1904, when the Orientalist Bartol’d was there, he noted an increased use of Turkic among the people (Bergne 2007: 15).

The increase in the use of Turkic can be explained by an internal drive among many Turkestanis to adopt pan-Turkic ideas. Moreover, an external push towards Turkic was brought on by the Russian administration’s penchant for employing Turkic speakers (Kazakhs and Uzbeks) as translators and members of staff (ibid.: 15–16). The policies of the Russian administration in this region point to deliberate attempts to bring about a Turkicisation, and finally, Russification of this region. An example of the latter is that in 1891, oblast governors were instructed that native administrative officials should use the Russian language in their administrative affairs, and that a good com-

6. In fact, it is very hard to say how these people identified themselves. As we shall see in section 2.9, there were many competing identities, such as Sart, Tajik, Uzbek, and Turkestani, which many times were conflated.
mand of Russian should be a merit when appointing people to future positions (Nourzhanov and Bleuer 2013: 22).

2.7 Educational systems and Jadidism

At the beginning of the 20th century, there were three types of schools in Turkestan. The traditional old schools; the Russian-type schools; and the so-called “new method” (usuli jadid) schools, favouring a combination of the old system with the Western one.

The Jadidis, i.e. the proponents of the “new method”, were Muslim modernists who operated within the framework of revolutionary or reformist movements that were common among Muslims in Russia and neighbouring countries during this period of time, i.e. the late 19th and early 20th centuries. The Jadidis were for the most part Pan-Turks, which made them the object of the tsarist government’s suspicion. In Bukhara, however, Jadidism was promoted, since the tsarist government saw it as a possible way of undermining the authority of the conservative clerics (ibid.: 23). The effects of public education efforts were rather modest in Bukhara and Russian Turkestan. By 1917, the literacy rate varied between one and two per cent, which was considerably lower than that of other parts of the region such as Iran, which according to some estimates had a literacy rate of five per cent by the beginning of the 19th century, and by 20 per cent by the 1920s (Amirahmadi 2012: 121). India, which at the time was under British rule, also fared better than Russian Turkestan in terms of literacy and educational efforts (Coates and Coates 1951: 53–54).

The Jadidi movement contributed to organising and shaping a future intelligentsia of Central Asia. One of these intellectuals was Sadriddin Ayni7 (1878–1954), who would later become the founder of Modern Literary Tajik. He was one of numerous Tajik intellectuals in Bukhara who were attracted by the new movement, despite its many pan-Turkic proponents (Hitchins 1987).

In adopting a Pan-Turkic view of Central Asia, Tajiks and Tajik history became obstacles that had to be discounted as parenthetical lapses brought about under the influence of oppressive Persian-speaking rulers in the region. For example, some of the Jadidis argued that the Tajiks were originally of Turkic stock, but had undergone a forced Persianisation on account of Persian being the administrative language in the Bukharan Emirate and Khiva Khanate. The logical conclusion, which even some Tajik intellectuals subscribed to, was that the Tajik language was a remnant of oppressive regimes and incompatible with modernity, and that the remedy was for Tajiks to “revert” to using Uzbek as their means of communication (Bergne 2007: 17–18). This message of eradicating the Tajik Persian language in fa-

---

7. Ayni’s novella Ahmadi devband (“Ahmad the Exorcist”) is analysed in chapter 5, p. 115.
vour of Uzbek met with opposition from some Soviet ethnographers and Tajik Persian-speaking intellectuals, such as Sadriddin Ayni (Brasher 2011: 108).

Bergne (2007: 125) argues that the question of individually aligning oneself with an ethnic group and identity was mostly a concern for the bilingual Tajik and Uzbek speakers of the cities, the so-called Sarts. The Tajiks inhabiting the plains of agrarian, pre-Soviet Central Asia had not yet formed an ethnic identity, let alone nationalistic sentiments. Moreover, apart from language, there was little to set Tajiks apart from Uzbeks.

Religious differences further fractured the Persian-speaking communities of the cities. According to modern estimates, the population of Bukhara was around 80,000–85,000 in the 19th century. The Shi’as, constituting approximately one tenth of the population, were mostly Eroni ‘Iranians’, descendants of the inhabitants of Merv who had been deported to Bukhara at the end of the 18th century by Shah Murād, the first emir of the Bukharan Emirate (Bregel 2003: 80). Sunni Tajiks in Samarkand and Bukhara did not feel that they had anything in common with the Tajik-speaking Jews, who numbered around 4,500 in Bukhara, or the Shi’a Eronis who lived alongside them in the cities. The Eronis, seen as both religiously divergent and ethnic outsiders, were at the bottom of the social hierarchy, making them natural targets for the Bolshevik cause. In 1919 Sayyid Rizo Alizoda (1887–1938), an Eroni Shi’a from Samarkand, started the first post-revolutionary Persian language newspaper Šū’lai Inqilob ‘Flame of the Revolution’, and he subsequently headed the Persian Section of the Communist Party (Bergne 2007: 23–24).

Šū’lai Inqilob and other Eroni endeavours that were undertaken after the revolution did not enjoy much success. Due to its low circulation, Šū’lai Inqilob had to be closed down in November 1919, only to be restarted again within the same month, this time with the help of prominent Jadidis such as Hoji Mu’in (1883–1942) and Sadriddin Ayni, who later would both become central figures in the formation of a new Tajik linguistic identity. The paper’s popularity hardly increased, despite this new initiative (ibid.: 24).

One possible explanation for the paper’s lukewarm support is that most Sunni Tajiks did not want to take part in an initiative that was led by an Eroni. Another explanation, put forth by Rzehak (2001: 76), is that Tajiks at this time did not see language as a distinguishing factor in their identity and instead preferred to see themselves as Turkestanis. This view is supported by the fact that Mu’in and Ayni, unlike Alizoda, avoided appealing to any “Persian nation” in their articles in Šū’lai Inqilob. Instead, they used the term Turkestani to describe themselves. Furthermore, to further demonstrate their adherence to a Turkestanı identity over a ethnic-linguistic one, alongside writing art-

---

8. An announcement regarding the Persian Section published in 1919 in Šū’lai Inqilob showcases the contemporary views on Persian-speaking ethnic groups. The announcement states that the sū’bai fors ‘Persian Section’, for Iranians, Afghans and Tajiks, has been opened in Tashkent. Moreover, it recommends the printing and publishing of literature for the fors ‘Persians’ and in forsī ‘Persian’ (Bergne 2007: 23).
icles in Persian they made contributions in Chaghatay\footnote{As explained in section 2.5, Chaghatay has been used with different meanings throughout the centuries. Chaghatay, sometimes also called Turki, refers here to the Turkic literary language in Central Asia before the creation of national languages, such as Kazakh and Uzbek.} to other journals and papers (Bergne 2007: 25).

### 2.8 National-territorial delimitations

In 1918, one year after the Bolshevik revolution, what had previously been the GGT became the Autonomous Turkestan Republic. In 1924 this too was dissolved and incorporated into the Soviet Union through a national-territorial delimitation which made present day Tajikistan a part of Uzbekistan, comprising the Autonomous Republic of Tajikistan and the Autonomous Region of Gorno-Badakhshan. Alongside new territorial boundaries, the Soviet system also aimed at forging national identities, languages, literatures and histories for each of the new nations (Roy 2000: 61).

Many scholars have seen a deliberate policy of “divide and conquer” in the Soviet Union’s national-territorial delimitation of Central Asia, but it could also be argued that the delimitation was greatly influenced by nationalist forces within Central Asia itself, and that the seemingly arbitrary territorial boundaries can be attributed to ignorance and an intermingled demographic landscape that did not lend itself well to the formation of clear-cut nation states (Nourzhanov and Bleuer 2013: 37).

Tajikistan was not turned into a Soviet Republic in its own right until 1929, when, in order to make it more economically viable, it was also given the province of Sughd with its capital city Khujand. But the cities of Samarkand and Bukhara, so important for Tajik culture and language, remained in Uzbekistan (Roy 2000: 61). It can thus be argued that an unfairly large Uzbek Republic had been created at the expense of areas and cities which had been, and still were, demographically dominated by Tajiks (Nourzhanov and Bleuer 2013: 38).

According to Roy (2000: 121), this was a deliberate attempt at “deculturation” of the Tajiks on the part of the Soviets. By severing the new nation state from its two main cities of culture and influence, and thus also cutting the Tajiks in Tajikistan off from their urban elite and Persian past, the Soviets wanted to create a new national identity the basis of which would be the rural populations, who were less marked by Persian identity.

The transfer of Tajik elites from Samarkand and Bukhara to Tajikistan proved troublesome, since the majority of them, Ayni being a case in point, preferred to remain in their towns of origin. The few who did relocate staffed the universities and media, but failed to become a political faction of their own in the midst of a rural majority who “lacked any vision of either state or nation” (ibid.: 121–122).
2.9 Tajik-Turkic identities and relations

Demographically, the Tajiks of Transoxania stand out as the only sizable ethnic group in the region who do not speak a Turkic language. The contemporary definition of a Tajik is a sedentary Persian-speaking Sunni Muslim on the eastern side of the Iranian border, i.e. in Afghanistan or Transoxania (Nourzhanov and Bleuer 2013: 27). Some further explanatory remarks must, however, be made regarding what peoples are subsumed under this ethnonym. The Pamiris, who inhabit the mountainous region of Badakhshan, are also seen as Tajik, even though they natively speak East Iranian languages and belong to the Isma’ili branch of Islam. The Twelver Shi’a minority in Uzbekistan, the descendants of captives taken from Marv and Khorasan, were called Eroni (Iranian) and not Tajik. On the other hand, the Tajik-speaking Jews of Bukhara and Samarkand have “always been seen as Tajik” (Perry 2009).

In Afghanistan, Tajiks make up the second largest ethnic group after the Pashtuns. As with the Eronis of Samarkand and Bukhara, the Twelver Shi’a faith and Iranian ancestry of the Fārsiwān in Herat also set them apart from the general Tajik population. Likewise, the Hazara, of Mongol descent and Twelver Shi’a faith, are also seen as constituting a separate ethnic group (ibid.).

Tajik historians trace the origins of the ethnonym “Tajik” to the Persian word tāj ‘crown’ (Eden 2017). A more linguistically plausible and agreed upon explanation is that the word tājik is derived from the Middle Persian tāzīk ‘Arab’. It is an adjectival derivation from the name of an Arabic tribe, Tayy, mirroring the derivation of rāzīk (New Persian: rāzi) ‘citizen of Ray’ from the city Ray. Together with the Arab conquest of Central Asia, there was also a large influx of Persians who settled in the area. What was once a label used exclusively for Arabs came to be applied to the Persians accompanying the Arabs, and then to everyone who spoke Persian (Perry 2009).

---

10. The use of Iranian Shi’a slaves in agriculture, domestic work and bureaucracy in Khiva and Bukhara was widespread well into the period of Russian dominance in the region (Dunbar 2018). Estimates as to the number of Iranian slaves in Bukhara during the 19th century range from 30,000–60,000 individuals (Eden 2017: 924). By some accounts, the Emirate of Bukhara depended on Iranian slaves to such an extent that the abolition of slavery had a crippling effect on the efficacy of the Emirate. Slaves constituted a considerable portion of the military and the bureaucracy. Upon their emancipation, many former slaves returned to their place of origin leaving behind a void that had to be filled with salaried personnel, resulting in further expenses for the Emirate (ibid.: 947).

11. Historically, the Tajik language as used by the Jews of Bukhara and Samarkand has displayed some idiosyncrasies in lexicon, spelling and pronunciation. Nowadays, however, there are no discernible differences between this type of Tajik and that of the general Tajik-speaking population in these cities, not even to the extent of there being a discernible accent (Ido 2017: 82–83).

12. From fārsī-zabān ‘Persian-speaking’, a term given to them by the Pashtuns.
One of the earliest attested uses of the term *tājik* is found in a manuscript of a Manichean hymn from ca 1000 CE. In this hymn, Persian-speaking Manicheans, who had sought refuge in Central Asia from Zoroastrian and Islamic persecution in Mesopotamia, are referred to as being *tājik* (Brasher 2011: 106). Around the same time, the term also appears in the works of the Persian-speaking historian Bayhaqi, where an Iranian man describes himself as Tajik to Masʿud of Ghazni in 1039 CE (Bergne 2007: 3–4).

It is impossible to fully grasp the significance of the word *tājik* without also mentioning its ethnic counterpart *tork* ‘Turk’. This opposition is also apparent in the cultures and languages of the two peoples. For example, in Ferdōsi’s Šāhnāme, Turān is the mythical homeland of the enemies of the Iranians—the Turks. Within the Turkic domain there are also testaments to the long-standing Iranian-Turkic relationship. Turkic proverbs recorded in the 11th by the Turkic lexicographer Mahmud al-Kāšgari point to a relationship which was sometimes symbiotic, sometimes hostile (Golden 2006: 17).

By the 19th century, the term *tājik* was used in ways that did not fit the early Soviet ethnographers’ intention of classifying the Persian-speaking population into an ethno-national group. The common term for the Tajik-Uzbek urban residents, Sart, was actively undermined in favour of the clear-cut ethnonyms Tajik and Uzbek (Brasher 2011: 107). The term Sart had undergone various changes in its meaning, just like the term Tajik. However, according to the Russian orientalist Bartol’d, by the end of the 19th century the term had come to denote “an Uzbekized urban Tajik” (Bergne 2007: 8).

### 2.10 Language policy

The 19th century saw the forming of national identities in Europe and Latin America, and after the 1920s, Turkey, Iran and to some extent Afghanistan also implemented reforms which aimed at creating nation states on a European model, with an emphasis on eradicating the *dhimmi* system, granting more rights to women and untangling traditional loyalties. The shift from empire to nation state also involved the suppression of ethnic minorities and their languages in favour of standardised official state languages. What sets the Soviet nations of Central Asia apart from other countries such as Turkey, is that this shift towards nation states and the accompanying language policies were externally driven, rather than an integral part of an independence movement. Moreover, the Soviets were also concerned with making the citizens feel like not only part of a particular nation, but also of the greater Soviet nation (Roy 2000: 74).

---

13. Under the *dhimmi* system, Non-Muslims were afforded special status in lands under Islamic rule. They were allowed to keep their religions and were exempted from military service by paying *jizya*, a special poll tax (Cahen 1991: 227).
The Soviet system of nation building was intrinsically linked to the formation and tailoring of a specific language for each nation with the aim of differentiating it from other languages both within and outside of the Soviet Union. The formation and delineation of the “Tajik” language is a case in point. The Tajik language lies at the far end of a continuum of Persian that stretches from western Iran, through Afghanistan and into Transoxania. The vernaculars of Persian spoken in these areas all share a common canon of Classical Persian literature. The literary Persian of Tajikistan (which later would become Modern Literary Tajik) remained more or less identical with the rest of the varieties of Persian until at least the 16th century (Perry 2005: 1). What binds this vast stretch of land together in time and place, culturally and linguistically, is Classical Persian, the type of New Persian written with an adapted Arabic alphabet that came to be dominant after the Islamic conquest of Central Asia.14 Classical Persian served as a *lingua franca* in the Iranian cultural sphere, and also developed, alongside Arabic, into a high-prestige language in its own right, giving rise to a plethora of classical poetry and literature that to this day is understood by Persian-speakers. Its sphere of influence, however, has historically been much wider than the current lands where its native speakers still reside; it was once used as an administrative language in the northern part of the Indian subcontinent and in Anatolia (Thackston 1994: ix).

Classical Persian can thus be seen as high-register language that spawned a literary canon common to all Persian speakers, wherever they may live. On the local level, however, spoken Persian has evolved differently. In Central Asia, surrounded by various Turkic-speaking groups, the Persian spoken in northern Transoxania displays a high degree of Turkic influence, not only in its everyday vocabulary, but also in its grammatical structure.

What had originally been prepositions in the literary language had shifted towards a postpositional status in the spoken Tajik of northern Transoxania. A prepositional phrase such as *ba šahr* ’to the city’ evolved into *šahr=ba* in the spoken vernacular (Roy 2000: 75). The vernacular Tajik of the northern part of Transoxania also had other distinguishing idiosyncrasies brought on by Turkic influence, such as constructing attributive relative clauses with the suffix -*agī*: *kitob=i xondagi=(y)am* ‘the book that I read’.15

The Soviet language policy was devised to highlight and officially promote such vernacular idiosyncrasies as a vital part of developing a new literary language for the nascent republic of Tajikistan. The Tajik language also underwent two changes of script, which further distanced it from the Persian

---

14. The earliest New Persian texts were written in a number of scripts (Hebrew, Manichean, Syriac, Pahlavi and Avestan) depending on the religious traditions of the authors. In fact, the oldest surviving New Persian documents are in the Judeo-Persian variety, written with the Hebrew script. Other documented varieties of Early New Persian belong to Manichean, Christian, and Zoroastrian speakers of Early New Persian (Paul 2017).

15. See section 4.2.1 for more examples of Turkic influences on Tajik.
varieties of Afghanistan and Iran and their shared literary canon. The first change, from the Perso-Arabic to the Latin script, was not imposed by Moscow, but was a result of Pan-Turkic fervour, which first began in Azerbaijan in 1922. Atatürk enthusiastically followed suit in Turkey in 1928. In the same year, Uzbekistan and Tajikistan, the latter still an autonomous appendix of the former, also implemented the change of alphabet with the approval of the Bolsheviks as “one of the progressive trends of the age” (Perry 1996a: 281). By the end of the 1930s, however, a unified script uniting Turkic-speaking countries within and outside of the Soviet Union was seen as a possibly disruptive Pan-Turkic threat to the Soviet hegemony. For the Turkic languages and Tajik, a second change of alphabet, from Latin to Cyrillic, was therefore implemented in 1939. Around the same time, materials in the Perso-Arabic script were forbidden (Perry 1996b: 571).

The supplementary letters for phonemes not found in Russian (six in the case of Tajik) were given a separate design for each Soviet nation in an effort to isolate the Turkic nations from each other and Tajikistan. For example, in Tajik, the supplementary letters were constructed with the help of additional cedillas. The voiced counterpart to the Russian letter č (če) /t͡ʃ/ was thus created by adding a cedilla to form a new letter: ҷ. In Uzbek, however, both the phonemes /ʒ/ and /d͡ʒ/ were represented by the Russian character ж (že) (Roy 2000: 77).

Russian (and other) linguists deemed the qualitative difference between the traditionally short vowels /i/ and /u/ and their long counterparts /ī/ and /ū/ to have been neutralised in Tajik. They therefore did not see any use of marking the contrastive difference between long and short vowels in the new alphabet. In poetry, however, the difference between the short and long vowels is crucial. In the new alphabet, the scansion of classical poetry became cumbersome and fraught with pitfalls. Moreover, producing poetry on a classical foundation with the new alphabet puts aspiring poets at risk of committing serious errors in metre unless they are aware of the original quality of the vowels. According to Perry (1996a: 284), the mistake of not transposing this difference into the new alphabet resulted in the “literal destruction of Tajik poetry”.

Moreover, since the Tajiks were no longer acquainted with the Perso-Arabic alphabet, the etymological relationships between words of Arabic origin that share the same root consonants could no longer be visibly appreciated. There was also a decisive break from most of the identity-building literary works of the past. Tajiks could no longer, by themselves, freely access such seminal works as Ferdawsi’s Šāhnāma, Saʿdi’s Golestān and the ghazals of Ḥāfeẓ. Instead, they had to rely on ideologically sanctioned, selectively edited versions of these classics (ibid.: 284–285).

Due to the high concentration of non-Tajiks, mostly Russians, in bureaucracy and radio, there was no incentive to develop a Tajik vocabulary for these domains. For example, Dushanbe had a high proportion of Russians and
other European immigrants. Around the year 1926, 76 per cent of the bureaucrats and technicians were non-Tajiks, and at the radio station, all were non-Tajiks (ibid.: 285).

Russian became a high-status language that continued to influence Tajik with loan words and calques, even to the point of replacing existing Tajik terms with Russian ones. Tajiks who wanted to provide their children with the best possible education put them in Russian-medium schools. The process leading towards bilingualism was endorsed by the Soviet authorities, the reasoning being that the sooner the masses learn Russian, the faster Communism can be established. It was, however, a very one-sided process; in 1989 in Dushanbe, only 2.3 per cent of the Russians knew Tajik, whereas 60 per cent of the Tajiks knew Russian (ibid.: 286–287).

In the 1980s, several factors contributed to awakening Tajik intellectuals to the endangered status of their native language. One was the war in Afghanistan, in which many Soviet Tajiks participated. The experience left them with a widened sense of the international nature of their native language and culture. Moreover, there was a growing awareness that very few of the national languages were protected by the constitutions of the various Soviet republics. In 1990, an All-Union law was passed on language status, but by that time 10 republics had already passed their own constitutional amendments to this end.

In 1989, a draft for a language law (qonuni zaboni) was published in the Tajik press for public discussion. It generated a lively discussion. On 9 September 1991, Tajikistan declared its independence as a sovereign state. With this came a need for redrafting the language law to fit the new situation. The new draft, which was published for debate in April 1992 and subsequently adopted, tasks the state with reviving and teaching the Perso-Arabic script in all educational institutions, and implementing an official reversion to the Perso-Arabic alphabet “in the near future” (ibid.: 301). This was put on hold, however, as a civil war ravaged the country between 1992 and 1997.

In the 1990s the relationship between Iran and Tajikistan deepened. Iran donated schoolbooks for learning the Perso-Arabic alphabet, opened libraries and contributed to cultural activities (Zafar 2016: 32). However, these kinds of efforts to revert to the Perso-Arabic script did not lead to an official reform. To this day, no efforts have been made to put the 1992 language law into effect, and the de-facto script of the country remains Cyrillic (ibid.: 30).

In short, Modern Literary Tajik (zaboni adabii hozirai tojik) was to a great extent manufactured under Soviet tutelage. It was loosely based on the literary Persian of Bukhara, and displays many features that are typical of the Northern dialects, such as syntactic structures calqued onto Uzbek usage. It was planned and exemplified by Soviet writers who now are held in high esteem in Tajikistan, such as Sadriddin Ayni, who ironically never ceased
to write in the Perso-Arabic script and continued to live in Bukhara, while writing for a budding nation state with Dushanbe as its capital.

![Map of Tajik Dialects](image)

*Figure 2.1. A simplified map of the different dialects of Tajik, as described by Windfuhr and Perry (2009: 420–421).*

Much research remains to be done in mapping out the different dialects of Tajik. However, one can roughly follow the topography of the Oxus basin and bifurcate the area into two parts, geographically and dialectally: the lowlands in the Northwest and the highlands in the Southeast. A more detailed grouping can be made in four parts as follows:\(^\text{16}\)

2. **Central**, covering the upper Zarafshon valley.
3. **Southern**, stretching south and east of Dushanbe, comprising the Kulob and Qarotigin regions, including Gharm, as far as the Pamirs.
4. **Southeastern**, comprising the Darvoz region.

After Dushanbe became the capital of Tajikistan in 1924, there was an influx of Tajiks from all over the region, and especially from Samarkand and Bukhara, since these two cities were allotted to Uzbekistan. For a long time Dushanbe lacked a stable demographic profile that would lend it a linguistic status of its own (Perry 2005: 3). Dushanbe nominally belongs to the Northern dialects, but since independence in 1992 it has seen an increased influx of writers and speakers from other regions.

---

\(^\text{16}\). This table is based on Perry and Windfuhr’s sketch of Tajik (2009: 421). It is very similar to an earlier classification given by Perry (2005: 3). The difference is that in the work from 2009, the Southeastern dialect is confined to the Darvoz region instead of covering the whole of Nagorno Badakhshan.
3. Evidentiality – definitions and classifications

The main purpose of this chapter is to provide a definition of evidentiality to serve as a basis for identifying and discussing evidential verb forms in Tajik. Since there is an ongoing debate about the relationship between evidentiality and related concepts that share an affinity with evidentiality, such as epistemic modality and mirativity, these will also be described from a typological and linguistic standpoint in an attempt to delineate the conceptual boundaries between them. Moreover, given that Tajik uses the Perfect and forms modelled on the Perfect as means for expressing evidentiality, the role of the perfect as a vehicle and originator of evidential meanings will be discussed.

3.1 Defining evidentiality

In this section we will briefly look at the different ways in which evidentiality can be defined: (1) notionally, (2) on the basis of its scope properties, and (3) as a grammatically coded notion. In brief, as far as the notion is concerned, I align myself with the majority of contemporary studies on evidentiality by defining it in terms of information source.\(^1\) As regards the scope properties

---

\(^1\) Referring to either information source or source of information when describing evidentiality is very common in the more recent literature dealing with this linguistic phenomenon. For example, Aikhenvald (2004: xi) describes evidentiality as “the grammatical means of expressing information source”. Bybee et al. (1994: 95) write that “evidential grams indicate something about the source of the information in the speaker’s assertion”, a postulation that the authors indicate is based on research by Chafe and Nichols (1986) and Willett (1988). Other works that also adhere to this view are Boye (2012: 2) and Anderson (1986: 273).
of evidentiality, only the type of evidentiality that scopes over propositions, i.e. propositional evidentiality, will be dealt with in this dissertation. Lastly, although evidentiality in a general sense can be both grammatical and lexical, unless otherwise qualified, evidentiality will refer solely to grammatical evidentiality.

3.1.1 The notional definition of evidentiality
This section will explore the different ways in which evidentiality is notionally defined and how this is reflected in the nomenclature used, what the terms in question have in common and what sets them apart, and why it is of importance to be cautious when defining evidentiality.

Although the term information source has gained traction among scholars in the last decades, there are also competing terms. For example, Boye (2012: 19) employs the term (epistemic) justification interchangeably with information source, and evidence; the latter term is also used by other authors, viz. Palmer (2001: 8).

At first glance, the terms (epistemic) justification, information source and evidence might all appear to capture the same notion. However, the term information source stands out among these three in that it is not tinged with any meanings of justification. As argued by Boye (2018: 264), despite his former employment of the term epistemic justification, the term information source is a better choice since it is neutral with regard to any type of commitment to the “evidentially modified proposition”, whereas the terms evidence and (epistemic) justification may imply that there is a commitment to the proposition. The reasoning behind this is that if one says, or indicates, that one has evidence for a proposition, or if one finds the proposition justified in any way, it follows that one also vouches for that proposition.2

This notwithstanding, these semantic differences are often ignored in the body of evidentiality research, and the terms are used interchangeably with a seemingly neutral stance as to their commitment value (ibid.: 264). For the above-mentioned reasons, the term information source will be used synonymously with evidentiality in this dissertation.

3.1.2 Scope
Scope, a term originally borrowed from logic, concerns the part, or extent, of a sentence with which an operator (for example a quantifier or negation)

---

2. Aikhenvald (2004: 4–5) argues that one cause of the conflation of evidentiality with reliability is the former term itself, due to its strong link to “evidence”, as it is used in a non-linguistic sense. Aikhenvald refers to definitions of Chafe and Nichols (1986: vii) and Matthews (1997: 120) as examples of this unorthodox taxonomy, since they also include the parameter of reliability in their descriptions of evidentiality.
combines in meaning (Szabolcsi 2011: 1605; Matthews 2014). In a sentence such as *She is perhaps here*, the adverb *perhaps* scopes over the entire clause. A fundamental line of reasoning for Boye (2012: 183) is that all meanings shared by the categories of propositional evidentiality and epistemic modality share *scope properties*. By this is meant that the common denominator for both epistemic modality\(^3\) and evidentiality\(^4\) is their ability to only scope over *propositions*, i.e. “meaning units that can be said to have a truth value” (Boye 2018: 264).\(^5\) Boye (2012: 185) asserts that *scope properties* are often mentioned when defining epistemic meanings or expressions, but that there is no consensus on what these properties actually are. There are, however, three terms which often occur in descriptions of these properties, namely *illocutions*\(^6\), *states of affairs* and *propositions*. While there is a great deal of literature on the nature of illocutions, there is far less on the differences between states of affairs and propositions.

In short, states of affairs can be said to occur or not occur, whereas propositions can be said to have a truth value. One way of illustrating the difference between states of affairs and propositions is to look at how they are contrasted in complements. If we take English as an example, the complements of verbs related to cognition and perception can take two forms depending on whether they express states of affairs or propositions. Consider the verb *to forget* and the two different types of complements it can take:

1. **State-of-affairs vs propositional complement**
   a. I forgot [to write the letter]. (State-of-affairs complement)
   b. I forgot [(that) I wrote the letter]. (Propositional complement)

   (After ibid.: 189).

In other words, what is forgotten in (1a) is an action, whereas in (1b) it is a proposition, i.e. a piece of information that can be true or false. Another way

---

\(^3\) See also van der Auwera and Plungian (1998: 82), where a distinguishing quality that separates epistemic modality from non-epistemic modality is said to be the former’s ability to scope over entire propositions.

\(^4\) Non-propositional evidentiality can however be seen in a small minority of the languages of the world, located primarily on the Northwest Coast of North America. In these languages, the evidential markers (mainly demonstrative pronouns and determiners) have a purported scope which is “limited to a noun phrase” (Jacques 2018: 109). In most cases these languages have a system distinct from that of propositional evidentiality (ibid.: 123). Boye (2018: 265) argues that the display of different systems for propositional and non-propositional evidentiality possibly also hints at the existence of a notional separation between them. There is still too little material available to draw definite conclusions regarding this matter.

\(^5\) This definition of what constitutes a proposition can be traced back to Aristotle, who in the fourth chapter of *On Interpretation* remarks that not every sentence is a proposition, but only those “in which there is a truth or falsity” (trans. Arens 1984: 23).

\(^6\) Actually, the term “speech act” is used by Boye (2012: 187), but he later (ibid.: 188) makes clear that this term will be understood as “illocutionary meanings”, which are the only type of speech acts that have meaning.
to describe the difference between the two sentences is that the complement in (1a) is a state-of-affairs, while the complement in (1b) is a proposition.

Evidentiality and epistemic modality, which are only concerned with meaning units that can be said to have a truth value, can thus only scope over propositions. With this knowledge in mind, one way to test whether a complement is of the state-of-affairs kind or the propositional kind is to see if the complement can function within the scope of an epistemic or evidential adverb.7 An evidential adverb, such as allegedly, cannot scope over the state-of-affairs complement in (1a): *I forgot allegedly to write the letter. Conversely, the evidential adverb allegedly can scope over the propositional complement in (1b): I forgot that I allegedly wrote the letter (Boye 2010: 293–294).

3.1.3 Evidentiality as a grammatically coded notion

A main point in Aikhenvald’s (2004: 7) definition of evidentiality is that it is a “category in its own right, and not a subcategory of any modality”. The term category, as used in this way by Aikhenvald, denotes a cross-linguistic descriptive category, which can be thought of as a “notional generalization over distinct but related linguistic meanings” used for the purpose of describing and comparing language-specific phenomena (Boye 2012: 9). Cross-linguistic descriptive categories such as tense, aspect and gender are defined independently of specific languages with the intention of facilitating comparisons between languages.

Aikhenvald (2018: 3) makes a distinction between real-world phenomena and grammatical categories. The grammatical category is seen as a grammatical manifestation of a real-world phenomenon. Tense is therefore a language-specific, grammatically expressed counterpart to the real-world phenomenon of time. By the same token, information source is seen as a real-world phenomenon which finds its grammatical manifestation in the grammatical category of evidentiality.

Aikhenvald (2004: 148) argues in particular for the notion that not all languages that can express evidential meanings lexically are truly evidential, arguing that just as sex distinctions can be expressed in a language like Hungarian or English (cow vs bull; hen vs rooster) this does not entail that there is a grammatical category of gender in these languages. Another example along the same lines is that time words such as “yesterday” and “today” cannot be seen as tense markers.

This distinction poses a number of challenges, however. Above all, the assumption that there is a clear-cut distinction between lexicon and grammar is a contested issue. Many linguists, particularly those working with

7. Also see Narrog (2012: 52–53) who makes a similar observation on the scope orderings of the speech-act-oriented modal expression I’m sure and the event-oriented modal expression can. The former can scope over the latter, but the inverse is not possible.
grammaticalisation theory (e.g. Lehmann 2015: 15) or cognitive linguistics (e.g. Langacker 1988: 19), assert that grammar and lexicon should not be thought of as constituting two separate domains, but rather as lying at opposite ends of a continuum. Still, even if one settles for a supposed division between grammar and lexicon, any delineation of the boundary between the two would ultimately depend on “the theoretical perspective one adheres to” (Squartini 2018: 273).

**Systems and strategies**

In Aikhenvald’s taxonomy, evidentiality proper is only possessed by a language if the marking of evidentiality is (a) grammatical and (b) obligatory (Aikhenvald 2004: 6). Languages which have optional grammatical marking of evidentiality fall into the category of languages with “evidentiality strategies”. This is a third category for languages that to some extent have grammatical marking of evidentiality, but its grammatical manifestation is not solely dedicated to the expression of information source (ibid.: 105).

<table>
<thead>
<tr>
<th></th>
<th>Grammatical</th>
<th>Obligatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidentiality proper</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Evidentiality strategy</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>No evidentiality</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

The task of deciding which languages belong in which group is fraught with difficulty when it comes to certain Eurasian languages, among them Persian and Tajik, which have developed non-first-hand meanings in their use of the perfect (ibid.: 38). Moreover, even if one makes a distinction between languages that have evidentiality proper, evidentiality strategies, and no grammatical evidentiality, some languages still fall between these categories (Plungian 2010: 42).

This opposition between languages that have “evidentiality proper” and those that have “evidentiality strategies” has not passed uncontested in the linguistic community. Diewald and Smirnova (2010: 4), for example, argue that Aikhenvald’s definition conflates two separate entities by letting a grammatical category totally eclipse a conceptual domain. This, they contend, is “analogous to equating the grammatical category ‘tense’ with the conceptual domain ‘temporality’” (ibid.: 4).

For the purposes of this dissertation, I will mainly be concerned with the grammatical coding of evidentiality. In other words, only the Tajik verb forms which have a marked evidential meaning will be under scrutiny, specifically the register II verb forms and the Presumptive. In almost all languages source of information can also be expressed lexically using adverbs with meanings such as “reportedly” and “apparently” (Aikhenvald 2004: 11).
However, these cases will only be analysed to the extent that they shed light on the evidential status of the proposition. A lexical evidential marker can scope over a proposition in conjunction with a grammatical marking of evidentiality in what is called *harmonic combination*—a term which was originally coined by Lyons (1977: 807–808) to describe how both lexical and grammatical expressions of epistemic modality used in tandem can scope over the same proposition in such a way that they reinforce the meaning of each other.

### 3.2 Delimiting evidentiality from related notions

Historically, it has not been uncommon for linguists to view evidentiality as epistemic (Boye 2018: 263), i.e. as marking the speaker’s attitude toward a particular proposition (Aikhenvald 2004: 3–5), or as a kind of mood (ibid.: 13). However, as this section will make clear, there are a number of logical reasons for these conflations. The purpose of this section is to account for two of the more prominent of these related notions, namely epistemic modality and mirativity. Moreover, since evidential markers typologically often originate from the perfect, the emergence of evidentiality from the perfect will also be discussed.

#### 3.2.1 Evidentiality vs epistemic modality

Epistemic modality is concerned with “the speaker’s judgement on the truth of a proposition” (Ziegeler 2006: 262), a definition which may seem somewhat confusing considering that etymologically, *epistemic* has to do with knowledge, as the original Greek *epistēmē* ‘knowledge’ informs us. However, in the philosophical literature, and by extension in the linguistic literature, the term epistemic is not only tied to knowledge, but also to belief (Boye 2012: 16).

It is thus not surprising that terms such as *degree of certainty*, *degree of commitment* or *degree of confidence* are used when describing epistemic modality (ibid.: 21). In fact, there is a strong philosophical tradition dating back to Plato of including justification and evidence in the definition of propositional knowledge. By propositional knowledge is meant the knowledge of a proposition. In epistemology, this is the type of knowledge that concerns facts and truths, as opposed to other types of knowledge such as knowledge of acquaintance (‘I know Jack from HR’), knowledge-where (‘I know where Jack is’), and knowledge-how (‘I know how to touch-type’) (Ichikawa and Steup 2018). However, as far as linguistics is concerned, the knowledge-where type

---

8. Lyons bases his observations on Halliday (1970: 331), who in turn never uses the term “harmonic combination” when describing modal pairs which reinforce one another when occurring in concord. For a related phenomenon, where the harmonious constituents are both grammatical evidentials, see Aikhenvald (2004: 87–95).
can also be said to be propositional since it fits Boye’s definition of being a meaning unit “that can be said to have a truth value” (Boye 2018: 264).

Using the criteria put forth in what Williams (2001: 16) calls “the standard analysis” of knowledge, a person can only truly know something if all the three conditions of belief, truth and justification are fulfilled. This tripartite analysis is often abbreviated as the JTB analysis for “justified true belief” (Ichikawa and Steup 2018):

\[
S \text{ knows that } p \text{ (where } S \text{ stands for an arbitrary person and } p \text{ for an arbitrary proposition) if and only if:} \\
1. S \text{ believes that } p. \quad \text{(The belief condition)} \\
2. p \text{ is true.} \quad \text{(The truth condition)} \\
3. S\text{'s belief that } p \text{ is appropriately justified.} \quad \text{(The justification condition)}
\]

(After Williams 2001: 16).

Whereas the truth condition is binary, and not epistemic in nature, the belief condition can be of varying degrees and the justification condition can be of varying types. Their internal variability gives rise to different types of “knowing” that differ in their internal make-up according to the subject’s strength of belief and type of justification.

A number of linguists (Aikhenvald 2004: 6–7; Palmer 2001; Narrog 2009) have acknowledged that in many languages there is a dovetailing between information source (evidentiality) and the speaker’s certainty about the information (epistemic modality). In other words, even if we treat evidentiality and epistemic modality as two distinct grammatical categories, we will inadvertently come across evidential markers imbued with epistemic overtones and, vice versa, epistemic markers imbued with evidential overtones (Wiemer 2018: 85).

Within the literature on evidentiality and epistemic modality, there are generally speaking three major trends in how the relationship between these two categories is viewed. Those who regard the two categories as separate entities can be said to hold a “disjunctive” view. Aikhenvald is an influential proponent of this view. Those who instead see them as two distinct but to some extent overlapping categories can be said to hold an “overlapping” view, and lastly, those who see one category as a subcategory of the other hold an “inclusive” view.9

A fourth way is proposed by Boye, who suggests a supercategory—epistemicity—which subsumes the distinct categories of evidentiality and epistemic modality. Boye (2012: 2) argues that one can indeed categorise both evidentiality and epistemic modality as cross-linguistic descriptive categories in their own right, but also suggests that these two categories can be seen

9. For a more detailed account of the opinions of scholars concerning these views, see Boye (2012: 1).
as belonging to a superordinate category of *epistemicity*, which subsumes expressions of degree of certainty and expressions of information source.

![epistemicity](image)

*Figure 3.1. According to Boye (2012), evidentiality and epistemic modality can both be subsumed under the broader category of epistemicity.*

In order to further explain the intimate relationship between epistemic modality and evidentiality, I will here give a brief account of Narrog’s (2009, 2012) theories regarding the interconnectedness of epistemic modality and evidentiality, and how Narrog’s model can be integrated with that of Boye.

Narrog essentially agrees with Boye regarding the notional relationship between evidentiality and epistemic modality. Narrog (2009:8–9) prefers to view *factuality* as the defining quality of modality as a grammatical category. In other words, all modal expressions, be they epistemic or non-epistemic, are concerned with marking a statement as non-factual; i.e. what is being modally expressed is not part of reality at the time of expression. Non-factual, as used here, thus denotes that something is “non-realized or non-actualized” (Narrog 2012:6). In other words, if the factual status of a proposition is marked as undetermined, it is modalised. Negations, being concerned with expressing negative facts rather than non-facts, are thus not modal (Narrog 2009:9).

To exemplify the ways in which non-factuality is the core quality for modal expressions, we can consider the two different ways in which the modal *must* is used in a sentence such as the one in (2).

(2) They must be afraid of you. (ibid.)

Depending on the context in which this sentence is uttered, it can either have an epistemic (“I gather that they are afraid of you”) or deontic reading (“I insist that you scare them”). What the two different readings have in common, however, is that the modal marker *must* marks a proposition as being “non-factual, or not belonging to reality at the time of speech” (ibid.:8–9).

Narrog (2012) proposes a two-dimensional model of modality, the constituents of which being made up of (1) a binary opposition of *volition*10 where modal statements are either volitive or non-volitive, and (2) a sliding scale based on *orientation*, where modal statements lie somewhere on a scale between speech-act orientation and event orientation.

---

10. This volitional status is also referred to as having an “element of will”. This binary distinction between volitional and non-volitional modality owes much to the theories of Jespersen ([1923] 1963:313–321).
A modal lacking an element of will, i.e., a non-volitive modal, must be regarded as epistemic, since it is the presence of an element of will that renders modal statements deontic. In order to illustrate the volitional status of modal sentences, Narrog (2012: 47) gives the following two examples:

(3) That guy must be dealt with. \((\text{Deontic, volitive})\)
(4) That guy must have a screw loose. \((\text{Epistemic, non-volitive})\)

In (3) we have a statement that can be said to contain an element of will; even though it is arguably not necessarily the will of the speaker, it does contain an inherent aim regarding what has to be done, whereas in (4), we cannot see any trace of an element of will.

As for the orientational scale (the y-axis in Narrog’s model), there is a sliding scale between event orientation and speech-act orientation, the latter being made up of three different parameters which constitute a nested scale: speaker orientation (subjectivity), hearer orientation (intersubjectivity) and discourse orientation. In other words, for a modal marker to be speech-act oriented it must concern either the speaker, the addressee, or the speech situation itself. Among these parameters, speaker orientation is the most important, since it is inevitably reflected in natural speech by the use of deictic expressions (such as tense and personal pronouns). Speaker orientation is thus intrinsic to natural speech, in essence forming an essential prerequisite for hearer orientation and discourse orientation, which are seen as occupying a higher position on the scale of speech-act orientation (ibid.: 49–50). In practical terms, this means that clausal mood and illocutionary force belong on the higher end of the orientational scale, with discourse orientation in many
cases being “the most advanced or final stage in speech act orientation” (Narrog 2012: 51). Conversely, on the lower end of the orientational scale we have event orientation, which is “concerned narrowly with the described situation itself and the participants of the event” (ibid.: 51).

The models of Narrog and Boye agree on certain core principles. First, speaker-oriented units of meaning (propositions) are more complex / advanced\(^{11}\) than event-oriented ones (states of affairs) (Boye 2012: 288; Narrog 2012: 50). In other words, there is an increase in complexity as the meaning unit goes through a process of “category climbing”, from being event oriented (nested in a state of affairs) to speaker oriented (nested in a proposition). Second, epistemic modality and evidentiality typically occur within speaker-oriented (propositional) non-volitive settings (Narrog 2012: 54; Boye 2012: 183).

\(^{11}\) Propositions are more complex on two accounts. First, propositions are conceptually more complex than states of affairs in that propositions have a higher level of “layered semantic structure”; i.e. a reference to states of affairs is inevitably intrinsic to a proposition whereas “states-of-affairs may be found in the absence of propositions” (Boye 2012: 287). Second, there is a cross-linguistic tendency for propositional complements to be more morphologically complex than state-of-affairs complements (ibid.: 288).

3.2.2 Evidentiality and mirativity

In 1878, the French philologist and diplomat Dozon described two special verb forms in Albanian for expressing mainly admiration, astonishment, and
to a lesser extent, irony. For this reason, he dubbed them “admiratives” (Dozon 1878: 226). For more than a century, this peculiar phenomenon of grammatically marking surprise would lie dormant within the linguistic community until DeLancey (1997) brought it to light again, this time dubbing it “mirativity”. DeLancey (1997: 36) defines mirativity as a grammatical category that marks a proposition as “new to the speaker” and “not yet integrated into his overall picture of the world”. In other words, mirativity is the grammatical marking of surprise or new information (Aikhenvald 2018: 18; Haan 2012: 1021).

Notwithstanding the vagueness of DeLancey’s definition—or rather, perhaps exactly because of it—the list of languages with purported mirativity has grown. Simultaneous with the adoption of mirativity as a grammatical category, the notion of mirativity as a cross-linguistic category has not been without its critics. Most vocal has been Hill (2012), who rejects the status of mirativity as a grammatical category in a series of articles. Others, like Zeisler (2017) and Peterson (2017), though not rejecting mirativity as categorically as Hill, still have certain misgivings regarding the definition and its conceptual boundaries.

There is a link between mirativity and evidentiality in so far as languages which exhibit evidential strategies “frequently develop mirative extensions” (Aikhenvald 2004: 204). Tajik is a case in point, since the Perfect is used for marking not only reportative and inferential propositions, but also mirativity (Perry 2005: 228). This is shown in the exclamation in (5), which a guest makes after having tasted a dish. The utterance is made in the Perfect, albeit with a present tense meaning.

(5) Čī xel bomaza bud-a=ast!
what type delicious be.pst-pp=cop.npst.3sg
How delicious it is! (Adapted from Baizoyev and Hayward 2004: 186)

Different theories have been put forward to explain the emergence of mirative extensions from forms that are already evidential. One is Akatsuka’s model of information incorporation (1985), which was originally devised to explain conditionals. According to this model, new information first enters the mind in the realm of Irrealis, a mental domain in which information is still uncertain and has not yet become true knowledge. Only after information processing has taken place can the piece of information move to the mental realm of Realis, and be seen as “knowledge”. Similar to this theory is that of the “unprepared mind” proposed by Aksu-Koç and Slobin (1986: 163).

According to these theories there exists within the cognitive system a mental space, dedicated to “new knowledge” which has not yet been integrated

---

with the mind. Consequently, the form one employs when speaking about
this unintegrated information is the same, regardless of whether it comes
from hearsay, inference or unexpected surprise (mirativity).\textsuperscript{13}

The type of mirativity that DeLancey (1997), and before him, Aksu-Koç
and Slobin (1986) describe, is the type of mirativity that is uttered in situ,
i.e. when the person who is experiencing a surprising situation, at that very
time, in that situation, is uttering a mirative statement. The sentence in (5)
is an example of this. In the words of Guentchéva (2017:569), this type of
mirativity is “a situationally-bound phenomenon which is actualized by the
enunciator and is always anchored in the situation of enunciation”. There are
other definitions of mirativity, however, where the mirative statement does
not necessarily have to be immediately uttered in situ, as can be seen below
in the classification made by Aikhenvald (2012):

1. sudden discovery, sudden revelation or realization (a) by the speaker,
   (b) by the audience (or addressee), or (c) by the main character
2. surprise (a) of the speaker, (b) of the audience (or addressee), or (c) of
   the main character
3. unprepared mind (a) of the speaker, (b) of the audience (or addressee),
   or (c) of the main character
4. counterexpectation (a) for the speaker, (b) for the addressee, or (c) for
   the main character
5. information new (a) to the speaker, (b) to the addressee, or (c) to the
   main character.

(Adapted from ibid.:437)

Aikhenvald does not provide the reader with any definition of the different
mirative categories listed above; e.g. what is the actual difference between
surprise and counterexpectation? At any rate, the categories describing the
type of “new information” are not as fundamental (I will treat them all as
“surprise” / “new information”) as the type of person (a, b, c) that can express
mirative statements.

Just as there are languages with “evidentiality strategies”, in which the
evidential markers are not used solely for marking evidentiality, one can also
speak of languages with “mirative strategies” (ibid.:462). As we have seen in
(5), Tajik can use the Perfect to express in situ mirativity. But as will become
apparent in section 4.1.2, Tajik also displays a so-called “3rd person mirative”,
i.e. (c) in the list above, in which an omniscient narrator marks the mirative
experience of the main character. In other words, Tajik marks mirativity
not only for (a) the speaker, but also for (c) the main character of a story.
Moreover, as will be argued in 4.1.2, it appears as if the register II forms in
Tajik in many cases mark personal evaluation (often post-situationally) rather

\textsuperscript{13}. See Aikhenvald 2004: 208–209 for a more thorough discussion and the proposal of three
interconnected “semantic paths” which give rise to mirative readings of evidentials.
than unexpected information in situ; for this reason the term *evaluative* is used in these cases.

### 3.2.3 Evidentiality and the perfect

Reference grammars and linguists often refer to the perfect as expressing a past action with “current relevance” (Bybee and Dahl 1989: 61). For example, Maslov (1990) defines the perfect as an aspecto-temporal form of the verb (“*vido-vremennaja forma glagola*”) denoting that a present state is the result of a preceding action or that a past action or state is considered important in the present. More succinctly, Comrie (1976: 52) posits that “the perfect indicates the continuing present relevance of a past situation”. Although other theories exist that describe the core function of the perfect, Dahl and Hedin (2000) and Lindstedt (2000) prefer the term “current relevance”. Lindstedt (2000: 378) sees “current relevance” as the “central and prototypical meaning of the perfect” while also noting that its expression is never the sole function of the perfect (ibid.: 378).

Indeed, in many languages of the world, alongside expressing its main meaning of current relevance, the perfect has also come to carry evidential meaning, being a so-called “evidential perfect” (Aikhenvald 2004: 38). The use of the perfect for evidentiality can be seen in languages spoken in Southeastern Europe and West Asia, such as Bulgarian, Georgian and Persian (Lazard 2001: 360; Anderson 1986: 275). If we take Persian as an example, not only can the Perfect carry a possible indirective meaning, but the verbal system also displays verb forms modelled on the Perfect which are decidedly indirective. Plungian (2010: 42) calls the forms modelled on the Perfect “perfectoid” and describes them as “typical for the binary systems of the ‘Balkan type’”. Historically, languages in which the perfect serves this kind of evidential function often develop into languages with full-fledged evidentiality (Aikhenvald 2004: 11).

Within the context of the “Great Evidential Belt” (as explained in chapter 1), it is reasonable to refer to the evidential perfect as an areal phenomenon, but the same role of the perfect is also found in languages that are not related to these languages, either genetically or areally. Typologically, there is a tendency for perfect forms to acquire reportative and inferential meanings (Lindstedt 2000: 375–376). Given the fact that the perfect has developed evidential meanings in languages which are not associated with the “Great Evidential Belt”, this development can also be seen as a general typological trend rather than an accidental syncretism, not unlike the counterfactual role accorded to the past tense in various languages (Izvorski 1997: 223–224; Comrie 1976: 108–110).
3.3 Types of evidentiality

The main opposition informing the category of evidentiality in natural languages is that between direct and indirect access to the source of information (Willett 1988: 57; Plungian 2010: 30). The direct evidentials are divided into three groups according to sense: visual, auditory, and other. The indirect evidentials consist of two groups based on the means of access: reportative and mental. Reportative can either be hearsay (i.e. second-hand) or folklore (i.e. third-hand). The mental evidentials mark information that has been arrived at through logical conclusions, either by piecing together clues gathered directly (inference) or by relying on one’s world knowledge (presumption).

According to Bergqvist and Kittilä (2020: 7), these different types of access have an inherent hierarchy as regards their evidential strength. By implication, speakers are expected to “provide the highest form of evidence for an assertion” (ibid.). For the direct evidentials, information acquired visually bears more weight than information gained through the other senses. The higher reliability of visually acquired information can also be seen in the ranking of the indirect evidentials. Bergqvist and Kittilä (2020: 7) assert that inferentially acquired knowledge is ranked higher than both reportative and presumptive knowledge because inferences are based on visually gained clues, whereas reportative information has been transmitted by others and presumptive information is acquired by processing general knowledge of the world.

![Diagram of evidential categories](figure3_4.png)

*Figure 3.4.* A representation of the evidential categories belonging to the direct / indirect dichotomy. Adapted from Willett (1988) and Plungian (2010), with some terminological adjustments.

To further illustrate the different types of access further, we can use the example of John, who in our examples will serve as the receiver of a certain piece of information (that his sister Jane is currently at home). In other words he becomes the “knower”, as it were, that his sister Jane is at home.
Examples of access types:

a. John is not at home, but he phones his mother and inquires about the whereabouts of his sister. His mother informs him that Jane is at home. \(\text{Reportative}\)

b. John is at home, and either sees or hears his sister. \(\text{Direct}\)

c. John enters the home and sees his sister’s shoes by the door and her overcoat hanging on the coat rack. \(\text{Inferential}\)

d. John is not at home, but he knows that his sister is usually home by 20:00. The time now is 21:00. \(\text{Presumptive}\)

In a language that marks inferential and presumptive statements differently, (6c) and (6d) would generate different markings. Thus, the main difference between inferential and presumptive statements is that the former are logical conclusions based on “evidence” perceived in situ, whereas the latter are logical conclusions based on the speaker’s knowledge about the world and ability to draw logical conclusions from it.

The indirect type of access comprises many different ways of acquiring knowledge, and in some languages the finer distinctions between inferential, presumptive and reportative statements are also reflected in the grammar. But for the majority of languages that grammatically mark indirect access to information, we see the use of single, opaque marker for all types of indirect access. The hearer is then left with the impression that the way this knowledge was acquired is somehow indirect, but without any further specification of how the speaker gathered it (inference, presumption, hearsay).

Lazard (1999) refers to indirect evidentiality as “mediative”, but also includes mirative statements under this term. A similar inclusion of mirativity is also seen in the term indirective used in the tradition of Turkic linguistics.\(^{14}\) Mediative and indirective are thus synonyms that refer to a set of opaque grammatical markers that can signal reportative, inferential, and mirative expressions. For the purposes of this dissertation, however, the term “indirective” solely refers to the types of indirective evidentiality found in table 3.3, primarily reportative and inferential, and not Johanson’s type of indirectivity, which also includes mirative expressions.

Other terms used are non-confirmative and non-first-hand. Within the Russian linguistic tradition we find the term zaglaznost’ ‘out-of-sightness’ and neočevidnyje formy ‘non-obvious forms’.\(^{15}\) In more recent years however, the term kosvennaja zasvidetel’stvovannost’ ‘indirect evidence’ has been widely used (Plungian 2010: 31).

---

\(^{14}\) Johanson seems to be the propagator of this particular term, preferring it over others (Johanson 2000: 62). Also, Comrie (2000: 3) refers to indirectivity as “Johanson’s term”. Aikhenvald (2004: 15) also links the use of this term to Johanson.

\(^{15}\) For examples of these, see Rastorgueva and Kerimova (1964: 71-72) and Dorofeeva (1960: 50), respectively.
3.3.1 Simple systems

The most simple of evidential systems only have two choices, whereas complex ones can have up to six. According to their binary opposition, Aikhenvald (2004:25) categorises the simple systems as follows:

- **A1.** Firsthand and Non-firsthand;
- **A2.** Non-firsthand versus “everything else”;
- **A3.** Reported (or “hearsay”) versus “everything else”;
- **A4.** Sensory evidence and Reported (or “hearsay”);
- **A5.** Auditory (acquired through hearing) versus “everything else”.

In the above classification, the use of “versus” implies that there is a relationship of juxtaposition between the two categories; the grammatical forms used for the first category are of a marked, non-neutral type. This type of juxtaposition, i.e. marked vs unmarked, does not exist in systems that are described with “and” (A1 and A4). In these systems, the two categories are on an equal footing, being neither marked nor neutral.

The main difference between A1 and A2 systems is thus the latter’s pronounced markedness of the non-firsthand forms. In languages with an A2 system, the unmarked forms, which are neutral as to their evidential status, exist as a backdrop to the marked, evidentially non-first hand forms. This marked juxtaposition does not exist in languages with an A1 system.

There is some disagreement, however, as to how one should interpret the unmarked forms. When it comes to Turkish, Aikhenvald (2004:40) adopts the analysis of Johanson (2003:275–276), who argues that the unmarked forms “always exhibits neutral uses” when the speaker “considers the evidential distinction unessential and thus chooses not to use it”. Therefore, use of the unmarked form -DI does not automatically signal that the source of information is direct or visual experience, even though the information may have come through such sources (ibid.:283). It must be noted, however, that this interpretation of the evidential system in Turkish is at odds with the “traditional” view whereby the unmarked form is interpreted as strictly marking direct experience, viz. Kornfilt (1997: 337–338) and Aksu-Koç and Slobin (1986: 165). Aikhenvald (2004:40) acknowledges, however, that it is not uncommon for individual languages to “reinterpret the ‘everything else’ term as a primarily ‘firsthand’ and ‘direct experience’” and thus to develop into A1 systems.

3.3.2 Complex systems

In some languages, we find a more refined system of grammatical evidentiality where the signalling of information source is further differentiated. One such language, which possesses an elaborate system for conveying information source is Tucano, a language spoken in the Amazon. It has four different
verbal forms for expressing visual, auditory, inferential and reported evidentials.

(7) Types of evidentials in Tucano; evidential markers are highlighted ( Adapted from ibid.: 52).

a. diâyî wa’î-re yaha-ámi.
   dog  fish-TOP.NON.A/S steal-REC.P.VIS.3sgnf
   The dog stole the fish (I saw it).  Visual evidential

b. diâyî wa’î-re yaha-ásĩ.
   dog  fish-TOP.NON.A/S steal-REC.P.NONVIS.3sgnf
   The dog stole the fish (I heard it).  Auditory evidential

c. diâyî wa’î-re yaha-ápi.
   dog  fish-TOP.NON.A/S steal-REC.P.INFR.3sgnf
   The dog stole the fish (I inferred it).  Inferential evidential

d. diâyî wa’î-re yaha-ápi’.
   dog  fish-TOP.NON.A/S steal-REC.P.REP.3sgnf
   The dog stole the fish (I was told).  Reported evidential

In the above example, we see how the means of acquiring the information in the utterance is clearly reflected in the morphology of the verb. If we were to translate the various sentences in example (7) into a language that does not mark grammatical evidentiality to the same degree, we would have to resort to lexical means in order to convey the different shades of meaning. The difference is that in English, for example, using lexical means to convey the same kind of sentiment is optional, whereas in Tucano the choice of any one of the evidential categories is obligatory.

In fact, native speakers of languages with grammatical evidentiality who also speak languages lacking grammatical evidentiality are known to complain about the inadequacy and lack of explicitness of these languages compared to their mother tongues. It is also not uncommon that speakers who cannot correctly use these different verbal forms are challenged by the hearers, and in some cases are deemed not worth speaking with, since they are perceived as unreliable (ibid.: 342–343).

3.4 Summary

For the purposes of this dissertation, I will view evidentiality as the grammatical marking of information source. In many languages there is a dovetailing between information source (evidentiality) and the speaker’s level of certainty regarding the information (epistemic modality). Boye (2012) proposes the superordinate category of epistemicity, which subsumes expressions of
degree of certainty and expressions of information source. Evidentiality and
epistemic modality have a common denominator in that they both scope over
propositions, which for linguistic purposes can be defined as meaning units
with a truth value. A fundamental difference between state-of-affairs com-
plements and propositional complements is that evidentiality can only scope
over the latter, and not the former. Moreover, since commands involve states
of affairs, evidentials cannot occur in genuine imperatives (Boye 2012: 194–
195).

As a grammatical category, modality can be defined by virtue of its being
non-factual; i.e. what is being modally expressed is not part of reality at the
time of expression. Moreover, modality can be further bifurcated internally
as regards the presence of an “element of will” in the modal statement. Epi-
stemic modality, which lacks any kind of volitional element, is by definition
non-volitive.

Mirativity can be defined as a grammatical category concerned with mark-
ing propositions as new or surprising. Languages which exhibit evidential
strategies often develop mirative uses for the evidential forms. Tajik is an
example of this, since the Perfect functions as a marker not only for report-
ative and inferential statements, but also for mirative ones. Moreover, as will
be discussed in section 4.1.2, the Tajik “mirative” deviates from the classical
definition in two respects. First, Tajik displays a so-called “3rd person mir-
ative”, in which an omniscient narrator marks the mirative experience of the
main character. Secondly, it appears as if the register II forms in Tajik in
many cases mark personal evaluation (often post-situationally) rather than
unexpected information in situ; for this reason the term evaluative is used in
these cases.
4. Evidentiality in New Persian and Turkic languages

This chapter consists of three parts. In section 4.1 a historical survey of evidentiality in New Persian will be given, which starts from Early New Persian before concentrating on the modern national varieties: Iranian, Afghan, and Tajik Persian.

There is a long history of language contact between New Persian and different instantiations of Turkic languages. The Turkic influence on New Persian is obvious in the lexicon, but the use of evidential forms in New Persian is also often attributed to influence from Turkic languages. For these reasons, section 4.2.2 will outline the main characteristics of evidentiality in Turkic languages. In section 4.2.3, the means of expressing evidentiality in Turkic languages and New Persian are compared and contrasted. Finally in section 4.3, a summary of the whole chapter is given.

4.1 New Persian

In a broad sense, “Persian” denotes the Indo-Iranian language New Persian, which came into existence in the wake of the Arab-Islamic invasion of the Sasanian Empire in the middle of the 7th century. By the year 700 ce, Middle Persian had evolved into New Persian (Windfuhr 1979: 169). The first phase of this language, Early New Persian, was characterised by the use of different scripts (such as Hebrew, Manichean, and Arabic) and a high frequency of dialectal features which informed the literary language. Moreover, Early New Persian had a very fluid verbal system compared to that of Modern Persian, in that the different verbal morphemes combined more freely in Early New Persian (Lazard 1963: 18; Paul 2017). The language of poetry became fairly standardised by the 10th century, whereas the standardisation of the prose language did not occur until the end of the 12th century (Windfuhr 1979: 166).

It is not improbable that many of the so-called archaic features which we find in Early New Persian, especially those produced under the auspices of the Samanid Empire, in fact reflect local peculiarities of the spoken language which were levelled out as Persian spread as a literary language into the Iranian plateau (Lazard 1963: 17). From the 13th century, these dialectal features gave way to a more or less homogeneous standard (ibid.: 22). This does not mean that Persian became forever fixed or standardised after the 13th century. On the contrary, the Persian language has undergone a continuous and
gradual development since its emergence, manifesting different types of style according to place and time (Paul 2017).

In common parlance, “Persian” is often used as shorthand exclusively for the official language of Iran (Utas 2009: 158). In order to distinguish this variety of Persian from New Persian as a general term, it will henceforth be referred to as “Iranian Persian”. The Persian varieties that constitute the official languages of Afghanistan (Dari) and Tajikistan (Tajik) will, apart from these shorter names, also be referred to as “Afghan Persian” and “Tajik Persian”. In this taxonomy, any reference to Persian without a modifier refers to New Persian as a general term, irrespective of contemporary national idiosyncrasies.

The main purpose of this section is to provide an introduction to the evidential status afforded to the Perfect and Perfectoid forms in New Persian. These verb forms can be analysed as constituting a second register corresponding to the Preterite, the Imperfect and the Pluperfect.\footnote{See, for example, Lazard (1985), Phillott (1919), Kasravi ([1938] 1996), and Perry (2000).}

Using the verb *kardan* ‘to do’, the two registers and their respective verb forms are displayed in table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>Preterite:</th>
<th>Perfect:</th>
<th>Imperfect:</th>
<th>Perfectoid Pluperfect:</th>
<th>Pluperfect:</th>
<th>Perfectoid Pluperfect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td><em>kard-∅</em></td>
<td><em>kard-e=ast</em></td>
<td><em>mi-kard-∅</em></td>
<td><em>mi-kard-e=ast</em></td>
<td><em>kard-e bud-∅</em></td>
<td><em>kard-e bud-e=ast</em></td>
</tr>
</tbody>
</table>

Table 4.1. *A rendition of the two registers, based on Lazard (1985: 27).*

The overview in this section will be presented chronologically, in accordance with the different stages of New Persian, which can be roughly divided into four consecutive interconnected periods. (1) Early New Persian, which is characterised by local idiosyncrasies, was written between the 8th and 10th centuries. (2) Classical Persian, written between the 10th and 16th centuries, was a relatively homogeneous literary language. (3) The Post-Classical period lasted from the 17th to the 19th century, during which many belles-lettres were written outside of Iran. (4) The period of National Languages, started in the 20th century and continues to this day (Windfuhr 1979: 166).

Specific verb forms for expressing epistemicity (i.e. evidentiality and epistemic modality) which have developed independently in these national varieties will also be dealt with, namely the *xāt* form in spoken Afghan Persian and the Presumptive in Tajik Persian.

4.1.1 Classical and Post-Classical Persian

Even though the Perfect and the Perfectoid forms have been attested in Classical Persian texts, their evidential status is still a matter of debate. For example, Windfuhr (1979) is of the opinion that the Perfectoid forms in Clas-
sical Persian have a strictly resultative meaning, whereas Kasravi ([1938] 1996:288) has a diametrically opposed view, seeing the Perfectoid forms, and even the Perfect, in both Classical and Modern (Iranian) Persian, as indicating that the proposition is nādide ‘unseen’. In one instance, Kasravi ([1938] 1996:288) gives an example from the Safarnāma of Nāṣer-e Xosraw, a travelogue from the 11th century, in order to prove the evidential nature of the Perfectoid Imperfect.2

Thorough studies on the prevalence and possible function of the indirective forms in Classical Persian have until recently been scarce. Lazard (1963) acknowledges the existence of the Perfectoid forms in texts from the earlier centuries of New Persian, but does not expound on their function. Utas (2000) conducted a study based on five different Classical Persian works3, which are mainly concentrated to the earlier-middle period of Classical Persian. Utas (2000:269) concludes his study by acknowledging that there does not seem to be any regular way of conveying inferential or reportative propositions in Classical Persian, either through the use of verb forms or by lexical means. He goes on to state that the Perfectoid forms we find in Modern Persian are “extremely rare” in Classical Persian, and do not appear to exclusively convey the evidential meanings of indirectiveness afforded to them in modern times (ibid.:269). As valuable a contribution to the field as it is, one cannot draw any far-reaching conclusions from Utas’s cursory study, as the five classical works included in his study are too few in number to fully give a comprehensive picture of the prevalence and function of the Perfect and the Perfectoid forms in Classical Persian. Moreover, gauging the study’s merits is made harder by his not mentioning the amount of pages that were analysed for the purpose of his research.

As explained above, Classical Persian spans over more than six centuries and covers a vast geographical area with varying dialectal differences that influenced the literary language. Utas’s study thus only presents us with a snapshot of how these forms occur and function in Classical Persian.

However, a more variegated picture of the emergence and use of the Perfect and the Perfectoid is presented in Lenepveu-Hotz’s (2014) analysis where she analyses a great body of Classical Persian texts written between the 10th and 15th centuries. The texts are from different genres and geographical locations. Among her ten texts we find commentaries on the Qur’an and the Bible, historical chronicles, and one scientific work (ibid.:27).

The results of Lenepveu-Hotz’s study show that indirective uses of the Perfect and the Perfectoids are mostly found in the more recent texts of the

---

2. For a more thorough presentation of this example, see page 78.
corpus, i.e. in the texts from the 15ᵗʰ to the 16ᵗʰ centuries, namely RA⁴, SX⁵ and TT⁶. But these forms already assume the value of the *passé révolu*⁷ and possibly even that of report by others as early as *Tārix-e Sistān*, which was composed between 1053 and 1063 (Lenepveu-Hotz 2014: 173). Due to the aim of her work, although she duly presents examples of inferential and reportative uses of the register II forms in her corpus, she does not look into the non-uses of these forms. In other words, the reader is presented with the cases in which they are used, but not those where their use could have been warranted by the indirective nature of the proposition in question. It is therefore difficult to draw any general conclusions based on her observations alone.

As for the Perfectoid Pluperfect and the Perfectoid Imperfect, these forms have been attested in texts belonging to the earlier centuries of New Persian.⁸ In the corpus of Lenepveu-Hotz, the Perfectoid Pluperfect appears in TJG⁹, RA, SX and TT. The Perfectoid Imperfect also appears in these same texts, except for RA.

As for Post-Classical Persian, this is an understudied field in general, and especially as regards evidentiality. However, the results of Hashabeiky’s (2021) study on the evidential use of register II forms in twelve Post-Classical Persian (16ᵗʰ–18ᵗʰ centuries) works (mainly historiographies and autobiographies) from Safavid Iran, the Khanate of Bukhara, and Moghul India indicate that these verb forms were systematically used to convey primarily reportative propositions (ibid.: 144). The use and frequency of the Perfect and the Perfectoids is directly related to the type of genre of the work. In works where the authors discuss contemporary events and people, these forms are used very infrequently. Conversely, these forms are more frequent in works that deal more with past events (ibid.: 144–145).

---

⁴. *Rawżat al-aḥbāb fi seyar al-nabi va al-āl va al-āshāb* (“The Garden of Friends in the Path of the Prophet, His Family and His Companions”) is a work written by Amir Jamāl al-Dīn ‘Āṭā’ Allāh b. Fāzī Allāh Ḥūṣaynī Dašṭakī Šīrāzī in 1484, probably in Herat. RA is written in fluid and simple prose, and details the life of Muhammad, his relatives, and his companions (Lenepveu-Hotz 2014: 38).

⁵. *Selsela-nāma-ye xᵛājagān-e naqšband* (“Genealogy of the Naqshbandi Masters”) is a work written by Nur al-Dīn Muhammad b. Ḥūṣayn b. ‘Abd Allāh b. Pīr Ḥūṣayn b. Šams al-Dīn Qazvinī in 1570. The work details the complete genealogy of the Naqshbandi masters and also gives some biographical information. Although the name of the obscure author seems to hint at this piece being written in Qazvin, this is just a conjecture, as nothing is known about this particular author (ibid.: 37–38).

⁶. *Taḏkirat al-tavārix* (“Biographical Collection of Chronograms”) was composed by ‘Abdallah Kābulī between 990/1582 and 1010/1601. The author is most likely from Kabul, making this text an example of South-Eastern Persian (ibid.: 38).

⁷. A French term, also used by Lazard (1985), to denote what in English would be called “the remote past”.

⁸. See, for example, Lazard (1963: 295, 342).

⁹. *Tārix-e jahān-gošā* (“History of the Conquerer”) was written by ʿAlāʾ al-Dīn ʿĀṭā-Malek Jovaynī between 1252 and 1260. Notwithstanding that it is well known that the author is born in western Khurasan, by virtue of his many sojourns in varying places, it is hard to determine in which dialect this text was written (Lenepveu-Hotz 2014: 34–35).
To summarise Lenepveu-Hotz’s study, the indirective use of the Perfect and the Perfectoids is more prevalent in the texts from the 15th and 16th centuries than in the earlier texts. The reportative use starts appearing in the 11th century, prior to that of the inferential, which possibly already existed by the 13th century, and most definitely by the 15th century (Lenepveu-Hotz 2014: 178). Lenepveu-Hotz does mention, however, that the inferential meaning may have emerged as early as the 11th century.10

Given the close contacts between Persian and various Turkic languages, which first began at the end of the 10th century, some scholars have seen evidence of Turkic influence in the indirective evidential forms in New Persian11. In Turkic languages, the first attested usages of the indirective forms hail from the 8th century. If we take the first attestations of the indirective in Persian to have occurred during the 10th to 11th centuries, this would mean that the effect of Turkic on Persian would have been rather immediate.

But we must also bear in mind that there does not necessarily have to be any outside influence for the perfect of a given language to acquire indirective overtones, as described in section 3.2.3. The link between the perfect and the indirective is attested in many languages, especially in those of the Balkans and the Middle East, but also in languages that have not been in contact with Turkic. In fact, “the earliest datable evidence for indirective use of a perfect in Iranian languages” can be found in Old Persian texts (Jügel 2020: 302), long before Turkic and Iranian languages came into close contact with one another. Specifically, the indirective use of the Old Persian Perfect can be seen in the second column of Darius’s Behistun inscriptions. When Darius I recounts the battles he participated in, he uses the Imperfect in the first person plural, but when recounting battles that one of his generals fought for him, he uses the Perfect in the third person plural (ibid.: 302).

According to Lazard (2001: 366), it is only because indirectivity was first spotted in Turkish, Bulgarian and Albanian that it was thought to be an areal phenomenon. Comrie (1976: 110) and later Bybee and Dahl (1989: 73–74) believe that the link between the perfect and the indirective arises through a shift from an action known by its result (the perfect) to an action known by report and inference (the indirective).

According to Lenepveu-Hotz, the emergence of indirective meanings from the Perfect and the Perfectoid forms in Persian can be adequately traced in an internal development, without having to resort to an explanation in which Turkic languages are the cause of indirectivity in Persian. This is not to say

---

10. The reason for this is al-Biruni’s juxtaposition of the Preterite with the Perfectoid Imperfect in a particular passage explaining how Ptolemy calculated the diameters of the moon, sun and earth. When describing how Ptolemy calculated the diameter of the moon, al-Biruni uses the Preterite. However, when al-Biruni is describing how Ptolemy calculated the diameter of the sun by observing eclipses, he uses the Perfectoid Imperfect, as if to imply that this particular piece of information has been reached through inference.

11. See, for example, Windfuhr (2005) and Perry (2005: 485), who make this connection.
that Turkic languages may not have acted as a catalyst or trigger for the emergence of indirectivity in Persian.

4.1.2 National varieties
The national varieties of Persian share a literary canon hailing from Classical Persian, consisting of works from a period which stretches from the 9th to at least the 15th century. With the emergence of modern nation states in the 20th century, however, the varieties of Persian in the capitals Tehran, Kabul and Dushanbe became all the more diversified, as local dialectal differences became more accentuated, and in the case of Tajik, were actively boosted by Soviet language planning. In this manner, new national varieties were formed (Jeremiás 1984:274). The Persian of Afghanistan was officially declared a national language under the name Dari. In like manner, the Persian spoken in Transoxania was dubbed Tajik and made into the official language of Tajikistan. Although the official language of Iran is still called “Persian”, in English there has also been increased use of the endonym “Farsi” (> fārsi ‘Persian’) for specifically Iranian Persian (Spooner 2012: 94–101).

Classical Persian

Afghan Persian

Iranian Persian

Tajik Persian

As a reaction to the peddling of Persian as three distinct national languages, we find scholars who maintain that the differences between these national varieties are only superficial, amounting to no more than the differences of accent one can find within a language (cf. Talattof 2015:216). This position, however, ignores the now existing grammatical differences between the varieties, some of which are possibly re-emerging local archaic forms (Jeremiás 1984).

Among the national varieties, Tajik stands out the most. Some of the idiosyncrasies of Tajik were implemented with the intent of creating a national language that would distinguish the language of Tajikistan from the varieties of Persian spoken in Afghanistan and Iran. For example, the Tajik language was not only refashioned in the Cyrillic alphabet, but was also engineered by Russian linguists to adopt new grammatical norms which accentuated the differences between it and the Persian of Iran and Afghanistan (Roy 2000: 75–76).

But we also have differences that have not been caused by deliberate engineering, but through the natural evolution of language. For example, the
national varieties all differ in the way they express the progressive form. We also see differences between the varieties in the area of epistemic and evidential expressions: Tajik Persian has the Presumptive (Perry 2005: 243–247) and spoken Afghan Persian the xat form (Perry 2002).

**Iranian Persian**

As mentioned above, the Perfectoids have historically played a rather minor role in comparison to their register I counterparts. This is also reflected in the grammars of Iranian Persian from the last two centuries; the Perfectoids are usually not afforded the same kind of attention as the more common verb forms on which they are structurally based. One reason why these verb forms may have been overlooked by grammarians, native and foreign, is that they “represent a deictic category of the verb not expected to exist in the contemporary Persian of Iran” (Windfuhr 1982: 263). That scholars were not expecting to find the Perfectoid forms in more recent Iranian Persian is also corroborated by an observation made in Grundriss der iranischen Philologie in which the German philologist Horn (1898: 154) remarks that the Perfectoid Imperfect, originally an archaic form found in older works such as those of Sa’di, “has just reappeared in more recent times”.

In order to give a brief overview of the differing opinions on the Perfect and the Perfectoid forms in Iranian Persian, a summary will be given of an analysis made by Windfuhr (1982) on this topic. This is followed by a more in-depth analysis of Kasravi’s theories. Lastly, Lazard’s discussions on this topic will be addressed.

**Windfuhr’s study**

In Windfuhr’s (1982) study “The Verbal Category of Inference in Persian”, a total of twelve books on Persian grammar written between 1919 and 1976, foreign and native, were analysed in order to ascertain what value is ascribed to the Perfect and the Perfectoid forms. All of the grammars were published in the 20th century. In chronological order they are:

1. Phillott (1919)
2. Jensen (1931)
3. Panj Ostād (1950)
4. Lambton (1953)
5. Rastorgueva (1953)
6. Lazard (1957)
8. Rubinčík (1963)
9. Pejsikov (1963)
10. Alavi-Lorenz (1967)

12. For example, compare the progressive form in Tajik Persian (Perry 2005: 223–227) with the one in Iranian Persian (Yousef 2018: 227–229; 237).
This list encompasses authors hailing from very different linguistic traditions and backgrounds. Among them, Windfuhr judges Phillott (1919) to be the most detailed and insightful, whereas the grammars of Lazard (1957) and Lambton (1953) have had the most influence on scholarship, and arguably still hold up as reference grammars to the present day. Alavi-Lorenz represents German scholarship, whereas the works of Rastorgueva (1953), Rubinčik (1963) and Pejsikov (1963) represent Soviet scholarship, the latter work being the only one in this collection devoted to the spoken language (detailing the Tehran dialect). The scholarship of native Iranian authors is reflected in the works of Panj Ostād (1950), Faršidvard (1968) and Sādeqi (1976).

Summarising the above-mentioned authors’ views on the Perfect, we see that there are three generally agreed upon functions of this form, the first two of which are well-known universal functions of it.

1. All the grammars agree that the Perfect denotes a past action continuing into the present.
2. Phillott, Lambton, Lazard: The Perfect expresses completed action in the remote past.
3. Lambton, Lazard, Pejsikov: The Perfect can also be used for the definite future; i.e. it is used for an action completed at a future point in time.

Drawing on the last function above, that of definite future, Sādeqi claims that the Perfect can have the function of “conjecture”, “assumption” and “absence of speaker”. Windfuhr (1982: 277), however, prefers to view “absence of speaker” as the defining trait of the sentences that Sādeqi lists as examples of “conjecture” and “assumption”.

Phillott, too, ascribes the quality of “absence of speaker” to the Perfect. By giving a sentence as in (1), Phillott (1919: 524) remarks that the use of the Perfect in (1) conveys the meaning that “the city is one of the ‘has beens,’ before the speaker’s time” whereas the use of the Preterite would have signified that the speaker “was in B. Abbas when it was an important place”.

(1) bandar-e ʿabbās ʾāshr=e moṭabar=i bud-e=ast
    harbour=ez ʿAbbas city=ez important=ez be.pst-pp=cop.npst.3sg
    Bandar-i ʿAbbas was once an important place. (Adapted from ibid.: 524)

Phillott (1919: 525) makes another similar comparison between the Perfect and Preterite forms of the copula budan, this time writing that the use of the

13. Windfuhr (1982) sees Sādeqi’s interpretation of the Perfect as having to do with “absence of speaker” as possibly being influenced by the interpretations of Phillott (1919).
Preterite (*bud-∅*) instead of the Perfect (*bud-e=ast*) would “signify that the writer or speaker saw what he is describing”.

Still writing about the Perfect, Phillott (1919: 524–525) notes a tendency for contemporary Persian to employ the Perfect where the Preterite would have been used in Classical Persian. It is a bit unclear what Phillott actually means here. Windfuhr (1982: 277) interprets Phillott as stating that the Preterite was used for “absence of speaker” in Classical Persian, whereas in contemporary Persian this would be marked with the Perfect. In any case, Phillott (1919: 525) goes on to write that the use of the Perfect in (2b) “clearly indicates that the writer was not present”.

(2) Comparison between Classical and Contemporary Persian. (Adapted from ibid.: 525)

a. peyğambar farmud-∅  
   prophet say.pst-3sg  
   The Prophet (allegedly said).  
   Classical

b. peyğambar farmud-e=ast  
   prophet say.pst-pp=cop.npst.3sg  
   The Prophet (allegedly) said.  
   Contemporary

In the context of the Perfect it is interesting to note that the two examples that Phillott provides to show the quality of “absence of speaker” both employ the Perfect of the verb *budan* ‘to be’. This is probably not because other verbs in the Perfect cannot have indirective meanings, as exemplified in (2), but rather that *budan*, being a copula, has a stronger markedness when used in its Perfect form, as compared to other verbs.

Most of the scholars referred to in Windfuhr’s paper analyse the Perfectoid Imperfect as carrying perfective and imperfective meanings at the same time, making it “a rare form of the present perfect expressing both durative/habitual and complete action” (Windfuhr 1982: 272). Farşidvard, Sādeqi and Phillott also see shades of ambiguity, uncertainty or non-witnessedness in this form. Farşidvard, for example, regards *ebhâm* ‘ambiguity’ as one of the basic functions of the Perfectoid Imperfect.

According to Windfuhr (1982: 275), Phillott gives the most detailed analysis of the Perfectoid Imperfect of all the above scholars. Phillott (1919: 536) remarks that this form is “rarely if ever used in classical Persian” and that it is “rare even in modern Persian”. Phillott (1919: 536) is of the opinion that the Perfectoid Imperfect acts like the Perfect in so far as it signifies that the speaker was not present, but it is imperfective rather than perfective. Moreover, he makes an important distinction between the markedness of the Imperfect vs the Perfectoid Imperfect, remarking that the Perfectoid Imperfect “signifies that the speaker was not present on the occasions mentioned.
or has no personal knowledge of the facts”, whereas the Imperfect “leaves these points doubtful” (Phillott 1919: 519). In other words, the Perfectoid Imperfect, according to Phillott, is a marked form, while the Imperfect is the default form when the speaker does not want to signal the evidential status of the proposition, not unlike Johanson’s analysis of the indirective forms in Turkish, as discussed in section 3.3.1.

As for the Perfectoid Pluperfect, according to Windfuhr (1982:269), it is more “marginal” than the Perfectoid Imperfect. There are two major lines of interpretation of the Perfectoid Pluperfect. Some, like Sādeqi and Lazard, prefer to interpret is as expressing actions in the remote past. Phillott, on the other hand, analyses the Perfectoid Imperfect as a way to express epistemic modality. Under the heading “The Presumptive Past Tense”, Phillott (1919:522) gives two sentences which are both “modern colloquial”. The first one (3a) employs the Pluperfect, prefaced by the adverb albatte ‘surely’, while the second one (3b), being more “vulgar”, dispenses with the adverb and instead solely relies on having the verb in the Perfectoid Pluperfect to convey the same type of assumption.

(3)  
Comparison between the Pluperfect and the Perfectoid Pluperfect. (Adapted from ibid.: 522)

a.  
albatte gonāh=i kard-e bud-∅ ke sezā=(y)aš=rā
surely sin=IND do.PST-PP COP.PST-3SG CLM punishment=3SG=DO
yaft-∅
find.PST-3SG
He must’ve committed some fault to be punished.

b.  
gonāh=i kard-e bud-e=ast ke sezā=(y)aš=rā
sin=IND do.PST-PP be.PST-PP=COP.NPST.3SG CLM punishment=3SG=DO
yaft-∅
find.PST-3SG
He must’ve committed some fault to be punished.

Windfuhr is of the opinion that Phillott’s interpretation of kard-e bud-e=ast is to be preferred over those made by Lazard and Sādeqi, since it can fit both his own examples and those of Sādeqi and Lazard, but not vice versa. The context of Phillott’s examples “clearly points to the function of assumption and excludes an interpretation of remote past”, whereas the examples of Sādeqi and Lazard, which are all gleaned from historical texts, can be reanalysed as pertaining to assumption or conclusion instead of a purported “remote” past (Windfuhr 1982:271).

Of all the scholars in Windfuhr’s study, Phillott is the one who most consistently affords shades of indirectivity to the functions of the Perfect and Perfectoid forms. Phillott seems to suggest that the indirective use of these
forms is feature of a recent, and in some cases, vulgar type of Persian. First, he posits that Classical Persian would favour the Preterite instead of the Perfect in such a sentence as “The prophet said that...”. Second, he writes that the use of the Perfectoid Pluperfect is a “vulgar” equivalent to a sentence prefaced with an epistemic adverb but with the verb in the Pluperfect.

The views of Kasravi
The Iranian Ahmad Kasravi (1890–1946) was an influential social thinker and pioneer of Iranian linguistics. Although he was an Iranian Azeri, his attitude towards his own native tongue was that it should be eradicated in favour of Persian. Throughout his life he consciously worked to develop and employ a purist form of Persian that did away with words of foreign origin and sought to revive old verb forms. In one of his essays, Kasravi expounds on the Persian verb system, focussing specifically on the function of the Perfect and the Perfectoid forms (Kasravi [1938] 1996). Kasravi mentions that the various and differing verb forms in Persian also exist in Azeri, albeit with a slight difference in usage (“bā andak-i jodāyi”). Although he wishes that Azeri (and all the other languages of Iran except Persian) will cease to exist (“mā mi-x’hāh-im az miān be-rav-ad va na-mān-ad”), he thinks that providing a table comparing Persian with Azeri would be beneficial for some of his readers (ibid.:280). According to Kasravi, what the Perfect and the Perfectoids have in common is that they are used for propositions that are nādide ‘unseen’.

Kasravi explicitly equates the Perfect and the Perfectoid forms with the verbal inflectional suffix -mIš in Iranian Azeri by juxtaposing these forms in a table using the third person singular of the verb “to write” – neveštan in Persian and yazmaq in Azeri. A shorter version which only takes into account the “unseen” forms is given below in table 4.2.

Table 4.2. The “unseen” forms of Azeri and Persian, as presented by Kasravi ([1938] 1996:281). Kasravi’s terms are given in the footnotes together with English translations.

<table>
<thead>
<tr>
<th></th>
<th>Iranian Azeri</th>
<th>Persian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect</td>
<td>yaz-miš / yaz-up</td>
<td>nevešt-e=ast</td>
</tr>
<tr>
<td>Perfectoid Imperfect</td>
<td>yaz-ar-miš</td>
<td>mi-nevešt-e=ast</td>
</tr>
<tr>
<td>Perfectoid Pluperfect</td>
<td>yaz-miš-miš</td>
<td>nevešt-e bud-e=ast</td>
</tr>
</tbody>
</table>

1 “The unseen past” – goḏašt-e=ye nā-did-e
2 “The unseen continuous past” – goḏašte=ye hamānzamāni=ye nā-did-e
3 “The unseen past past” – goḏašte=ye goḏašte=ye nā-did-e.
4 Corrected from nevešt-e bud-∅.

In this context, it is also worth mentioning the Soviet-Azerbaijani scholar Alizade, who also observed this category in Persian. In an article from 1965 he makes a similar juxtaposition of verb forms as Kasravi, albeit without including the Perfect in his analysis. As for the Perfectoids, he characterises them as “rare and literary only” (Windfuhr 1982: 283).
Kasravi holds that certain people who have misunderstood the Perfect have called it *godašte=ye nazdik / māzi=ye qarib* 'the near past', but that this term does not carry any meaning ("*ma’nā=ye bā-x’od na-dār-ad*"). Given that the Perfect in Persian is primarily used with a resultative meaning, it is possible that Kasravi’s declaration that the term “the near past” is meaningless is due to a misunderstanding on his part. The adjective “near” would in this context thus refer the “current relevance” of the Perfect, and not with proximity in time.

Kasravi instead prefers to describe the Perfect as the “unseen” past rather than as “near” past, describing the form as “more often used in cases where things have occurred but the speaker has not seen them with the eyes” ("*in gune=rā biš-tar dar jā=yi mi-āvar-and ke kārhā=yi ruy dād-e vali soxangu ānhā=rā bā dide na-did-e=ast*") (Kasravi [1938] 1996:274).

To further illustrate this point, Kasravi presents the reader with a contrastive pair of sentences, one of which is constructed with the Preterite and the other with the Perfect. According to Kasravi, the use of the Perfect in (4a) signals that the speaker was not at home during the described event. In like manner, the use of the Preterite in (4b) will impart to the hearer that the speaker was at home during the event.

(4) a. *diruz kes=i āmad-e dar=e xāne=ye mā=rā zad-e o ṣad riyāl pul dād-e o raft-e*

    yesterday someone=IND come.pst-pp door=ez house=ez we=do
    hit.pst-pp and 100 rial money give.pst-pp and go.pst-pp

    Yesterday someone (allegedly) came to our house, knocked on the door and handed over 100 riyal in money and went away.

b. *diruz kes=i āmad-∅ dar=e xāne=ye mā=rā zad-∅ o ṣad riyāl pul dād-∅ o raft-∅*

    yesterday someone=IND come.pst-3sg door=ez house=ez we=do
    hit.pst-3sg and 100 rial money give.pst-3sg and go.pst-3sg

    Yesterday someone came to our house, knocked on the door and handed over 100 riyal in money and went away. (Adapted from ibid.: 274–275)

Kasravi further emphasises that the difference between *rafte* and *raft* is not one of tense, but of whether the act took place before the eyes of the speaker or not. Kasravi acknowledges that this form is also used in other circumstances, but he is of the opinion that it “initially was not used for anything but this meaning” ("*dar naxost joz bahr=e in ma’ni na-bud-e=ast*") (ibid.: 275)

Kasravi ([1938] 1996: 282) also adds that sometimes the Perfect is used even in cases where the speaker has seen the event, but the listener has not. In these cases, the Perfect is used pre-emptively out of concern for the listener who has not witnessed the event.
Moreover, he says that sometimes a question is asked in the Perfect, and therefore the answer is also in the Perfect. For example, a teacher may ask his student: dorost=rā nevešt-e=ī? ‘Did you write the right [answer]?’ The student answers: nevešt-e=am ‘I have [written]’.

Furthermore, Kasravi is of the opinion that when something is brought forth as an introductory remark in order to say something else, the Perfect is used, as in example (5). In other words, Kasravi inadvertently acknowledges that the Perfect is used for past actions that have current relevance.

(5) farš-i ke xarid-e=am zud-tar be-foruš-∅ ke rug=IND CLM buy.PST-PP=COP.NPST.1SG quick-COMP IMPV-sell.NPST-2SG CLM na-mān-ad NEG.SUBJ-remain.NPST-3SG
Sell the rug that I have bought as soon as possible so that it will not remain. (ibid.:282)

According to Kasravi, the Perfect is also employed in certain cases when a lot of time has passed since the event. Kasravi gives the example of a sentence uttered in a conversation between two brothers, as can be seen in example (6). Like in example (5), one could argue that once again “current relevance” is in play in the use of the Perfect in example (6), notwithstanding that Kasravi does not mention this.

(6) dah sāl ast pedar=emān dar gozašt-e=ast ten years COP.NPST.3SG father=PCIPL prev.PST-PP=COP.NPST.3SG
Our father passed away ten years ago. (ibid.:282)

A justified question regarding the use of the Perfect for unseen events, which Kasravi ([1938] 1996:282) himself raises and answers, is that if the Perfect is only used for unseen events, then why do history books only use the Preterite, although the historians have not seen most of the events they describe, and most of things they recount are from hundreds and even thousands of years ago. Kasravi’s answer is that since they are retelling what other people have seen, this is not a problem. A problem with this explanation is that it assumes that historians only write down events as told to them by first-hand witnesses, and that these same historians would therefore not record the events with “unseen” verb forms, but instead retain the “seen” verb forms in their own accounts, as if quoting verbatim from the eye-witnesses. In other words, Kasravi acknowledges that not all unseen events are described with “unseen” verb forms. Although he does not state it explicitly, it is clear that in his view, historiography has its own rules for verb forms, and that writing history with “unseen” verb forms would be inappropriate.

Kasravi ([1938] 1996:282–283) writes that if a proposition has not been witnessed with the speaker’s own eyes, but the speaker has acquired the piece of information through thinking (“be andiše dar yāft-e=ast”), the Preter-
ite cannot be used. Instead it must be written in such a way as to make it clear how the information was acquired (“bāyad bā zabān-i nevešt ke dānest-e šav-ad az če rāh be dast āmad-e”). Here, Kasravi is hinting at an inferential function of the Perfect, but he does not elucidate this matter further.

Moving on to the Perfectoid Pluperfect, Kasravi gives an explanation of the different particles that make up the Persian verbs. The function of the past stem bud is to make it “more past” ("goḏašt-e-tar"), the function of the silent suffix -he (-e) is to convey that it is “unseen”, and the function of mi- is to convey “presentness” (“hamānzamāni”). Thus, the Perfectoid Pluperfect functions like the Pluperfect, but with an added unseen meaning (Kasravi [1938] 1996:277–278). In like manner, the Perfectoid Imperfect, is a combination of the mi- particle and the -e particle. In order to illustrate this form, Kasravi ([1938] 1996:288) quotes an example from Nāṣer-e Xosraw’s Safarnāma, where the author writes about the reservoirs of Mecca. By using the Perfectoid Imperfect in (7), the author is signalling that the method of gathering water is reported information, while the reservoirs being empty belong to the category of “seen”.

(7) ba āb=e bārān ke az darra-hā furō mē-āy-ad por to water=ez rain CLM from valley-PL down IMPF-come.NPST-3SG full mē-kard-a=and va dar ān tārīx ke mā ānjā IMPF-do.PST-PP=COP.NPST.3PL and in DEM.DIST date CLM we there bud-ēm tohī bud-and. COP.PST-2PL empty COP.PST-3PL

They would fill them with rainwater that flowed (lit. flows) down from the valleys, and when we were there they were empty. (Xosraw 1956:89)\(^{14}\)

To summarise, Kasravi draws upon his knowledge of his native language, Azeri, to make a point about how the verbal system works in Persian. In his comparison, the forms that are traditionally seen as indirective in Azeri correspond to the Perfect and Perfectoid forms in Persian.

The “unseen” quality of the Perfect, which, along with its ending in -e was grafted onto the Imperfect and the Pluperfect to make the Perfectoid Imperfect and the Perfectoid Imperfect, is the pivotal part of Kasravi’s explanation of the past forms in Persian. In Kasravi’s analysis, the Perfect is used when the speaker has not seen the events with his own eyes, but Kasravi also mentions that sometimes, in order to compensate for the fact that the hearer has not seen the event, the speaker uses the “unseen” forms.\(^{15}\) Ultimately, some of Kasravi’s explanations for the use of the Perfect are so convoluted and far-

---

14. The passage that Kasravi quotes has some slight differences from the one quoted here. The main point of the quotation—the use of the Perfectoid Imperfect—remains the same, however.

15. This type of pre-emptive use of the evidential form also exists in Uzbek (Straughn 2011:158).
fetched that his purported original use case of the Perfect, “the unseen past”, becomes impossible to justify.

The views of Lazard

The most prolific writer on evidentiality in Persian is Gilbert Lazard. Throughout the years he has written on evidentiality as a general concept, and on how this specifically manifests itself in Persian and Tajik in the form of the “mediative” – a term coined by Lazard himself to refer to the different shades of meaning given by the perfect and its derivatives. In one of his articles (Lazard 1985), he presents an overview of the origins and the functions of the Persian forms which he calls “distanced”. The purpose of this section is to summarise and comment on this article.

Lazard groups the Persian past tenses into two parallel paradigms. In group one, we have the simple past form, i.e. the Preterite, which Lazard calls “the aorist” – kard-∅ ‘s/he did’; the Imperfect (durative past) mi-kard-∅ ‘s/he used to do / was doing’; and the Pluperfect kard-e bud-∅ ‘s/he had done’. Specific to the Iranian type of Persian, there is also the Past Progressive - dāšt-∅ mi-kard-∅.

There is also a second register of verb forms which are modelled on the first group and have the perfect participle marker -e appended to them: the Perfect kard-e=ast; the Perfectoid Imperfect mi-kard-e=ast; the Perfectoid Pluperfect kard-e bud-e=ast, called by Lazard passé surcomposé (lit ‘the double-compound past’); and the Perfectoid Past Progressive dāšt-e mi-kard-e=ast. In table 4.3, all of these forms, which make up the two registers, are juxtaposed with their French equivalents, as provided by Lazard (1985: 27).

| I   | kard-∅ | mi-kard-∅ | dāšt-∅ | kard-e bud-∅ |
| I   | kard-e=ast | mi-kard-e=ast | dāšt-e mi-kard-e=ast | kard-e bud-e=ast |
| I   | «fit» | «faisait» | «était en train de faire» | «avait fait» |

Lazard agrees with Windfuhr (1982) that the second group of verbs, which Windfuhr calls inferential, express a “distanced past”. This “distance” can be temporal, i.e. with the events situated in a remote or distant past, but they can also be indirect in the sense that the speaker does not have personal experience of them. Lazard classifies the verbs of the second group into four types: (a) facts reported by others (“faits rapportés par autrui”); (b) inference proper (“inférence proprement dite”); (c) remote past (“passé révolu”); and (d) retrospective (“rétrospective”) (Lazard 1985: 28).

Lazard posits that the quality of “distance” is a unifying trait of the forms in register II, even though outwardly they differ from one another. In the passé
révolu, the distance is temporal, while in the case of inference, the distance lies in the fact that the speaker did not experience the events directly. In the case of the reportative, it is the relaying of someone else’s information that creates the distance between the speaker and the event. It is not entirely clear how the retrospective function manifests distance. Lazard notes that this category was added for statements that cannot be placed in the other three categories, but unfortunately he does not provide any definition of it or examples of its usage. He does, however, state that it “takes stock of past events” (“[i] fait le bilan d’événements passés”) (Lazard 1985: 39).

Lazard is of the opinion that these four usages derive their value from the “proper” function of the perfect, i.e. its resultativeness. According to the classical definition, the perfect describes a present state resulting from a process in the past. It is a semantically complex form which therefore involves both the present and the past and thus lends itself to expressing various nuances. In many languages it has more or less completely drifted toward becoming a preterite\(^\text{16}\). Lazard is of the opinion that, in Persian, we have the aorist acting as the Preterite and still being very much alive, whereas the Perfect has taken on the meanings of inference and distanced past (ibid.: 38).

Thus, in the words of Lazard, we have a second register mirroring that of the first, but with the added quality of “distance”. As mentioned above, register II constitutes a parallel system to register I. The registers differ only in that the register II forms have an added function of conveying various degrees of “distance”—the corresponding pairs are of equal value regarding their aspect. In register II, the Perfectoid Imperfect and the Perfectoid Pluperfect have acquired the status of proper markers of “distance”. This is not the case with the Perfect, however, as it possesses both the function of resultativeness and that of “distance”.

This dual nature of the Perfect in turn makes its positioning within the verbal system somewhat problematic, as it essentially straddles both registers, being either resultative or “distanced”. According to Lazard, the Perfect has a present tense meaning when it is used in a resultative sense. The only difference between the Perfect and the Present is that the resultative Perfect also carries the meaning of perfectivity (completed process, resulting state). Moreover, Lazard is of the opinion that it is used to express not only the present but also the future. To exemplify this, Lazard takes an example from the fable of the hare and the tortoise, where the hare declares:

\begin{quote}
(8) Tā to be deraxt be-res-i man dowr-e zamin=rā until you.2sg to tree subj-reach.npst-2sg I around=ez earth=do peymud-e=am. traverse.pst-pp=cop.npst.1sg
\end{quote}

By the time you reach the tree, I will have gone around the world.

(After ibid.: 40)

\(\text{16. This happened in Persian: Old Persian: krtam > Middle Persian and New Persian kard}\)
The logical conclusion drawn by Lazard is that the Perfect is on the same level ("sur le même rang") as the Present, and in like manner, the Pluperfect corresponds to the Imperfect. In Lazard’s classification of the Persian verbal system, the Perfect belongs to two subsystems depending on what meaning it conveys. One is a present perfect and the other a past of the second register (distanced). According to Lazard, there is no other choice but to give this form two different positions in the verbal system. The reasoning behind this, according to Lazard, is the existence of the derived forms, namely the Perfectoid Imperfect and the Perfectoid Pluperfect. Since they functioning solely as “distanced” forms, they also highlight the “distanced” quality of the Perfect, from which they have acquired their “distanced” meanings as well as their morphology. The Perfectoids thus have uses that are extensions of the true sense of the Perfect ("extensions du sens propre du parfait") (ibid.: 41).

Regarding the dual nature of the Perfect, Lazard suggests that one is tempted to suggest that there are two homonymic forms, but this supposition would in the end be overly categorical and divisive, since there are cases where it is impossible to tell the resultative meaning apart from the “distanced” one. The Perfect thus inhabits a position with dual meanings, between which a tension is created (ibid.: 41).

Afghan Persian

The Persian of Afghanistan (Dari) is the national variety of Persian that has received the least amount of academic attention as regards the register II forms. In writing this section, the following works describing the grammar of Afghan Persian have been consulted:

1. Bogdanov (1948)
2. Farhadi (1955)
3. Dorofeeva (1960)
4. Glassman (1972)
5. Kiseleva (1985)

It is interesting to note that only Ioannesyan (1999) mentions the existence of the Perfectoid Pluperfect. As for the Perfectoid Imperfect, only Dorofeeva (1960) and Ioannesyan (1999) mention this form, but neither of them affords it any evidential status. Dorofeeva describes the Perfectoid Imperfect as the “continuous form” ("dlitel’naya forma") of the Perfect, which in turn is “the result of the action which is currently available”; i.e. she assigns both of them a resultative meaning. She categorically rejects that these forms are non-witnessed as “they are not found in that particular function which characterises them in Tajik (the non-obvious forms)” (Dorofeeva 1960: 50). The most interesting of her examples of these forms is found in (9), where the
Perfectoid Imperfect (*mē-zist-a-ast*) is used in the proposition that describes the time period when Aristotle lived, but not in the proposition that he was the teacher of Alexander the Great. Dorofeeva does not comment on why one verb form was chosen over another in this particular example. However, if we take into account the fact that the proposition containing the Perfectoid Imperfect is prefaced by a lexical marker for uncertainty (*dar ḥodud=ez ‘circa’*), and the actual time span is off by two years since his death in 322 BC is “widely attested” (Shields 2012:3), this could indicate that the Perfectoid Imperfect is used here in a reportative way to distance the speaker from the factuality of the proposition regarding when he lived. The proposition that he was the teacher of Alexander the Great, however, is not marked in the same way as being “up for debate”.

(9) Arestō moʿallem=e Sekandar=e Maqduni pesar=e Filip
Aristotle teacher=ez Alexander=ez Macedonian son=ez Philip
bud-∅ va dar ḥodud=ez sēṣad o haštād o čahār tā
cop.pst-3sg and in limit.pl=ez three.hundred and eighty and four until
sēṣad o bist o čahār qabl=e milād
three.hundred and twenty and four before=ez birth
mē-zist-a-ast.
impf-live.pst-pp=cop.npst.3sg

Aristotle was the teacher of Alexander the Macedonian, son of Filip, and he lived circa 384–324 BC. (Adapted from Dorofeeva 1960: 73)\(^{17}\)

In Ioannesyan’s (1999: 71) sketch of the Herati dialect of Persian, he writes that the Perfect denotes “an accomplished action” which is closely connected to its result at the moment of speech. As for the Perfectoid Imperfect, Ioannesyan describes it as externally identical to the form called “Continuous Perfect” (“*dlitel’nnyj perfekt*”) in other grammatical descriptions of Persian. He holds, however, that its meaning does not differ from that of the Imperfect.

In one of Ioannesyan’s example sentences, both the Perfectoid Imperfect and the Perfectoid Pluperfect are employed. In the translation of (10) Ioannesyan indicates that the function of these forms is to signal that the events took place “in the distant past” (“*v dalekom prošlom*”). It is not unlikely that the Perfectoid forms in this sentence are reportative, but without having more details about the context in which it was uttered, it is impossible to say.

---

\(^{17}\) Dorofeeva’s example is taken from *Erfān*, 1955, no. 5, p. 51. The transcription provided by Dorofeeva gives the wrong date—374 BC instead of 384 BC—for the birth of Aristotle. Moreover, her Russian translation gives 322 BC as the date of his death, rather than 324 BC, which is stated in the original.
When he became violent, he used to hit his own wife, make her scream and moan. (Adapted from ibid.: 73)

In passing, Kiseleva (1985: 90) employs the Perfectoid Imperfect in one example, but only for the purpose of highlighting how the prefix mē- carries the meaning of continuity or repetitiveness. In other words, she focuses more on the prefix mē- than on its combination with a Perfect stem and its possible evidential meanings. In describing how the Perfect is sometimes used with a present tense meaning, the following example is given:

(11) Inak ma'ljum šod-∅ ba rāh=e ġalat raft-a=i. 
now apparent become.pst-∅ to road=ez mistake go.pst-pp=cop.2sg
It now became apparent that you’ve taken the wrong road. (Adapted from ibid.: 94)

In the light of the fact that the Perfect can function as a marker of inference in other types of Persian, the above sentence could perhaps be analysed as an example of the inferential use of the Perfect, especially given that the Perfect here is prefaced by a lexical marker of inference – “it became apparent”. In other words, Kiseleva may have inadvertently given an example of how the Perfect in Afghan Persian can function as a marker of inference or realisation.

A peculiarity of the spoken language of Kabul and Herat, which is not found in the Persian of Iran and Tajikistan, is the use of the auxiliary verb xᵛāstan in ways which “do not correspond morphologically or syntactically to standard Persian usage” (Perry 2002). According to Perry (2002), this idiom can convey meanings of speculation, inference, and presumption. In other words, it apparently carries both epistemic and evidential meanings.

According to Perry (2002: 321), when the xāt form is used inferentially, it ostensibly always has a future reference:

(12) Emrōz besyār kam abr as barf na-xāt bārid. 
today very little cloud cop.npst.3sg snow neg-will rain.pst
There’s very little cloud cover today—I don’t think it’ll snow. (Adapted from ibid.: 321)

When used in a presumptive sense, the xāt construction can have both a future reference as in (13), and a past reference as in (14).
(13) I dige xod=e me na-xāt raft-a.  
    he EMPH with=ez I NEG-will go.PST-PP  
He won’t go with me. (Adapted from Perry 2002: 323)

(14) Paysē zyātagi=ra tu gereft-a xāt me-raft=i.  
cash.ez extra=DO you.sg take.PST-PP will IMPF-go.PST=COP.2SG  
(I suppose) it was you who always took the extra cash. (Adapted from ibid.: 323)

Dorofeeva (1960: 50–51) describes this form as expressing doubts and assumptions, and states that it can also be used with present and future tense meanings. She sees this form as having emerged as a result of influence from neighbouring Afghan and Indian languages, and points to a similar construction in Pashto. Glassman (1972: 160–163) gives a rather lengthy description of xāt and its uses. Raonaq (2003: 87) also describes this construction, which he calls a “mode conditionell” or “futur hypothétique”. Farhadi (1955: 86) only mentions it in passing, as does Ioannesyan (1999: 75).

Tajik

The Tajik verbal system is highly influenced by Turkic languages, mostly Uzbek (Perry 2005: 484–486; Soper 1987: 38–329). Like the other varieties of Persian, it also has a so-called perfectoid system for denoting indirectivity, i.e. a second register of verbal forms modelled on the Imperfect and Pluperfect. The Tajik system also displays three main characteristics that are not found in the other national varieties of Persian. (1) The Perfectoid Imperfect can also function with non-past reference, making it tense neutral. (2) The Perfect and the Perfectoid verb forms can also be used with mirative functions. (3) Tajik also has a specific verb form expressing doubts and conjectures called the Presumptive.

Table 4.4 summarises the analysis of 12 different grammars and textbooks written in English, Russian, and Tajik. From the summary, we see that there is seemingly unanimous consensus regarding the indirective functions of the Perfect and the Perfectoid forms. Such uniform support for the evidential nature of these forms cannot be found in the literature on the other varieties of Persian, including Classical Persian. Nearly all of the sources listed in table 4.4 ascribe reportative and inferential functions to the Perfect and Perfectoid forms, the exceptions being Baizoyev and Hayward (2004) and Khojayori and Thompson (2009), who do not mention the Perfectoid Imperfect, the Perfectoid Pluperfect or the Presumptive as existing verb forms. Instead, they only mention the Perfect and ascribe to it a reportative function. The lack of further descriptions of the Perfectoid Imperfect and Perfectoid Pluperfect can perhaps be explained by the nature of these two books—they are geared towards new learners of the language, with an emphasis on the most basic elements of the grammar.
Table 4.4. The purported indirective uses of the Perfect (P) and the Perfectoid Forms in Tajik grammars and course books. For the categories of reportative, inferential, and mirative, a check mark indicates that all of the above-mentioned verb forms have been mentioned as having this particular use. As for the Presumptive, a check mark indicates that this form has been mentioned.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arzumanov (1989)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Baizoyev and Hayward (2004)</td>
<td>P</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efimov et al. (1982)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Ido (2005)</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khojayori and Thompson (2009)</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lazard (1956)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Maxadov (1999)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Perry (2005)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Rastorgueva ([1963] 1992)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Rastorgueva and Kerimova (1964)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Rustamov and Gafforov (1985)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Rzehak (1999)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Seménova (2008)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

If we disregard these two books, all of the remaining authors describe the Perfectoid forms and agree that (1) the Perfect and the Perfectoid forms can be used with reportative and inferential meanings, and (2) the Presumptive is a verb form for expressing various degrees of doubt and conjecture. As for mirative uses of the Perfect and Perfectoid forms, only Rastorgueva and Kerimova (1964), Lazard (1956), and Perry (2005) mention this function.

The Perfect

Perry gives two main functions of the Perfect in Tajik, those of resultativeness and indirectiveness. According to Perry (2005: 218), the resultative function of the Perfect can usually be translated by the English Perfect, which it closely resembles, except for the fact that it can function together with specific time references:

\[ (15) \quad \text{Man dirūz ba rafiq=am xat navišt-a=am.} \]
\[ \text{I yesterday to friend=PC.1SG letter write.pst-pp=cop.npst.1SG} \]
\[ \text{Yesterday I sent (lit. have sent) a letter to my friend. (After ibid.: 218)} \]

18. In fact, Perry uses the term “Non-Witnessed functions”, which includes reportative, inferential, and mirative uses of the Perfect.
Colloquial Tajik also displays a variant of the Perfect which ends in -agī and omits the personal ending. Its function, according to Perry (2005: 218), is “invariably that of a resultative”.19

The second function of the Perfect, that of expressing non-resultativeness (i.e. reportative, inferential and mirative functions), is shared with the Perfectoids. Reportative statements with the Perfect can be prefaced with lexical markers like “they say that” or “I’ve heard that”, as can be seen in (16) and (17).

(16) Yakbora ḡoib šud-∅; aknun me-šunav-am, ki ba suddenly absent become.PST-3SG now IMPF-hear.NPST-1SG CLM to Dara=i Nihon raft-a=ast. Dara=ez Nihon go.PST-PP=COP.NPST.3SG He disappeared suddenly; now I hear that he has gone to Darai Nihon. (Adapted from Rastorgueva and Kerimova 1964: 77)

(17) ...lekin kotib xabar dod-∅, ki u=ro... Rašidov jeg but secretary news give.PST-3SG CLM he=DO Rashidov scream zad-a=ast va guft-a=ast, ki ba xud=aš hit.PST-PP=COP.NPST.3SG and say.PST-PP=COP.NPST.3SG CLM to self=PC.3SG kor=i zaruri dor-ad. work=EZ important have.NPST-3SG ...but the secretary reported that Rashidov had called him and said that he had an important matter with him. (Adapted from ibid.: 77)

Reportative Perfects can also be used without any lexical marker. Pinpointing the exact function of the Perfect in these cases is difficult, because, as mentioned above, the main use of the Perfect is resultative. There are, however, some indicators that signal that the Perfect is used in an indirective sense. In (18), the verbs for “to throw” and “to hear” are in the Perfect, but the verb for “to scold” is in the Perfectoid Imperfect, which further emphasises the indirective nature of the segment. Moreover, the narrative style of the text, which begins with the indefinite “one person”, further accentuates the reportative use of the Perfects in this segment.

---

19. In order not to leave out a potentially indirective form from the analysis, both Perfects will be taken into account and treated equally when looking for indirective statements in part II. Also, see section 4.2.3 for similarities with Uzbek.
The inferential Perfect can occur without an explicit lexical marker. In (19), the proposition that they had not denied her request is supposedly in the Perfect because it is based on an observation of her happy state.

(19) Az xursandi=i ū fahmid-an dušvor na-bud-∅, ki from happiness=EZ she understand.PST-INF difficult NEG-COP.PST-3SG CLM xohiš=i vay=ro rad na-kard-a=and. wish=EZ she=DO denial NEG-do.PST-PP=COP.NPST.3PL Judging by (lit. from) her joy it was not hard to guess (lit. understand) that they had not denied her request. (Adapted from ibid.: 77)

In (20), we also have an observation leading to an inference in the Perfect, but this time the connection between the observation and the inference is more explicitly marked by the use of the phrase “it became apparent that”.

(20) Ū dar avval=i gap zad-an=aš, – oča, – gūyon ba she in first=EZ speech hit.PST-INF=PC.3SG – mother – saying to zabon=i mahalli ovoz bar ovard-∅. Az in ma’lum language=EZ local voice PREV bring.PST-3SG from DEM.PROX apparent šud-∅, ki duxtari=yak e az korgar-on=i mahallī become.PST-3SG CLM daughter=EZ one=IND from worker-PL=EZ local bud-a=ast. be.PST-PP=COP.NPST.3SG At the beginning of her speech, she uttered the word “mother” in the local language. From this it was obvious that she was the daughter of one of the local workers. (Adapted from ibid.: 77–78)

Mirative expressions
According to Perry, the Perfect forms of the verbs doštan ‘to have’ and budan ‘to be’ constitute a special case in Tajik, since they are often found with mirative functions and “always refer to the present time, or general truths” when they are used in an indirective sense (Perry 2005: 228). Tajik is the only variant of Persian where the Perfect, usually of budan, can be used with a mirative sense, as is seen in example (21), where the speaker exclaims how delicious
the food is after having just tasted it. In this example, the mirative function is classical in that it is an immediate reaction to an in situ context, as explained in section 3.2.2. Notice how the proposition, despite being in the Perfect, has present tense meaning.

(21) Či xel bomaza bud-a=ast!
what type delicious be.pst-pp=cop.npst.3sg
How delicious it is! (Adapted from Baizoyev and Hayward 2004: 186)

The type of mirativity found in Tajik is described by Perry (2005: 219) as “a fact unexpectedly discovered or newly appreciated as relevant”. In other words, the mirative in Tajik is not only confined to immediate discoveries in situ, but also applies to sentiments which are newly appreciated after the event has taken place. The Tajik mirative thus differs from the original type of mirative, as described by DeLancey (1997) and Aksu-Koç and Slobin (1986). Moreover, it is more appropriate to refer to some types of Tajik “miratives” as being evaluatives, since the meaning they convey is the speaker’s (or main character’s) personal appreciation or evaluation of a fact, not its unexpectedness. Mirative statements and evaluative statements will thus be treated separately in this dissertation. Mirative statements signal new information based on direct experience, while evaluative statements signal personal opinion based on direct experience. In (22), we see how the Perfect is used with present tense meaning to describe a personal evaluation of someone’s character.

(22) Odam=i xub bud-a=ast.
man=ez good be.pst-pp=cop.npst.3sg
He is a good man. (Adapted from Perry 2005: 228)²⁰

In (23), the same kind of personal judgement is visible in a scenario where a person is being scolded by his former superior. Of course, one could argue that the Perfect is used here because the proposition is reportative, or inferential. But the prevalence of the Perfect of budan in conjunction with very negative or positive personal traits indicates a usage whereby personal opinion is marked to make the statement more forceful.

²⁰ Note that Perry gives the translation for this sentence as “He’s a good man (as it turns out)” which may lead the reader to believe that this proposition is based on an inference. However, Perry (2005: 228) lists this sentence as an example of how the Perfect of budan can be used in “miratives and gnomic statements”. 

88
A closely related phenomenon to the marking of information as “newly appreciated as relevant” is the use of register II forms after epistemic verbs such as “to hear” and “to be informed” and epistemic phrases such as “it became apparent that” (ibid.: 228). In fact, this is a type of “3rd person mirativity”, as explained in section 3.2.2.

Incidentally, the use of register II for marking surprise or new information is identical to the use we find in early Classical Persian texts, such as (25), which is taken from Tarjoma-ye Tafsiri-ye Ṭabari, a text from the 10th century.22

Perry does not give an explicit example of the gnomic, or general truth, condition. The closest sentence to fit this category among the examples he gives is (26). However, it could also be argued that the sentiment in example (26) is not a general truth, but a very personal view of a general truth, and that

21. Perry gives the translation to this one as “Oho! So (I hear) you’re a troublemaker, a tattle-tale!”, which makes the reader believe that the Perfect is used here in a reportative sense, which it of course can be. However, Perry (2005: 228) lists this sentence as an example of how the Perfect of budan can be used in “miratives and gnomic statements”.

22. The text is actually a Persian translation from Arabic commissioned by the Samanid ruler Mansūr b. Nuḥ, who gathered scholars from all over Transoxania to complete the task (Lazard 1963: 42–43).
the use of the Perfect of *budan* here is in fact of the same type we saw in (25); i.e. its use is evaluative.

(26) In navozandāgi xudd=i mullogī barin *bud-a=ast*. Playing an instrument is just like reading and writing. (Adapted from Perry 2005: 229)

*Two special use-cases of the Perfect*

In the spoken language one can find two special uses of the Perfect. The Perfect of *budan, bud-a=ast*, can also be used in a quotative way, i.e. to quote someone’s speech verbatim (Rastorgueva and Kerimova 1964: 82). When used in this sense, *bud-a=ast* marks the end of the quoted speech. In (27), the author has also provided the reader with quotation marks in order to further accentuate the quotative nature of the passage. One can however question the author’s inclusion of *bud-a=ast* within the quote, since it is not part of the quoted speech, but rather marks the end of it.

(27) ∅-gūy-∅ ki «yanga=am va Murodxon-ako=m ba IMPV-say.NPST-2SG CLM aunt=PC.1SG and Murodkhon-uncle=PC.3SG to šumo nigaron hast-and, darraV bo duxtar=aton you.PL waiting COP.NPST-3PL immediately with daughter=PC.3PL rasid-a ∅-rav-ed *bud-a=ast*» arrive.PST-PP SUBJ-go.NPST-2PL be.PST-PP=COP.NPST-3SG

Go and say “My aunt and uncle Murodkhon are waiting for you; go there immediately with your daughter”. (Adapted from ibid.: 82)

(28) Pūlod ako, modar=am xūrok tayyor kard-and, ro Pūlod uncle, mother=PC.1SG food ready do.PST-3PL road ∅-š-ed *bud-a=ast*. subj-become.NPST-2PL be.PST-PP=COP.NPST-3SG

Uncle Pūlod, my mother has prepared food, “come and join us” she says. (Adapted from ibid.: 82)

The Perfect of *šudan* (*šud-a=ast*) is used as a fixed phrase meaning “that’s good enough” or “that’s enough”.23

(29) Kor=e kard-a xabar=i salomati=aš ∅-ras=ad, work=IND do.PST-PP news=EZ health=PC.3SG subj-reach.NPST=3SG *šud-a=st* become.PST-PP=COP.NPST-3SG

It’s enough if news of his health comes in any way. (Adapted from ibid.: 83)

---

23. The spelling *šud-a=st*, as seen in the example, is probably a colloquial variant of *šud-a=ast*. 

90
The same type of expression also exists in Uzbek, modelled on the Uzbek *bo'l-* 'to be; to become'. It is morphologically identical to its Tajik counterpart: *bo'p-* (>bo'l-ib-di) (Soper 1987: 238).

**The Perfectoid Imperfect**

Two things set the Perfectoid Imperfect apart from the Perfect. (1) Though modelled on the Imperfect, this verb form is indifferent to time and can express "habitual or iterative actions in present, past or future time, or acts of general validity" (Perry 2005: 229). (2) Whereas the Perfect can have resultative readings, the Perfectoid Imperfect is used solely as an indirective.

As can be surmised from its morphology, the primary use of the Perfectoid Imperfect is for indirective habitual or iterative actions performed in the past. This is seen in (30), where the habitual actions are rendered in the Perfectoid Imperfect to signal that the accounts are reportative.

(30) Dar zavod-ho=i paxta... kor me-kard-a=ast va har in factory-PL=EZ cotton work IMPF-dopst-PP=COP.NPST.3SG and every sol tobston ba qišloq=amon bar me-gašt-a=ast.
year summer to village=PC.1PL Prev IMPF-turn.pst-PP=COP.NPST.3SG
He (is said to have) worked in the cotton mills and (to have) returned every year in summer to our village. (Adapted from ibid.: 230–231)

The use of the Perfectoid Imperfect with present and future meanings is an innovation of Tajik, which cannot be found in the other varieties of Persian. In (31) and (32), the Perfectoid Imperfect is used with present and future tense meanings, respectively.

(31) Hozir dar kadom kolxoz dehqoni me-kard-a=ast.
now in some kolkhoz farming IMPF-do.pst-PP=COP.NPST.3SG
Now (I gather) he’s working as a peasant on some collective farm. (Adapted from ibid.: 229–230)

(32) Odam-on me-gūy=and, ki Oqpodšoh bača-ho=i mo=ro ba jang people-PL IMPF-say.NPST=3PL CLM Oqpodshoh child-PL=EZ we=DO to war me-firistod-a=ast.
IMPF-send.pst-PP=COP.NPST.3SG
People say that Oqpodshoh will send our children to the war. (Adapted from ibid.: 230)

As mentioned above, Perry also considers this form to be used in statements of general validity. However, as argued above for the case of gnomic statements expressed with the Perfect, one can argue that an indirective form is used in cases such as (33), not because the statement holds general validity, but rather because it is a personal evaluation or judgement, albeit concerning a "general" phenomenon.
Xud=aton fikr ∅-kun=ed, umr=i odamizod misl=i ob=i self=PC.3PL thought subj-do.NPST=2PL life=ez human like=ez water=ez jüy guzašt-a raft-an me-girift-a=ast u ammo rivulet pass.pst-PP go.pst-INF IMPF-take.pst-PP=COP.NPST.3SG and but xud=i odam bexabar me-mond-a=ast. self=ez person unaware IMPF-remain.pst-PP=COP.NPST.3SG

Just think about it: a person’s life is flowing by like a river, but the person himself / herself is unaware of it. (Adapted from Perry 2005: 230)

There is also a Subjunctive Perfectoid Imperfect, which replaces the Perfectoid Imperfect in constructions requiring the Subjunctive. Just like the Perfectoid Imperfect, it is indifferent to tense (ibid.: 238). In (34), the implied conditional clause takes the Subjunctive Perfectoid Imperfect. It is in the subjunctive because it is a conditional clause, whereas the Perfectoid Imperfect is used because the proposition is based on an inference.

(34) Xub, šumo na-me-xūrd-a ∅-boš-ed, man good you.2PL NEG-IMPF-eat.pst-PP subj=COP.NPST-2PL I ∅-xūr-am. subj-eat.npst-1SG
Well, if (as it seems) you’re not going to eat, I’ll eat. (Adapted from ibid.: 237–238)

The Perfectoid Pluperfect

Just as the Perfectoid Imperfect is an indirective counterpart to the Imperfect, so too the Perfectoid Pluperfect is an indirective counterpart to the Pluperfect. In (35), a son is reporting the statements of his father. While the second statement is reportative and marked as such by the use of the Perfectoid Pluperfect, the first statement is in the Perfect, ostensibly to mark its current relevance.

(35) Padar=am xat u savod došt-a, dar vaqt-ho=i father=PC.1SG writing and literacy have.pst-PP in time-pl=ez javoni=aš čand sol dar madrasa-ho=i Buxoro istiqomat kard-a youth=PC.3SG several year in madrasa-pl=ez Bukhara residency do.pst-PP bud-a=ast. be.pst-PP=COP.NPST.3SG
My father knew (lit. has known) how to read and write; during his youth he had spent some years at the madrasas of Bukhara. (Adapted from ibid.: 232)

Perry describes the example in (36) as an inferential use of the Perfectoid Pluperfect. However, since the inferred proposition is prefaced by fahmid-em ‘we understood’, one could also argue that this is actually an example of how register II forms are used after epistemic verbs to mark realisations, as in example (24).
From what his mother said, we gathered that he had moved to the city a long time ago. (Adapted from ibid.: 232)

**The Perfectoid Past Progressive**

A peculiarity of Tajik is that it employs the Perfect Participle of the verb *istodan* 'to stand' to form progressive forms in the past and non-past. The sentence "He is walking" is thus expressed as *raft-a istod-a=ast*, while the Past Progressive replaces *ast* with *bud-∅*.

Perry (2005: 233) describes the Perfectoid Past Progressive form as the "Non-Witnessed counterpart of the Past and Present Progressive tenses", thus implying that it can function with both past and non-past tense meaning.²⁴ Morphologically, it consists of the Past Progressive (*raft-a istod-a bud-∅*), but with an appended -a making it a Perfectoid (*raft-a istod-a bud-a=ast*).

²⁴ This function is also seen in Uzbek, where the corresponding form -(a)yotgan ekan also functions with past and non-past meanings (Soper 1987: 213).

(37) vay kitob xond-a istod-a bud-a=ast, ki man
    he book read.pst-pp stand.pst-pp be.pst-pp=cop.npst.3sg clm I
dar=ro taq-taq kard-am
    door=do knock-knock do.pst-1sg
He was (evidently) reading a book when I knocked at the door. (Adapted from ibid.: 233)

**The Past and Non-Past Presumptive**

Tajik exhibits a specific verb form for expressing "an unsubstantiated conjecture or assumption" (ibid.: 243). This is a peculiarity of the Northern dialects of Tajik that subsequently became standardised in the written Tajik of the Soviet period. It is not found in Afghan and Iranian Persian. Rzehak (1999: 87) refers to this verb form as the Presumptive ("der Präsumtiv"). Perry (2005: 243) acknowledges this name, but prefers to call it "the Conjectural mood", perhaps in an attempt to stay true to the Tajik term, which also refers to it as a "mood": *siğa-i ehtimoli*, "the probabilitive mood" (Rustamov and Gafurov 1985: 209). Ido (2005: 63) calls it "the Speculative".

According to Perry, what sets the Presumptive apart from the inferential use of the Non-Witnessed forms is that it is "inferential by implication". This means that the proposition presented with the Presumptive is a supposition based on contextual clues. Unlike cases where the Non-Witnessed forms are used, "the evidence is not usually presented or even alluded to in the Conjectural sentence" (Perry 2005: 243). The Past Presumptive is morphologically created by appending -agist to the past tense of the verb, thereby creating
a new past stem onto which personal endings are appended: \textit{raft-a-gist-am}, \\
\textit{raft-a-gist-i}, \textit{raft-a-gist-∅}, etc.

\begin{enumerate}[(38)]
\item \textit{Ū az šahr omad-a-gist-∅.}\\
he from city come.pst-pp-presu-3sg\\
He must have come from the city. (Adapted from Perry 2005:244)
\end{enumerate}

\begin{enumerate}[(39)]
\item \textit{Yagon kor=i ganda kard-a-gist-∅!}\\
some work=ez bad do.pst-pp-presu-3sg\\
He must have done something bad! (Adapted from ibid.:245)
\end{enumerate}

The present-future form of the Presumptive, i.e. the Non-Past Presumptive, is prefixed with a \textit{me-}: \textit{me-kard-a-gist-am}, \textit{me-kard-a-gist-i}, \textit{me-kard-a-gist-∅}, etc. According to Perry (2005:245), this seems to be the most frequently occurring form in Tajik literature. In (40), \textit{me-burd-a-g-em} is a dialectal form of the literary \textit{me-burd-a-gist-em}.

\begin{enumerate}[(40)]
\item \textit{balki dar oyanda hamroh=i ū... zindagī ba sar}\\
maybe in future with=ez he life to head\\
\textit{me-burd-a-g-em.}\\
IMPFBring.pst-pp-presu-1pl\\
Maybe in the future he and I will live our lives together. (Adapted from ibid.:245)
\end{enumerate}

Likewise, the suffix \textit{-gist} can also be appended to the Present Progressive (\textit{kard-a istod-a=ast}) to create a Present Progressive Presumptive: \textit{kard-a istoda-gist-∅}.

\begin{enumerate}[(41)]
\item \textit{ba gumon=i on-ki düst=i mühtaram=i vay holo az}\\
to supposition=ez dem.dist-clm friend=ez respected=ez he now from\\
xob bar xost-a va ċoi nūsid-a istod-a-gist-∅,\\
sleep prev rise.pst-pp and tea drink.pst-pp stand.pst-pp-presu-3sg\\
dar=ro taq-taq zad-∅.\\
door=do knock-knock hit.pst-3sg\\
Presuming that his good friend had already risen and would be drinking tea, he tapped at the door. (Adapted from ibid.:247)
\end{enumerate}

\subsection*{4.2 Turkic influences on evidentiality in New Persian}

The close contacts that historically have existed between New Persian and Turkic languages are commonly cited to explain the emergence of evidentiality in New Persian.\textsuperscript{25} It is not surprising that the emergence of evidentiality in New Persian is sought in its contact with Turkic languages, because there is a strong and old tradition of recognizing that Turkic languages display evidentiality, going back to the 11\textsuperscript{th} century, when Mahmud al-Kāšgāri

first described the verbal inflectional suffix \(-mIš\) in Turkic. Moreover, the same evidential use of this verbal inflectional suffix can be seen in the earliest written remnants of Turkic, namely the Orkhon inscriptions dating from the 8th century (Tekin 1968: 192; Erdal 2004: 272–278).

The purpose of this section is to provide an introduction to the topics of Turkic-Persian language contact, how evidentiality is expressed in Turkic languages, and how this relates to New Persian generally, and Tajik specifically.

4.2.1 History of Turkic-Persian language contact

There has been contact between speakers of Iranian and Turkic languages since pre-Islamic times. The earliest contacts occurred between nomadic groups on the Eurasian steppes (Johanson 2010: 655). The extent of early language contact has not yet been documented, but key religious texts extant in both Turkic and Iranian languages indicate that there was some kind of early oral contact and bilingualism. For example, Buddhist literature of Inner Asia is found in both Sogdian26 and Old Turkic. Moreover, the religious texts of Manichaeism, a gnostic Iranian religion which spread far into Inner Asia and won many adherents from different backgrounds, exist in Old Turkic and the Middle Iranian languages Sogdian, Parthian and Middle Persian.

The Sogdian speakers of Samarkand and its environs were no strangers to Turkic speakers; however, it is not until after the Islamisation of Iran and Central Asia that linguistic convergence27 is noticeable between Iranian and Turkic languages. When Central Asia came under the rule of the Caliphate, Persian ousted Sogdian as the elite language, and the Persian speakers of the region found themselves in constant contact with Turkic speakers “as slaves, immigrants, or conquerors” (Perry 2005: 485). Centuries of close contact have led to Eastern Persian (Tajik) borrowing many Turkic words and also displaying syntactic features characteristic of Turkic. Likewise, Uzbek has lost one of the most distinctive traits of Turkic languages, vowel harmony, while also copying the vocabulary and syntax of Persian (ibid.: 485). Moreover, the sound systems of Tajik and Uzbek have converged almost completely (Perry 2001: 193). The different varieties and dialects of Tajik vary in in terms of how many Turkic elements they have adopted. The Northern dialects display the most pervasive Turkic influence of all. The southern dialects of Tajik, which

26. The Northeastern Middle Iranian language that dominated Transoxania before the advent of Islam. See section 2.1.

27. Convergence is the process whereby languages in contact develop in ways that make them more similar to one another (Thomason 2001: 89). Chronologically, the first feature of language convergence is lexical borrowing, including lexical affixes. This feature is also universally present in language convergence, while structural morphs are situated at the other end of the spectrum of tradability (Perry 2006).
are spoken in the southernmost areas of Tajikistan, near the Afghanistan border, have seen the smallest amount of Turkic influence (Doerfer 1991).

Such is the pervasive effect of Turkic upon Northern Tajik in particular, that the Turcologist Doerfer (1967:57) even called it a Turkic language “in statu nascendi”. Many Turkic features became part of the national language of Tajikistan when speakers of Bukharan Tajik and other Northern dialects were tasked with devising a Tajik Persian national language in the 1930s.

The number of Turkic loan words in literary Tajik is very high and far outnumbers Turkic loan words in Iranian Persian. According to Doerfer (1991), Iranian Persian displays a fairly large number of loanwords from the sphere of state and government, whereas Tajik Persian contains a larger number of loanwords from lower social levels, possibly indicating that the Turkic loanwords in Iranian Persian were transmitted among the upper classes and officers. On the other hand, the influx of Turkic loanwords into Tajik came about through contact with the Uzbek common people.

The Turkic loan words in Tajik can be sorted into two groups. Words belonging to the first group are attested in Classical Persian and are thus quite old. Persian yal ‘mane’ (Tajik yol) belongs to this group and is attested from the 11th century. In the other group we find Turkic loan words that are exclusive to the Tajik variety of Persian, such as ang ‘intelligence’ (ibid.).

The influence of Turkic on Tajik is not confined solely to the vocabulary. Tajik also displays an array of Turkic function words, such as the question particle -mi occurring at the end of questions in colloquial Tajik:

(42) Zardolu na-me-xar-ed=mi?
apricot NEG-IMPF-buy.NPST-2PL=Q
Won’t you buy some apricots? (Adapted from Perry 2005:295)

The Turkic interrogative pronoun kim has been paired with Tajik simple interrogatives to form a series of Turkic-Persian hybrids functioning as indefinite pronouns:

(43) Vay az kim-čī norozī ast.
she from something unhappy COP.NPST:3SG
She’s unhappy about something. (Adapted from ibid.: 130)

Another Turkic feature of Tajik that sets Tajik apart from the other variants of Persian is the use of postpositions. In the Northern dialects one finds pre-

---

28. Although Doerfer is correct in stating that many of the Turkic borrowings in Iranian Persian came about through the influence of “higher” strata of society, such as the domains of warfare and state administration, many of the borrowed military words are now obsolete. Among them we find words such as yuzbāši ‘centurion’ from yuz ‘hundred’ and bāš ‘head’ in a genitive construction, and qolčāq ‘gauntlet’ from qol ‘arm’ and the nominal suffix -CAK. The overwhelming majority of Turkic words used in everyday Iranian Persian bear no connection to the higher strata of society; on the contrary they have lower sociolinguistic status than their Persian counterparts (Perry 2001).
positions like *dar* ‘in’ and *ba* ‘to’ placed after the noun phrase. In literary Tajik the postposition *barin* ‘like, resembling; as if’ has no prepositional prototype and can solely function as a postposition. It corresponds in meaning to the longer attested *ezāfe* constructions *misl=i* and *monand=i*:

(44) Ahmad barin šatranjboz=ro man to hol na-did-am.
    Ahmad like chess.player=DO I until now neg-see.PST-1SG.
    I have never seen a chess player like Ahmad. (Adapted from ibid.: 101)

 Literary Tajik displays several Turkic syntactic structures, often involving non-finite verb forms. In these cases, Tajik usually substitutes the Uzbek participle in *-ib* for the Tajik perfect participle in *-a*. An example of this can be seen in the Turkic styles of quoting speech verbatim as displayed in (45), where the perfect participle *guft-a* marks the end of the quoted speech (ibid.: 323–324).

(45) Xud=i û ki=st? - guft-a man az Rahim Qand
    self=ez he who=NPST.3SG - say.PST-PP I from Rahim Qand
    pursid-am
    ask.PST-1SG
    “Who is he, exactly?” I asked Rahim Qand (“…saying…I asked”) (Adapted from ibid.: 323–324)

For reference, in (46) we see how the Uzbek *deb* functions in the same capacity as the Tajik *gufta* in (45). In Uzbek, *deb* and *deya*, the gerund forms of the verb *demoq* ‘to say’ can mark the end of a direct quote (Bodrogligeti 2003: 594; 1046).

(46) Ota=m men-dan tog’a=ng shahar-dan qachon
    father=PC.1SG I-ABL maternal.uncle=PC.2SG city-ABL when
    kel-di-∅ de-b so’ra-di-∅.
    come-PST-3SG say-CVB ask-PST-3SG
    My father asked me, “When did your uncle come from town?”. (Adapted from ibid.: 1223–1224)

The Turkic influence can also be seen in the way Tajik uses infinitives in nominalisation of sentential complements. In these constructions, the finite verb is changed into the corresponding infinitive:

(47) �akhir az qahr  cî guft-an=aš=ro
    She from indignation what say.PST-INF=PC.3SG=DO
    na-me-don-ad.
    neg-IMPF-know.NPST-3SG.
    She is so furious that she does not know what to say (lit. “from indignation she does not know her-saying-what”). (Adapted from Perry 2005: 313)
A number of progressive constructions are only found in the Southeastern branch of the Turkic languages. What they have in common is that the perfective converb -(i)b attaches to one of the four verbs tur- ‘to stand’, o’rir- ‘to sit’, yur- ‘to walk’, and yot- ‘to lie down’ to create a progressive form (Straughn 2011:78). Due to heavy influence from Uzbek and Uyghur, Tajik also developed a similar construction using the perfect participle of the verbs istodan ‘to stand’ and xoraftan ‘to lie down’ (cf. Uzbek tur- and yot-, respectively) (Johanson 2005:212).

In the Western part of the Persian-speaking domain, Turkic language contact has also been significant. The Seljuks (1037–1194) brought about a massive Turkicization of the Iranian speakers of Southern Caucasus. The rulers of the Safavid dynasty (1501–1722) spoke Turkish in the court, even though they retained Persian as the language of bureaucracy and literature. Foreign visitors to the Safavid court remarked that “spoken Turkish was so common among all classes in Iran as to be the lingua franca” (Perry 2001:194).

The Turkic influence on Iranian Persian can essentially be broken down into three domains: phonology, lexical borrowing (and reborrowing), and syntax. In a process that was completed by the 15ᵗʰ century, the Persian of Western Iran saw the merging of the two phonemes /q/ and /ɣ/. Perry (2001:193) considers this to be a possible result of Turkic influence. The phoneme /q/, represented in writing with the letter qāf (ق), was originally a uvular plosive native to Turkic and Arabic, but not to Persian. The phoneme /ɣ/, however, is a uvular fricative native to Arabic, Turkic and Persian, represented in writing with the letter ġeyn (غ).

The lexical borrowing of Turkic words into Iranian Persian is only visible in a handful of everyday words used in Iranian Persian today. Perry (2001:196) notes that there is a tendency for words of Turkic origin to occupy “a lower sociostylistic register than their Persian counterparts”. For example, the original Persian word āheste ‘slow’ is more formal and literary compared to the Turkic synonym yavāš (originally yāvāš), which is the word mostly used in the spoken register (ibid.:196). Closely related to lexical borrowing is reborrowing, which occurs when a Persian or Perso-Arabic term is borrowed into Turkic and then passes back into Persian. In the field of etymology, a reborrowed word is sometimes called a Rückwanderer ‘a returner’. The word for the Shi’a passion play in contemporary Iranian Persian, taʿziye, originally meant “consolation, condolence” in Arabic. After having passed through Persian in the form of taʿziyat, it gained the meaning of “funeral; Moharram rites” in (Azeri) Turkish while simultaneously transforming its -at ending to the current -e ending. It then passed back into Persian and finally gained its current meaning “Moharram drama” (ibid.: 194–195).

Lastly, the reversal of word order in titles is also seen by Perry as an example, albeit rather subtle, of the effect of Turkic on Iranian Persian. The original (pre-Safavid) way of referring to royalty was to prepose the word šah ‘king’ to the name: Šah Esmāʿīl. But from the end of the Safavid dynasty
in Iran a new standard is set where the title is postposed: Nāder Šāh. In the words of Perry (2001: 196), this change is due to an “unconscious adoption of Turkish syntax”.

4.2.2 Evidentiality in Turkic languages

Despite having varied systems, almost all known former and present stages of the Turkic languages have grammatical means of expressing evidentiality (Johanson 2018: 510). Within the academic study of Turkic languages, and especially the research of Johanson, the kind of evidentiality that is grammatically marked on the verb is usually referred to as “indirective” by virtue of the information source being “indirect” (ibid.: 511). As outlined in section 3.3, this cluster of evidentiality has been referred to by different names in different traditions. For example, within Iranian studies, the same type of indirect evidentiality is commonly referred to as mediative by Lazard (1956, 1999). For Turkic languages, Johanson (2018: 511–512) divides the grammatical marking of indirectiveness into three groups according to the type of information source of the proposition. Whereas the first two sources of information are rather straightforward and transparent—reportative and logical inference – the third one, dubbed “perceptive (or experiential)”, does not lend itself to easy categorisation in evidential terms. According to Johanson (2018: 512), the propositions belonging to this third group are based on “firsthand knowledge, direct sensory perception of the event, or indirect perception on the basis of traces or results.” If we interpret Johanson’s statement about “firsthand knowledge” as referring to conclusions based on previous knowledge, that would indeed make the group indirective. The same goes for propositions based on “indirect perception on the basis of traces or results”, i.e. a type of conclusion / inference based on visible clues. The problem here is that Johanson also includes “direct sensory perception of the event”, and this is incompatible with the supposed indirect quality of this group. In an article from 2003, Johanson provides the reader with an example from Turkish to illustrate the experiential nature of this group:

---

29. Johanson (2018: 511–512) uses the term “reportive”, but unfortunately he also indicates that this type of indirectiveness is also called “quotative”, a term best avoided since it is not evidential, as explained in section 3.3.

30. It must be noted here that some interpretative liberties have been taken in order to present the terms in a unified manner. In fact, Johanson (2018: 511–512) writes that evidentials scope over “the narrated event or its effect”, but since only propositions can be evidential (see section 3.1.2), this term will be used.
(48) Sentence said by somebody in situ (in the state of listening):

İyi çal-iyor-muş.
good play-INTRA-IC
S/he is, as I hear, playing well. (Johanson 2003: 275)

By using the above example to illustrate the third group of indirective usages, it appears that Johanson, in fact, is describing mirative or evaluative expressions.

According to Johanson (2018: 512), the indirective markers of Turkic do not fit into the dichotomy of firsthand and non-firsthand information, since their “primary task is not to express the external origin of the addressee’s knowledge”. Rather, Johanson argues that even though Turkic displays a contrast between the marked indirectives and their unmarked counterparts, this does not, in and of itself, mean that unmarked verbs consistently denote situations that are personally known to the speaker (ibid.: 512).

Two types of evidential markers

Indirectivity in Turkic is indicated by two types of markers. One consists of inflectional markers (suffixes occurring after verbal stems) which are used to create post-terminals. These are ambiguous as to their evidential nature, as they can vacillate between evidential and non-evidential readings. There are three inflectional markers that can signal evidential readings in Turkic: miš, gan and ib-dir.

The second type consists of stable markers of evidentiality taking the form of copula particles (enclitic elements added to nominals), which are ambiguous between past and non-past time reference and are unable to carry stress. The main types are ār-miš and ār-kān (ibid.: 513).

31. The use of the term “addresser” here follows the tradition of Jakobson (1960: 353) wherein “[t]he addresser sends a message to the addressee”.

32. Post-terminality is a term primarily used by Johanson to denote “a view on events from an orientation point posterior to their inherent relevant limit” (Johanson 2006: 172). The meaning of post-terminals is akin to that of perfects in British English or Scandinavian languages; i.e. they refer to past events which have relevance in the present.

33. Both Johanson (2018: 512) and Csató (2000: 37) employ the term “pitch”, even though the term “stress” is used in the main body of literature on Turkish word prosody when referring to word-level prominence (Ipek 2015: 25). However, Underhill (1986: 11) and more recently, Levi (2005), have proposed that Turkish is a pitch-accent language. For a discussion on this issue, and why it is preferable to speak of Turkish as a stress-accent language, see Ipek (2015: 23–27).
Table 4.5. A summary of the two groups of indirective markers in Turkic.

<table>
<thead>
<tr>
<th></th>
<th>miš, gan and ib-dir</th>
<th>ār-miš and ār-kān</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Inflectional markers, post-terminals</td>
<td>Copula particles</td>
</tr>
<tr>
<td>Time reference</td>
<td>Mostly past time</td>
<td>Ambiguous between past and non-past</td>
</tr>
<tr>
<td>Indirective ambiguity</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Stressable</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>

The ability to carry stress is thus one of the distinguishing characteristics of the inflectional markers of evidentiality in Turkic, a quality which is of great importance in cases where the two types of markers are otherwise indistinguishable. In Turkish, for example, the Turkic ār-miš is realised as -(y)-mlş, which in its written form sometimes coincides with the inflectional marker -mlş. There is a difference, however. The simple inflectional marker -mlş in most cases has past time reference and carries the stress. Thus, the Turkish sentence ğülmüş can be analysed either as having the stress on the last syllable, with the meaning “s/he apparently laughed” or as having stress on the first syllable: “it is/was apparently a rose” (ibid.: 513).

Uzbek

This section is largely based on the work of Straughn (2011), whose theories on Turkic evidentiality differ from those of Johanson. First, Straughn prefers to speak in terms of “non-confirmativity” rather than “indirectivity”. Secondly, unlike Johanson, he does not see Johanson’s unmarked forms as being uniformly neutral. For example, the Uzbek past tense in -di is analysed by Straughn as conveying marked confirmativity, whereas Johanson would characterise it as a neutral form.

According to Straughn (2011), the uses of the marked non-confirmative Converbial Past -(i)b signals that the speaker does not vouch for, or confirm, the utterance. Bodrogligeti (2003:691) calls this verb form the “subjective past tense of the indicative mood” and describes it as being used when “the speaker is not quite sure about the actual occurrence of the past action”. Bodrogligeti gives the following four reasons why a speaker may choose to
use this form: the statement about the event is made through “sudden inference”, the statement is a conclusion drawn from “attending circumstances”, the speaker is “wondering about its validity”, or finally, the speaker “simply presents it as a fact known from history”. In the words of Straughn (2011: 157), the main function of the Converbial Past is to express a non-confirmative meaning, i.e. doubt, hearsay (especially reported speech introduced with “according to...”), inference, or admirativity. It is limited in its uses, however, since it inherently and necessarily expresses the past tense (ibid.: 157).

When the Converbial Past combines with the first person marker, it usually denotes unintentional acts. Thus, it is common when speaking of forgetting, falling asleep or other acts which one does not deliberately perform (Straughn 2011: 79). In (50), we see how the speaker uses the Converbial Past -i\(b\) form in the first person singular after realising that he has made a mistake.

\[
\begin{align*}
\text{(49)} & \quad \text{Qoʻshshayev-ni “tez yordam” ol-ib ket-ib-di.} \\
& \quad \text{Qoʻshshayev-acc quick help take-cvb leave-cpst-3sg.} \\
& \quad \text{The motor ambulance took Qoʻshshayev away (I believe). (Adapted from Bodrogligeti 2003: 693)}
\end{align*}
\]

\[
\begin{align*}
\text{error do-cpst-1sg address-acc incorrect write-cvb take-cpst-1sg} \\
& \quad \text{(It seems) I made a mistake. I put down the address incorrectly. (Adapted from Bodrogligeti 2003: 693)}
\end{align*}
\]

Note that whereas qil-ib-man and ol-ib-man in (50) are finite verbs (in the Converbial Past) capable of expressing the above-mentioned indirective meanings, yoz-ib in (50) and ol-ib in (49) are not functioning as evidentials, but as converbial bases for compound verbs.\(^{34}\) The very name Converbial Past refers to its being based on the perfective converb in -i\(b\), the original function of which is to act as a serial verb, devoid of any indirective meanings (cf. the Tajik Perfect):

\[
\begin{align*}
\text{(51)} & \quad \text{Oʻrmon-ga bor-ib oʻtin kes-di-m.} \\
& \quad \text{forest-dat go-cvb firewood cut-pst-1sg} \\
& \quad \text{I went to the woods and cut some firewood. (Adapted from Bodrogligeti 2002: 151)}
\end{align*}
\]

According to Straughn (2011: 62), the Perfect in -gan is unmarked with regard to the speaker’s stance on the information being relayed. In this sense, it stands in an oppositional relationship to the Simple Past in -di, which is used in cases where the speaker wants to mark the confirmative nature of the proposition. Another quality that sets the two past tense forms apart, according to Straughn, is that the Simple Past in -di is more likely to be used in

\(^{34}\) Many of these constructions also exist in Tajik as calques. A case in point is the compound verb in (50) yoz-ib ol-, which also exists in Tajik as navišt-a girift-an (Perry 2005: 468).
cases where the events described are accompanied by greater temporal detail. On the other hand, all things being equal, the Perfect in -gan takes precedence when there is a greater temporal distance to the events described (ibid.: 74–74). Notwithstanding this, two out of the five examples given by Bodrogligeti (2003: 696–697) are time-specific historical facts with the Perfect in -gan, an instance of which is given in the following:

(52) Shoir Muqimiy og’ir kasallik-dan so‘ng 1930 yil 25 may-da vafot poet Muqimiy grave illness-ABL after 1930 year 25 May-LOC death et-gan-∅.
do-PRF-3SG.
The poet Muqimiy died on May 25, 1930 after a grave illness.

Bodrogligeti (2003: 694) calls the verb form in -gan the “Present Perfect” and indicates that although its main purpose is to express “a past action with emphasis on its result at the moment the statement is made” it can also be used to express “a past action known only from indirect information not from direct experience”.

However, if the Perfect in -gan is used in conjunction with the auxiliary clitic -dir, it expresses “probability, doubt, or uncertainty about a past event”, sentiments which can be emphasised by modal particles such as balki ‘maybe; chances are’ and ehtimol ‘probably; could be’ (ibid.: 697).

(53) Yoz-gan-dir-lar.
write-PRF-AC-3PL
They may have written. (Adapted from ibid.: 698)

Table 4.6. Features of Past Tense Forms in Uzbek, according to (Straughn 2011: 82). “Definiteness” denotes temporal detail such as “in the year 1492”, “yesterday”, etc.

<table>
<thead>
<tr>
<th></th>
<th>Confirmativeness</th>
<th>Definiteness</th>
<th>Temporal distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past -di</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Perfect -gan</td>
<td>∅</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Converbial Past -(i)b</td>
<td>–</td>
<td>∅</td>
<td>∅</td>
</tr>
</tbody>
</table>

The copular ekan and emish

The unmarked Perfect in -gan is closely related to the particle ekan which is derived from the perfect *-GAn of the copula *er- (Straughn 2011: 89; Bodrogligeti 2003: 777).

This particle is solely used for marking indirectivity. There are, however, differing views as to the nature of this indirectivity. According to Straughn (2011: 90), it is used either with the intention of signalling a non-firsthand information source, or to convey admirativity. Bodrogligeti (2003: 777), on the other hand, states that it is used when the proposition is either based on
logical inference or on a “subjective appraisal made on the basis of indirect information”.

Unlike the Converbial Past in -(i)b, the particle ekan does not constitute a finite verb form in and of itself. Rather, it can combine with a number of time/tense bases of finite verb and can also constitute the nominal predicate. These qualities of the particle ekan in Uzbek, which also exists as eken in Kazakh, set these two languages apart from other languages in the Eurasian evidentiality belt, such as Balkan Slavic and Turkish, in that their ability to also express present and future indirective tenses breaks their ties to the past tense from which they have sprung, and thereby making the marking of indirectivity independent of tense. Bodrogligeti (2003:778) lists the following uses of the particle ekan. It can act as the nominal predicate when added to nouns or adjectives:

(54) Men xursand ekan-man.
    I    pleased ic-1sg
    Apparently, I am pleased. (Adapted from ibid.: 778)

Together with the perfect participle in -gan, it can create the Indirect Present Perfect:

(55) Yoz-gan ekan-man.
    write-prf ic-1sg
    Apparently, I have written. (Adapted from ibid.: 779)

According to Bodrogligeti (2003:782–785), when the particle ekan is added to the finite forms of the presumptive future tense, it can have four different semantic meanings. Two of these are indirective, inferential as in (56) and reportative as in (57).

(56) Ertaga dala-ga chiq-ar ekan-miz.
    tomorrow field-dat go.out-aor ic-1pl
    (Apparently,) tomorrow we will go out to the fields. (Adapted from ibid.: 782)

(57) Hamkasb-lar-ning aytishicha, Nodir Qodir, hatto ota-si-ga ham
    colleague-pl-gen according.to Nodir Qodir even father-pc.3sg-dat also
    “ukan”la-b gapir-ar ekan-∅.
    call.younger.brother-cvb speak-aor ic-3sg
    According to what his colleagues say, Nodir Qodir would address even
    his own father as “my younger brother”. (Adapted from ibid.: 782)

To the above inferential and reportative uses of ekan, Straughn (2011:158) also adds the categories of “assumption” and “admirativity”. Another use of ekan is to form evidential questions, a type of question that is used to indicate the expectation that the hearer will not have firsthand information for the answer or that the hearer will “base the response on the best possible grounds”
The particle *ekan* can also form rhetorical questions, a function we do not find in Turkish or Macedonian, for example. Straughn proposes that these rhetorical questions should be grouped together with admirativity as emotive use of the language, in line with Jakobson’s (1960) account of the functions of language.

The indirective particle *emish* can also be added to adjectives and nouns to form the nominal predicate:

(58)  
U g’amgin *emish-∅.*  
he sad ic-3sg  
He may be sad. (Adapted from Bodrogligeti 2003:788)

Like *ekan*, it can also follow the time/tense bases of finite verbs to form the Indirect Past (59) and the Indirect Present Perfect (60).

(59)  
Yoz-ar *emish-∅.*  
write-aor ic-3sg  
He probably wrote. (Adapted from ibid.:790)

(60)  
Yoz-gan *emish-∅.*  
write-cvb ic-3sg  
He has probably written. (Adapted from ibid.:791)

According to Straughn, if the speaker wishes to explicitly signal that the proposition is reportative, *emish* may be used, since it is never used for other types of indirectivity. The view that *emish* is solely reportative runs contrary to the general description of *emish* found in many grammars of Uzbek. Bodrogligeti (2003:788), for example, writes that the speaker uses *emish* (whether as nominal predicate or with finite verbs) when “the statement is doubtful, is the result of a hasty inference, lacks authenticity, or simply transmits a historical fact”. However, when Bodrogligeti (2003:789–793) describes the use of *emish* to create the Indirect Past and the Indirect Present Perfect the only function mentioned is its reportative use. Notwithstanding Bodrogligeti’s ambiguous descriptions, Straughn (2011:109) bases his position on Johansson’s (2003:279) claim that the copula marker ār-miš “expresses corresponding reportative meanings” as well as on corroborating findings of his own. Thus, according to Straughn, the proposition in (61) can only have a reportative meaning, whereas Bodrogligeti sees in *emish* the above-mentioned range of functions. Interestingly, Bodrogligeti translates the function of *emish* in these cases with “may be” or “may” in his grammar from 2003 and as “apparently” in his grammar from 2002:
4.2.3 Similarities in the evidential systems (Tajik - Uzbek)

This section aims to present the similarities between the Tajik and Uzbek evidential systems and draws heavily upon the work of Soper (1987). Uzbek equivalents to Tajik forms will be presented rather than vice versa.

Table 4.7. Tajik verb forms and their corresponding Uzbek equivalents, as detailed by Soper (1987). Note that this is not an exhaustive list of possible corresponding verb forms, as the Uzbek verbal system is more elaborate than that of Tajik. The model verb for making the table more concrete is ‘to go’ (raftan, ket-). P. = Perfectoid.

<table>
<thead>
<tr>
<th>Tajik</th>
<th>Uzbek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preterite</td>
<td>raft-∅</td>
</tr>
<tr>
<td>Imperfect</td>
<td>me-raft-∅</td>
</tr>
<tr>
<td>Pluperfect</td>
<td>raft-a bud-∅</td>
</tr>
<tr>
<td></td>
<td>raft-a-gī bud-∅</td>
</tr>
<tr>
<td>Present Progr.</td>
<td>raft-a istod-a=ast</td>
</tr>
<tr>
<td></td>
<td>raft-a istod-a-gī</td>
</tr>
<tr>
<td>Past Progr.</td>
<td>raft-a istod-a bud-∅</td>
</tr>
<tr>
<td></td>
<td>raft-a istod-a-gī bud-∅</td>
</tr>
<tr>
<td>Perfect⁴</td>
<td>raft-a=ast</td>
</tr>
<tr>
<td></td>
<td>raft-a-gī</td>
</tr>
<tr>
<td>P. Imperfect</td>
<td>me-raft-a=ast</td>
</tr>
<tr>
<td></td>
<td>me-raft-a-gī</td>
</tr>
<tr>
<td>P. Pluperfect</td>
<td>raft-a bud-a=ast</td>
</tr>
<tr>
<td></td>
<td>raft-a-gī bud-a=ast</td>
</tr>
<tr>
<td>P. Progr.⁵</td>
<td>raft-a istod-a bud-a=ast</td>
</tr>
<tr>
<td>Past Presumptive</td>
<td>raft-a-gist-∅</td>
</tr>
<tr>
<td>Non-Past Presumptive</td>
<td>me-raft-a-gist-∅</td>
</tr>
</tbody>
</table>

¹ Colloquial form derived from ket-ib edi-∅.
² There is, however, a form that more closely resembles the Tajik morphology: ket-ib tur-ib-di, which means “he constantly goes” as opposed to “he is going” (Bodrogligeti 2003: 739).
³ Colloquial form derived from ket-a-yot-ib edi-∅. ⁴ The Tajik indirective Perfect can also correspond to ekan or emish. ⁵ Despite its morphology, this form is used is with both past and non-past meanings. ⁶ This is the form that corresponds most strictly with the Tajik in terms of morphology. However, the form ket-gan ekan-∅ is also given as a possible equivalent by Soper (1987).
The perfect and the pluperfect

There is a longstanding tradition among the Tajiks themselves of equating the Uzbek converb in -(i)b with the Tajik perfect participle (ibid.: 222). There is also good supporting evidence for making this connection. Soper (1987: 237–238) brings up two examples of corresponding uses where the Tajik Perfect and the Uzbek -(i)b mirror one another. The first is to convey continuous states with verbs that can be both stative and dynamic. In Tajik and Uzbek, verbs such as “to stand”, “to lie down” and “to sit” belong to this category of dynamic-stative verbs (Perry 2005: 221; Sjoberg 1963: 113). The sentence he is standing is thus rendered as vay istod-a=ast in Tajik and u tur-ib-di in Uzbek. The second is their use in forming the colloquial expression šud-a=ast / bo‘-p-tti (→bol-ib-di) with the meaning “it’s enough; OK; alright”.35 The Uzbek stem bo‘l- means “to be, become”. Note that the Uzbek -di in the above expression is the third person singular ending for the Converbial Past / Subjective Past in -(i)b.

Alongside the perfect participle in -a, e.g. raft-a, guft-a, Tajik also displays a second perfect participle ending in -agī, e.g. raft-a-gī, guft-a-gī, which can function predicatively without a personal ending in the colloquial language. In this capacity, it corresponds to the Uzbek Perfect in -gan; for example the sentence “he has left” is rendered as vay raft-a-gī in colloquial Tajik and u ket-gan in Uzbek (Soper 1987: 222–223). As explained in the previous section, the Converbial Past in -(i)b can be used to signal indirectivity, whereas the Perfect in -gan is unmarked for any type of indirectivity. Similarly, Rastorgueva (as quoted in ibid.: 224) holds that colloquial Tajik has evolved in such a way that the predicative use of the perfect participle in -agī lends the sentence a resultative meaning. This is a shift that further accentuates the indirective nature of the Perfect in -a. Perry (2005: 218), undoubtedly inspired by Rastorgueva, also describes the predicative use of -agī as being “invariably that of a resultative”.

We thus have two perfects in both Tajik and Uzbek. One type of perfect, ending in -a in Tajik and -(i)b in Uzbek, is ambiguous as to its indirective status. Another type ends in -agī in Tajik and -gan in Uzbek. The difference between them is that -gan is unmarked as to its indirectivity in Uzbek, whereas its Tajik counterpart allegedly has gone a step further and assumed the function of being a marked resultative.

The correspondence between the Tajik perfect participle marker in -a and the Uzbek gerundive marker in -(i)b, and between the Tajik second perfect in -agī and the Uzbek gerundive marker in -gan, is also seen other forms involving the Tajik perfect participle. For example, the morphology of the Tajik Pluperfect -a bud-∅, adding the past tense to the perfect participle, is also seen in the Uzbek equivalent -(u)v-di-∅, which is derived from the perfect participle -(i)b and the past tense in -edi-∅. However, Soper (1987: 221)

35. For an example of how this is used in Tajik, see page 90.
remarks that the exactly parallel morphological structure of \textit{raft-a bud-∅} and \textit{-\(u\)v-di-∅} appears to be “due to an accident of history rather than to cross-language influence”.

In like manner, the Tajik Pluperfect in -\textit{agī} corresponds to -\textit{gan edi-∅}, which is made up of the (perfect) converb in -\textit{gan} and the definite past in -\textit{edi}. Many scholars see the two Uzbek variants as being semantically interchangeable.\footnote{See Soper (1987:219–220) for the differing views on this issue.} It is interesting to note, however, that in literary Uzbek, -\textit{gan edi-∅} is preferred over -\textit{(u)v-di-∅}, and conversely, -\textit{(u)v-di-∅} is preferred over -\textit{gan edi-∅} in the colloquial language (Soper 1987:220). In Tajik, however, the inverse is true, since the perfect participle in -\textit{agī} is more prevalent in the colloquial language, and plays a minor role in the literary language (Perry 2005:218).

\section*{The Perfectoids}

According to Soper (1987:213), the Perfectoid Imperfect, which in Tajik can have past as well as non-past meaning, corresponds to the Uzbek -\textit{(a)r ekan}, which also carries past and non-past meaning. Example sentences with this verb form in Tajik and Uzbek have already been given on pages 91 and 104, respectively. For the sake of comparing and contrasting, they are also given here:

\begin{enumerate}
\item[62] The sentence “Apparently, we will go”.
\begin{enumerate}
\item Me-raft-a=em.
\begin{tabular}{l}
IMPF-go.PST-PP=COP.NPST.1PL
\end{tabular}
\item Ket-ar \textit{ekan-miz}.
\begin{tabular}{l}
go\text{-AOR IC-1PL}
\end{tabular}
\end{enumerate}
\end{enumerate}

As has been noted above, \textit{ekan} is morphologically derived from the perfect *-\textit{GAn} of the copula *\textit{er-}. With this in mind, the morphological similarities between the two corresponding forms emerge more clearly. Taking the above pair as an example, -\textit{a=em} corresponds to \textit{ekan-miz}, and the Tajik imperfective marker \textit{me-} corresponds to the Uzbek Aorist -\textit{ar}.

The Perfectoid Pluperfect corresponds to -\textit{gan ekan}. Example sentences of this form in Tajik and Uzbek have already been given on pages 92 and 104, respectively. For the sake of comparing and contrasting, they are also given here:

\begin{enumerate}
\item[63] The sentence “Apparently, I had written”.
\begin{enumerate}
\item Navišt-\textit{a} \textit{bud-a=am}.
\begin{tabular}{l}
write.PST-PP be.PST-PP=COP.NPST.1SG
\end{tabular}
\item Yoz-gan \textit{ekan-man}.
\begin{tabular}{l}
write-PRF IC-1SG
\end{tabular}
\end{enumerate}
\end{enumerate}
Soper (1987: 213) tentatively links the Perfectoid Pluperfect in -agī to the Uzbek -gan bolīb. By all accounts, the Perfectoid Imperfect with -agī seems to be synonymous with its counterpart in -a (ibid.: 244). As far as the Uzbek form is concerned, it seems to be rather rare, being rarely described in Uzbek grammars as part of the regular verb paradigm (ibid.: 244). These two forms are given here for the sake of comparing and contrasting:

(64) The sentence “He had gone, it seems”.
   a. Raft-a-gī bud-a=ast.
      write.pst-pp be.pst-pp=cop.npst.3sg
   b. Bor-gan bo'l-ib-di.
      go-prf be-cpst-3sg

The Perfectoid Progressive in Tajik is used with both a past and non-past meaning (Perry 2005: 233). The same kind of tense neutrality is also seen in its Uzbek equivalent -(a)-yot-gan ekan-∅ (Soper 1987: 213). Morphologically, the Uzbek form is based on -(a)-yot-gan-∅, which corresponds to the Tajik Present Progressive in -agī.

The Presumptive
Soper (1987: 242) mentions that Rastorgueva equates the Tajik Presumptive with the Uzbek -gan dir-∅. Morphologically, this pairing makes sense, as both constructions involve a perfect participle (-agī / -gan) and a copula stem (-st / -dir) to which personal endings are added. The parsing of a sentence such as “they may have written” thus exhibits perfect symmetry: yoz-gan-dir-lar / navišt-agī-st-and. While Soper (1987: 242) raises no objection to these similarities, he is of the opinion that -gan ekan-∅ should be added as an Uzbek equivalent of the Tajik Presumptive, since they can be synonymous at times. Note, however, that Soper also suggests that -gan ekan-∅ is an Uzbek variant of the Tajik Perfectoid Pluperfect.

Soper identifies the Tajik non-past Presumptive as a counterpart of the Uzbek verb form in -adigan ekan-∅. The participle -adigan is not described by Bodrogligeti (2003) as specifying a special case in conjunction with ekan, however.

(65) Rossiya-dan ishchi-lar kel-adigan ekan-∅.
    Russia-ABL worker-PL come-prog 1C-3SG
    Workers are (apparently) coming from Russia. (Adapted from Straughn 2011: 115)
4.3 Summary

New Persian exhibits three verb forms which, morphologically, end in the perfect participle marker -a: the Perfect, Perfectoid Imperfect and Perfectoid Pluperfect. Compared to the “mother forms” from which they are derived, the verb forms ending in -a are generally thought to mark the proposition as indirective, i.e. as based on second-hand information or inference. The Perfect can be said to straddle the two registers, as its evidential status is ambiguous; it can either be purely resultative in meaning, or function as a marker of indirectivity.

The function of the register II forms in Early New Persian and Classical Persian has been a subject of debate. Until recently, studies on this subject have been scarce and their results inconclusive. Recent corpus studies, however, have shown that indirective use of the Perfect, the Perfectoid Imperfect, and the Perfectoid Pluperfect is more prevalent in younger texts, i.e. from the 15th to 16th centuries. The reportative use emerged in the 11th century, prior to the inferential, which possibly already existed by the 13th century, but most definitely by the 15th century. However, it is possible that the inferential meaning emerged much earlier than that, i.e. in the 11th century.

The function of the register II verb forms have also been the subject of studies on the national varieties of New Persian, i.e. Iranian, Afghan and Tajik Persian. There is a general consensus among the most influential scholars of Iranian Persian in the 20th century that the register II forms have indirective meaning.

Scholars of Afghan Persian, the national variety that has received the least amount of academic attention as regards the register II verb forms, generally do not see any indirective uses in these verb forms. The examples given by these scholars, however, do not satisfactorily prove that the register II verb forms are devoid of indirective meanings. Rather, they can be interpreted as confirming that these verb forms have indirective meanings in Afghan Persian.

If we take into account that the register II forms carry indirective meaning in Classical Persian, and that they still function in this capacity in the modern national varieties of Iranian and Tajik Persian, it is very unlikely that Afghan Persian does not also display some form of indirective meanings for these verb forms. In light of this, Kiseleva’s categorical rejection of any indirective function associated with the Afghan Perfectoid Imperfect seems more like an ideological remnant of Soviet policy meant to emphasise real and imagined differences between languages spoken within the Soviet Union (Tajik Persian) and outside of it (Afghan Persian).

There is a general consensus among scholars of Tajik Persian that the register II forms carry reportative and inferential meanings. Their mirative use, however, is only mentioned in three of the consulted sources. In a majority of the grammars consulted, the Presumptive forms are also described.
Although one does not need to look to Turkic languages to find reasons for the emergence of indirectivity in New Persian. There is no doubt that Turkic languages have influenced Persian to a great extent, and especially the Tajik variety.

Despite their fundamentally different ways of marking indirective propositions, Tajik and Uzbek exhibit some strikingly similar verb forms. In general, the Tajik Perfect (which is ambiguous as to its resultative / indirective nature) corresponds to the Uzbek -(i)b, which also carries the same qualities of ambiguity. Moreover, the Tajik variant of the Perfect ending in -agī can function predicatively without personal ending, and which by all accounts is resultative, corresponds to the resultative non-indirective Uzbek -gan. Similar correspondences are also found in the formation of other verb forms in Tajik and Uzbek; for instance when the Tajik form contains a perfect participle in -a, its Uzbek equivalent has -(i)b in its place, and conversely, when the Tajik has a perfect participle ending in -agī, the Uzbek has -gan. The close connections between Tajik and Uzbek often lead to a one-to-one correspondence in the morphology of the verb forms, such as in language pairs for the Perfectoid Imperfect, the Perfectoid Pluperfect and the Presumptive.
Part III:
Presentation and analysis of the written sources

In this part of the thesis six different Tajik texts will be analysed with respect to the occurrence and function of, primarily, the Perfect, the Presumptive raft-a-gist-∅, and the Perfectoids (i.e. the Perfectoid Imperfect me-raft-a=ast and the Perfectoid Pluperfect raft-a bud-a=ast).

Table 4.8. Texts used in part II. A page number in **boldface** indicates that the whole text was analysed.

<table>
<thead>
<tr>
<th>Time-frame</th>
<th>Works</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>Ayni ([1928] 1948); Jalil (1949)</td>
<td>28; 41</td>
</tr>
<tr>
<td>Middle</td>
<td>Mirzo ([1989] 2009); Tursun ([1988] 2013)</td>
<td>35; 70</td>
</tr>
<tr>
<td>Recent</td>
<td>Safar (2014); Nazarova (2015a)</td>
<td>112; 148</td>
</tr>
</tbody>
</table>

The Tajik texts analysed in this part are six novels and stories taken from three different time periods: an early (1920s–1940s), a middle (1980s), and a more recent period (2010s). These were chosen because they represent three different stages in the development of Modern Literary Tajik. The texts from the first period represent a time in Tajik history when the focus was on consolidating a new type of Modern Tajik which was written in an adapted Cyrillic script and exhibiting several grammatical idiosyncrasies adopted from the Northern dialects. The second period may be conceived of as an intermediate stage when society was beginning to open up. The authors of this period were all raised with the literature of the first period, but the ongoing Soviet-Afghan war accentuated the fact that the Tajik variety of Persian was surrounded by a greater area spanning multiple countries where varieties of
Persian were spoken. At the same time, the opening up of society due to perestroika allowed for new cultural influences from Afghanistan and Iran (Foltz 2019: 140). The authors of the third period are writing in a society that is vastly different from that of their predecessors. Tajikistan’s emergence as an independent country in the 1990s reduced the influence of Russian, and increased the introduction of more words from Iran and the production and importation of materials written in the Perso-Arabic script (Perry 2005: 489; 491).

For each text, the analysis is conducted as follows. Following a short introduction to the story, the reportative, inferential, evaluative, mirative, and presumptive uses of the Perfect, Perfectoids, and Presumptive are listed. When needed, other categories that do not fit into the above-mentioned ones will also be listed, such as the use of these verb forms for sarcasm, questions, gnomic statements, and proverbs.

There will also be a special section for the use of these verb forms when they occur after epistemic verb and phrases. This is to ascertain what kinds of epistemic verbs and phrases that are more commonly used with register II forms.

Naturally, grouping the sentences in this kind of way is difficult, since it is not always clear what group they belong to. For this reason, the section “After evidential verbs and phrases” will be listed separately from the other categories unless the examples belonging to this category cannot be further analysed as belonging to another category. For example, a sentence introduced with “it is clear that…” could be analysed as either an inference or a mirative statement. A sentence introduced with “he understood that…”, however, will be analysed as mirative since it more clearly emphasises the processing of new information (i.e. mirativity) than the reaching of conclusions from in situ clues (i.e. inference).
5. Analysis of *Ahmadi devband*

5.1 Introduction

Ayni’s story *Ahmadi devband* (‘Ahmad the Exorcist’) revolves around a young man, Ahmad, who challenges the popular superstitious beliefs prevalent in Central Asia at the beginning of the 20th century. The story was first published in a periodical called *Rahbari Doniš* in 1928 (Majnun 2016). Like Ayni’s other stories featuring a young boy as a protagonist, such as *Maktabi kūhna* (‘The old school’) (1934) and *Yatim* (‘The orphan’) (1939), it is a semi-autobiographical work aimed at a younger audience. Even if the setting of the story and the character Ahmad bear many resemblances to the life of the author, Ayni never explicitly states in *Ahmadi devband* that the story is autobiographical. He has confirmed that this is the case, however, in the autobiographical work *Yoddoštho* (‘Reminiscences’) (Ayni [1949] 2009:39), which has also been translated into English (Aĭnī et al. 1998).

This chapter will present the different categories of usage of the register II forms as they are found in *Ahmadi devband*. Passages that contain register shifts (i.e. from register I to register II forms and vice versa) will receive extra attention. The bulk of the register II forms occur in hearsay environments, but the text also exhibits cases where the register II forms are used for reportativity, inference, mirativity, questions, sarcasm, and gnomic statements.

5.1.1 Synopsis and narrative structure

*Ahmadi devband* consists of 28 pages of text and can be divided into four parts. The first part, comprising roughly five pages, describes the village of Soktare and especially its haunted burial ground and adjacent cave. The Khojas1 of this village were famous for being *devbands* ‘exorcists’. With the help of sorcery and incantations devbands could compel jinns and other evil spirits to do their bidding. People suffering from all kind of ailments, mental and physical, would seek their help.

---

1. The Persian word xᵛāje ‘master’ is sometimes transcribed as Xoja, Khwaja, Kjodja in Central Asian contexts. In Central Asia it denotes a patrilineal group and typically emphasises their descent from a holy person, such as the Prophet Muhammad, his family, the first caliphs, or Sufi saints (Malikov 2018:121–122). In a pre-Soviet Central Asian society like the one Ayni describes in *Ahmadi devband*, the Khojas, like members of other holy lineages, enjoyed high social status, fulfilling societal roles such as performing religious services and acting as arbiters. Another important role, which Ayni also mentions in *Ahmadi devband*, is that of healing the sick (ibid.:123).
The author begins the second part by describing how the adults and young people of the village would gather at least once a week to talk, read poetry and tell one another stories to alleviate the boredom of the winter season, when work slowed down due to the cold and snow. It is during one of these meetings that the reader first encounters Ahmad. We are not presented with any introduction to who Ahmad is; instead the story begins in medias res, with Ahmad sitting in an inn with a group of men. The group of men are startled as they hear a shriek from a man who had just left the inn to relieve himself. The men run out and find the man lying unconscious. After coming to, he claims to have seen a jinn. Since Ahmad is the only person in the village who doubts the stories of jinns and demons, he sets out on a quest to prove to the village people that supernatural beings, if they exist, cannot harm humans. He does this by going to the burial ground in the middle of the night. In order to prove that he in fact accomplished what he set out to do, he buries a knife at the cemetery for the village people to uncover the following day.

In the third part, the reader is presented with stories from Ahmad’s childhood which describe the impact his father had on him when it comes to not being afraid of jinns.

In the concluding fourth part, we learn that Ahmad is now studying in Bukhara and reading books that give him a scientific explanation to the mysterious light phenomena that occur at cemeteries. The moral of the story is that one should not frighten children with tales of jinns, since they have an adverse effect on their development and make them anxious.

5.2 Reportative

Despite its brevity, the story makes abundant use of the register II forms, the bulk of which occur in two longer narrative passages detailing hearsay accounts. The first, and largest, of these accounts occurs in the opening pages of the story. The second passage is presented to the reader in part three, when Ahmad’s father tells a story about Ahmad’s great-grandfather using register II forms (Ayni [1928] 1948:73).

Lexical evidential markers such as gūyo ‘it is said’ or other means of indicating that the information is hearsay, such as phrases like “they say that”, are also interspersed throughout these narrative passages.

In the first part of the story, Ayni describes the Soktare village and its surroundings. Up to the point where the sentence in (1) is presented to the reader, all of the propositions have used register I forms. However, Ayni switches to a register II form when describing the possible symbolism of the seven flags that adorn the burial place. The shift to the Perfect (bud-a=ast) serves as a contrastive marker, singling out an indirective statement. The proposition that informs the reader that the flags symbolise “the seven per-
fect Pirs” is clearly marked as hearsay, not by lexical means, but solely by virtue of employing the register II form *bud-a=ast* with this specific piece of information. The preceding and subsequent facts, i.e. that the graveyard is known by the name Haftalam, and that the flags are always hoisted there, are given with register I forms.

(1) On mazor=ro Haftalam niz me-gūy-and, zero bar dem.dist graveyard=do Haftalam also impf-say.npst-3pl because on bolo=i on mazor haft alam=i (yalav=i) qutos-dor ki up=ez dem.dist graveyard seven flag=ez banner=ez yak.tail-having clm alomat=i haft pir=i komil bud-a=ast, hameša bar-po sign=ez seven pir=ez perfect be.pst-pp=cop.npst.3sg always on-foot bud-∅. cop.pst-3sg.

They also call that graveyard Haftalam², since seven flags adorned with yak tails³ were always hoisted there, symbolising the seven perfect Pirs.⁴ (ibid.: 49)

In (2), the author is quoting the people’s speech verbatim. Orthographically, this is indicated by the guillemets (« ») which enclose the cited speech. Grammatically, the author also places the perfect participle of *guftan* ‘to say’ immediately after the quotation to mark the end of it.⁵ It is interesting to note that the quoted statement is not marked with a register II form; i.e. the villagers themselves use *ast*, and not *bud-a=ast*, when talking about the supernatural creatures, and when the author indirectly reports the villagers’ beliefs, such as in (3), he uses register II forms.

(2) «in ḡor xob-goh=i dev-on, nišeman-goh=i pari-(y)on dem.prox cave sleeping-place=ez jinn-pl dwelling-place=ez fairy-pl bošišgoh=i ajina-(g)on va orom-goh=i aždaho=i dwelling-place=ez jinn-pl and resting-place=ez dragon=ez tilism-bon ast» guft-a dar zabon=i mardum šuhrat talisman-guarding be.npst.3sg say.pst-pp in language=ez people fame došt-∅. have.pst-3sg

“This cave is the lair of demons, the dwelling-place of fairies and jinns and the resting place of treasure-guarding dragons”, it was said and known amongst the people. (ibid.: 51-52)

---

². Literally “seven flags”.
³. The original reads *kutos*, a typographical error for *qutos*.
⁴. The word given within brackets (*yalav*) is found in the original text. It is not uncommon for Ayni to pedagogically supply the reader with synonyms directly following a lesser-known word, as in this case, or an explanatory footnote.
⁵. This is a Turkic way of quoting speech, as detailed in section 4.2.1.
Ayni consistently uses register II forms when describing the cave and the lore associated with it. Sometimes, the register II forms are prefaced with *gūyo* ‘it is said’, as in (3). The term *gūyo*, which is found interspersed throughout the passage, serves as a further reminder that the information in this passage is hearsay.

(3) [...] *gūyo* pas az aftod-an=i yak qism=i on [...] it.is.said after from fall.PST-INF=EZ one part=EZ DEM.DIST 
**gor-dev-on=e**, ki dar on jo maskan cave-demon-PL=IND CLM in DEM.DIST place dwelling 
*došt-a=and*, az devband=i on zamon iltimos have.PST-PP=COP.NPST.3PL from exorcist=EZ DEM.DIST time pleading 
*kard-a=and* [...]

(4) Dar ġarbi=i tal čorboġča, sarhavz va miqdor=e jangalzor in west=EZ hill garden pool.of.water and amount=IND forestation hast-∅, ki *gūyo* šab-ho dev-on va pari-(y)on ba on COP.NPST-3SG CLM it.is.said night-PL demon-PL and fairy-PL to DEM.DIST jo bar omad-a havo *me-xūrd-a=and* va bazm place PREV come.PST-PP air IMPF-eat.PST-PP=COP.NPST.3PL and feast 
*me-kard-a=and*. Ovoz=i ġazalxonī va IMPF-do.PST-PP=COP.NPST.3PL sound=EZ poetry.recitation and 
nașmanavozī=i pari-(y)on šab-ho az on jo šunid-a music.making=EZ fairy-PL night-PL from DEM.DIST place hear.PST-PP 
*me-šud-a=ast*. IMPF-become.PST-PP=COP.NPST.3SG

To the west of the hill there is garden, a pool of water, and a patch of forest, where they say the demons and the fairies come out to get some fresh air and feast. The poetry recitation and music making of the fairies could be heard coming from that place at night. (ibid.: 52)

In (5), Ayni goes on to use register I verb forms to describe how people would visit the area during daytime, for pilgrimage and sight-seeing, but no one would dare go there at night. But when Ayni is explaining what would happen if anyone did venture out to this place at night, in (6), he switches to a register II form. Ayni thus makes a clear distinction between things that ac-
tually took place (pilgrimage, sight-seeing) and hearsay (that the night-time visitor would be stricken with a calamity).

(5) Rūzona ba niyat=i ziyorat, yo sayohat har kas ba in by.day to intent=ez pilgrimage or sight-seeing every person to dem.prox mazor va ba nazdik=i in gor omad-a graveyard and to nearness=ez dem.prox cave come.pst-pp ∅-rav-ad ham, šabona kas=e=ro yoro=i on subj-go.npst-3sg even by.night person=ind=do bravery=ez dem.dist na-bud-∅, ki dar in davr u peš qadam-guzor neg-cop.pst-3sg clm in dem.prox turn and before step-place.npst ∅-šav-ad.
subj-become.npst-3sg

Even if people—for pilgrimage or sight-seeing—would visit the burial ground and the vicinity of the cave during the day, no one would dare walk around there at night. (ibid.: 53)

(6) Xususan agar kas=e šabona ba qasd=i ozmoiyiʃ, yo ba Especially if someone=ind by.night to intent=ez test or to sifat=i šakkok=i dar in jo biy-o-yad, albatta ba yak quality=ez doubt in dem.prox place sub-come.npst-3sg surely to one balo girifor me-šud-a=ast.
calamity afflicted impf-become.pst-pp=cop.npst.3sg

Especially if anyone would come to this place as an experiment or full of doubts, he would certainly be stricken with a calamity. (ibid.: 53)

Ayni goes on to describe how people possessed by demons and jinns would come from afar to visit the Khojas. This is recounted with register I forms in (7), but in (8), when discussing the lineage of the Khojas and their ability to heal people, register II forms are used in conjunction with the lexical marker gūyo for the second of the two propositions.

The reason for the change in register is because the author himself witnessed how people would visit the Soktare village, which is not the case for the purported lineage of the Khojas and their healing powers.

(7) Devona šud-a-(g)on, ajina zad-a-(g)on va umuman bemor-on az insane become.pst-pp-pl jinn hit.pst-pp-pl and generally ill-pl from joy-ho=i durdast ba xona-ho=i xoja-(g)on=i Soktare omad-a place-pl=ez remote to house-pl=ez Khoja-pl=ez Soktare come.pst-pp me-xobid-and.
impf-sleep.pst-3pl

Those who were possessed by demons or jinns, and generally sick ones, would come from afar to visit the Khojas and stay the night at their place. (ibid.: 53–54)
In (9), we again see how reported speech is conveyed using register II forms. A character in the story recounts a story which his mother had told him. He uses a register I form for the verb denoting “used to say”, but the reported speech is conveyed through register II forms.

The speaker in (9) continues the story about his mother, but shifts to register I forms in (10) when describing how his mother treated the possessed. We thus learn from the shift to a register I verb form that this information is something he witnessed himself, rather than stories he heard from his mother, as was the case in the previous example.

In (11), Ahmad’s father tells him a story about Ahmad’s great-grandfather. The register II forms are used, which signals that Ahmad’s father did not witness these events himself.
5.3 Inference

After laying the groundwork for the main story by describing the Soktare village and its surroundings, Ayni introduces the character of Ahmad and recounts how he became known as a *devband*.

The events preceding that in (12), is that Ahmad and his friends are sitting inside a tavern during the winter season. One of them goes out of the house to relieve himself. Shortly after going outside, the men hear him scream and they think to themselves that maybe a thief has come. This proposition is expressed using register II forms.

(12) 

> Bo šunid-an=i in ovoz “magar duzd=e yo noxalif=e with hear-INF=EZ DEM.PROX voice Q thief=IND or ruffian=IND omad-a=ast” gū-(y)on ahl=i majlis kord-ho va come.PST-PP=COP.NPST.3SG say.NPST-PRSP people=EZ assembly knife-PL and kaltak-ho=ro gifit-a hama yakbora berun david-and. cudgel-PL=DO take.PST-PP all immediately out run.PST-3PL

Hearing this sound, members of the group said “has a thief or ruffian⁶ come?” and took their knives and cudgels and ran out all at once. (ibid.: 58)

The same group of men mentioned in (12) search the area, and in (13), the author describes how they see their friend lying sprawled on a dung heap with a bull snuffling and snorting over him. The first verb in the subordinate clause, aftod-a=ast, is in the Perfect, which is to be expected since verbs in subordinated clauses following sensory verbs (i.e. “to hear”, “to see” etc.) keep the tense of the direct experience. To exemplify, a sentence such as “I saw that he was happy” would literally be rendered as “I saw that he is happy” in Tajik. However, the second verb, bar omad-a bud-a=ast (the Perfectoid Pluperfect) signals an inference on the part of the spectators: “which (apparently) had broken free from the stable...”.

---

⁶ The word *noxalif* is nonstandard. It is probably a typographical error for *noxalaf* ‘ruffian, degenerate’ (Nazarzoda et al. 2008a: 940).
They saw that he lay fallen on top of a heap of manure, and that a big black bull, which (apparently) had broken free from the stable, was sniffing all over him, going “hu... u... hu... u...” (Ayni [1928] 1948:58)

In (14), we also see how the register II forms are used to signal inferences. The author is still describing the scene in which the men have run outside looking for an assailant. The author argues that if the group of men had seen footprints in the snow, they would have inferred that a thief had come and then escaped. As we can see in the quoted speech given by Ayni, they would have expressed the inferred information using the Perfect.

In (15), the reader is presented with yet another register II form for marking an inferential statement. One of the villagers tells a story from his youth, when he had an accident while playing with fireworks. The register II form is used to signal that the proposition is an inference drawn from the fact that the firework blew up in the speaker’s hand.
5.4 Mirative

Even after the villagers acknowledge that Ahmad did in fact visit the burial mounds (since he buried a knife there for them to recover during the day), they still do not believe that he managed to muster up the courage to do this. Rather, they think that he has struck a deal with the supernatural beings and has thus become a devband. In (16), the Perfect is used two times to express the mirative nature of their realisations, first that they had been ignorant, and second that Ahmad, in their eyes, had become a devband.

(16) Lekin hanūz on-ho bovar na-me-kard-and, ki Ahmad in but still DEM.DIST-PL belief NEG-IMPF-do.PST-3PL CLM Ahmad DEM.PROX kor=ro bo dîl u gurda=i xud kard-a=ast, but=DO with heart and kidney=EZ own do.pp=COPI.NPST.3SG but me-guft-and: “Mo bexabar bud-a=em, Ahmad devband IMPF-say.PST-3PL we unaware be.PST-PP=COPI.NPST.1PL Ahmad exorcist šud-a bud-a=ast”, become.PST-PP be.PST-PP=COPI.NPST.3SG

But they still didn’t believe that Ahmad had done this through his own courage. Instead, they said “We were in the dark, Ahmad had become a devband.” (ibid.:70)

A friend of Ahmad claims to have seen a jinn in the shape of a flame. Ahmad approaches the flame and stomps it out with his foot. The onlookers are amazed at this. In (17), the author describes how they all thought that Ahmad was a devband. As can be seen in the original, only the words Ahmad devband are within quotation marks. One could argue that this is an example of using the bud-a=ast to end a quotation. However, this sentence is not uttered by any of the characters, so it cannot be considered sufficiently colloquial to warrant such an interpretation. Since it is a register II verb form used in a mirative sense, it would make more sense to include it within the quotation marks, as has been done in the translation.

(17) Hama guft-and, ki “Ahmad devband” bud-a=ast. all say.PST-3PL CLM Ahmad exorcist be.PST-PP=COPI.NPST.3SG
Everyone said “Ahmad is an exorcist.” (ibid.: 76)

5.4.1 Epistemic verbs and constructions

Sometimes, epistemic verbs and construction are used together with register II forms in a type of third person mirative sense, when the new information made such an impact on the speaker that the register II form of the mirative is retained even in accounts by the narrator. An example of this can be seen in (18), which takes place after Ahmad’s nocturnal visit to the burial place and the crowd sees that he did in fact finish what he set out to do. The realisation
that Ahmad completed his task is retained on the verb form of the proposition following the phrase “it became clear as day that...”.

(18) Rūz šud-∅, oftob bar omad-∅, ba hama čun rūz ravšan
day become.pst-3sg sun PREV come.pst-3sg to all like day bright
šud-∅, ki Ahmad kornoma=ro tamom kard-a
become.pst-3sg clm Ahmad task=do complete do.pst-pp
bud-a=ast.
be.pst-pp=cop.npst.3sg

It dawned, the sun rose, for everyone it became clear as day that Ahmad had finished what he set out to do. (Ayni [1928] 1948: 70)

Ahmad’s father confirms that Ahmad has indeed seen jinns, but also says that the reason he is afraid is that he does not know that jinns cannot hurt people. In (19), the register II forms are prefaced by *ma’lum me-šav-ad* ‘it is clear that’.

(19) Durust, in čiz-ho=e ki did-i jin va ajina
ture dem.prox thing-pl=ind clm see.pst-2sg jinn and jinn.pl
ast, – guft-∅ ba Ahmad padar=aš, – lekin ma’lum
cop.npst.3sg – say.pst-3sg to Ahmad father=pc.3sg – but apparent
me-šav-ad, ki tu az yak čiz bexabar
become.impf-npst-3sg, clm you from one thing unaware
bud-a=ī va binobar in az in-ho
be.pst-pp=cop.npst.2sg and because.of this from dem.prox-pl
tarsid-ā=ī.
to.be.scared.pst.pp=cop.npst.2sg

“True, the things which you saw were jinns,” Ahmad’s father said to him, “but it’s obvious that you have been unaware of something, and because of this you were frightened of them.” (ibid.: 73)

It might be worth comparing (19) with a similar utterance involving the same people. However, the utterance in (20) lacks *ma’lum me-šav-ad*, and the verb is in a register I form. This may well be a coincidence, but it could also indicate that the proposition is not of the kind that is derived from any type of realisation.

(20) Tu az in voqea bexabar bud-ī, binobar in
you from dem.prox incident unaware cop.pst-2sg because.of dem.prox
tarsid-ī, on-ho ba nazar=at namo-(y)on
to.become.scared.pst-2sg dem.dist-pl to view=pc.2sg appear.npst-prsp
šud-a=and.
become.pst-pp=cop.npst-3pl

You were unaware of this incident, that is why you were frightened [and] they appeared to you. (ibid.: 74)
Ahmad reads a couplet of Saᶜdi to the effect that jinns cannot mingle with humans. In (21), the register II form is prefaced by *fahmid-∅* 'he understood'.

\[(21)\]

Ahmad dar maⁿi=i in bayt fikr kard-a az in, čunin Ahmad in meaning=ez this couplet thought.do.pst-pp from such *fahmid-∅*, ki dev […] ba odamī zarar rasond-a understand.pst-3sg clm demon […] to mankind detriment bring.pst-pp na-me-tavonist-a=ast.

NEG-IMPF-can.pst-pp=cop.npst.3sg

When Ahmad thought about the meaning of this couplet, he understood it thus, that […] demons cannot hurt humans. (ibid.: 74)

Ahmad is reading about the natural sciences. In one book he reads how in certain circumstances phosphorus, which is found in bones, can appear as fire. The new information he learns from the books is given in register II forms. The examples below all come from the same few pages that detail Ahmad’s new discoveries.

\[(22)\]

Ahmad dar yak=e az kitob-ho=i fanni *xond-∅*, ki yak Ahmad in one=ind from book-pl=ez technical read.pst-3sg clm one juz’=i ustuxon fosfor nom modda=i kimiyoī part=ez bone phosphorus name element=ez chemical *bud-a=ast*, ki šab-ho, xususan dar havo-ho=i garm be.pst-pp=cop.npst.3sg clm night-pl=ez especially in weather-pl=ez warm ba nazar=i odam monand=i otaš namo-(y)on to view=ez human like=ez fire appear.npst-prsp me-gardid-a=ast.

IMPF-become.pst-pp=cop.npst.3sg

Ahmad read in one of the science books that one constituent of bones is a chemical element called phosphorus, which at night appears to people as fire, especially in warm climes. (ibid.: 77)

But the newly learned information does not need to be prefaced by any lexical marker such as *xond-∅* 'he read’ in (22). In (23), which follows (22), the marking of new information is done solely by the use of a register II form.

\[(23)\]

Har-goh ki ustuxon pūsid-a xok ∅-gard-ad, in every-time clm bone rot.pst-pp dust subj-become.npst-3sg dem.prox čiz padidor *me-šud-a=ast.*

thing apparent impf-become.pst-pp=cop.npst.3sg

Whenever bones rot and become dust, this thing appears. (ibid.: 77)

Directly following (23) comes (24), where there is a juxtaposition between register I and II forms. There is no epistemic verb or construction to trigger the use of the register II form here, but through context (following the two prior examples) it is clear that the reason behind using a register II form here is
that it is newly learned information. The recently acquired information, that phosphorus is the cause of the light phenomenon, is marked with a register II form. This is to be contrasted with the unmarked proposition, i.e. that graveyards are full of bones, which acts as a counterweight to the preceding proposition. In other words, the proposition containing knowledge known from before (“there are bones in the graveyard”) is conveyed with a register I verb whereas the newly learned knowledge, which explains the cause for the light phenomena seen at graveyards (“it is for this reason that…”) is marked with a register II verb.

(24) Šab-ho dar mazor-ot namud-an=i otašpora az on night-pl in graveyard-pl appear.pst-inf=ez ember from dem.dist sabab bud-a=ast, ki dar on jo ustuxon=i reason be.pst-pp=cop.npst.3sg clm in dem.dist place bone=ez pūsid-a bisyor ast. rot.pst-pp many cop.npst.3sg

The reason for the appearance of glowing embers in the burial grounds is that there are a lot of rotten bones there. (Ayni [1928] 1948:77)

In (25), xond-a bud-∅ 'he had read' introduces a subordinate clause which contains a register II form, marking the information as new.

(25) Ahmad dar ta’rix-ho xond-a bud-∅, ki madrasa=i Badalbek Ahmad in history-pl read.pst-pp cop.pst-3sg clm madrasa=ez Badalbex dar jo=i mazor=i kūhna=e bino yoft-a=ast. in place=ez graveyard=ez old=ind building find.pst-pp=cop.npst.3sg Ahmad had read in the annals of history that the Badalbek madrasa was built on an old burial ground. (ibid.:77)

Likewise, in (26), donist-∅ 'he understood' introduces a subordinate clause, which in turn contains a relative clause. The epistemic verb does not, however, scope over the relative clause, as can be seen by the use of register I forms. The proposition that informs the reader that the cause of light was phosphorus, however, is marked with a register II form.

(26) Pas donist-∅, ki on otašpora-ho=e ki dar on jo so know.pst-3sg clm dem.prox ember-pl=ind clm in dem.prox place namo-(y)on šud-a hama=ro dar tars andoxt-a appear.npst-prsp become.pst-pp all=do in fear throw.pst-pp bud, ba ġayr az fosfor čiz=i digar=e cop.pst.3sg to except from phosphorus thing=ez other=ind na-bud-a=ast. be.neg-pst-pp=cop.npst.3sg

He thus understood that those glowing embers which had appeared there and frightened everyone, were nothing but phosphorus. (ibid.:77)
In (27) we find a similar statement as in (26), the difference being that the verb in the subordinate clause is not in a register II form. This is possibly due to the proposition in (26) being newly incorporated information, while the proposition in (27) constitutes a certitude.

(27) [...] digar yaqin paydo kard-∅, ki “dev, pari va ajina” nom [...] other certain apparent do.PST-3SG CLM demon fairy and jinn name čiz=e dar dunyo nest-∅.
thing=IND in NEG.COP.NPST-3SG
 [...] he found out for sure (= he became convinced) that there are no such things as “demons, fairies and jinns” in the world. (ibid.: 77)

5.5 Sarcasm

In (28), Ahmad, who does not believe stories about jinns and demons, reacts sarcastically to such a story by using the Perfect of budan.

(28) In-ho ham dev yo ajina bud-a=and - diya, -
DEM.PROX=PL also demon or jinn bc.PST-PP=COP.NPST.3PL - EMPH -
guft-a Ahmad pas az yak pičing parrond-an pursid-∅: [...] say.PST-PP Ahmad after from one sarcasm throw.PST-INF ask.PST-3SG [...]
“So these were demons or jinns then,” Ahmad said, and after displaying his sarcasm, he asked: [...] (ibid.: 64)

Prior to the sentence in (29), Ahmad had related how he visited the burial ground together with his dog and found out that the sound of feasting fairies was nothing more than howling jackals that ran away when he approached them with his dog. In (29), a person in the audience asks him a question, using the Presumptive. The reason for using the Presumptive here can best be translated with the English “I presume...”, spoken in a sarcastic manner.

(29) Az uvvos=i sag=i tu dev-on=i darun=i ãor az na’ra from barking=EZ dog=EZ you.2SG demon=PL=EZ inside=EZ cave from bellow zad-an va aždarho az nafas kašid-an ham hit.PST-INF and dragon from breath draw.PST-INF also
remain.PST-PP-PRESU-3SG - say.PST-3SG Rūzimurod jest have.PST-PP
“I presume the sound of your dog’s barking also stopped the demons in the cave from roaring and the dragon from breathing?” Rūzimurod asked in jest. (ibid.: 69)
5.6 Questions

In (30), a shepherd is relating a story about how he once saw glowing embers, but is interrupted by another villager, who uses a register II form to ask him about the nature of these glowing embers. This is an example of how the register II forms are used in questions, possibly in order to further accentuate the uncertain nature of this phenomenon.

(30) – On otašpora-ho či bud-a=ast? – gů-(y)on
– DEM.DIST ember-PL what be.PST-PP=COP.NPST.3SG – say.NPST-PRSP
Ašūr suxan=i guppon=ro burid-a taajjub kard-∅.
Ashūr speech=ez shepherd=DO cul.PST-PP surprise do.PST-3SG
“What were those glowing embers?” Ashūr asked surprisedly, interrupting the shepherd. (Ayni [1928] 1948:65)

Interestingly, the answer to the question in (30) is given with a register I form: ajina bud-∅ ‘they were jinns’. This further corroborates the theory that the villagers themselves talked about the supernatural beings using register I forms, as we also have seen in example (2).

5.7 Gnomic statements

Gnomic statements are statements of general truths. In Tajik, the Perfect of budan ‘to be’ and doštan ‘to have’ are used for these kind of statements. In (31) we see how the Perfect of budan is used to mark such a gnomic statement. As argued on page 91, it makes more sense to view these statements as a subset of the evaluative function of the Perfect rather than an independent usage. In other words, it is the personal evaluation or judgement that triggers the Perfect form, not the quality of “general truth” per se. The personal judgement is also amplified by the use of the emphatic diya.

(31) Agar aql na-∅-boš-ad, sar – kadu=i xušk, dahan – tubak=i
if reason NEG-SUBJ-COP-3SG head – pumpkin=ez dry mouth – pot=ez
bekor, va čašm - šiša=i xoli bud-a=ast diya!
worthless and eye – glass=ez empty be.PST-PP=COP.NPST.3SG EMPH
If reason is lacking, then the head is nothing but a dry pumpkin, the mouth a worthless pot, and the eye an empty glass vessel. (ibid.:70)

5.8 Proverbs and sayings

The Perfect of guftan ‘to say’ in the third person plural, guft-a=and ‘they have said’, is appended at the end of adages and sayings.
5.9 Summary

As mentioned in the introduction above, *Ahmadi devband* is partly autobiographical, and many of the incidents recounted in it can also be found in Ayni’s autobiography. With this piece of information in mind, the analysis of the interplay between the use of register I and register II forms takes on a new dimension. Another important aspect to bear in mind is that *Ahmadi devband* was probably not meant to be read solely for enjoyment. It is clear from both the text itself and the moral of the story that Ayni intended it to serve as vehicle for raising both the literacy and the enlightenment of a newly formed nation. As regards the text itself, difficult or unusual words are often elucidated, either with synonyms in brackets or explanatory footnotes. As for the message, the story decidedly opposes superstition and encourages the reader to find scientific explanations of otherwise inexplicable phenomena in nature.

The story is thus a carefully crafted didactic piece of literature, both in text and in message. It is therefore not unlikely that Ayni would be just as conscientious when it comes his use of verb forms. In fact, Ayni’s use of the register II forms is extremely consistent throughout the text and it comes as no surprise that the story reads as a textbook example of how to use these forms.

5.9.1 Reportative

The majority of the register II forms found in *Ahmadi devband* are used reportatively, and they are of two types. With the first type, the author is
describing the beliefs of the villagers, while the second type occurs in the dialogues of the characters in the story.

The first type occurs frequently in the opening paragraphs of the story, when the author is describing the villagers’ superstitious beliefs, and the lore surrounding the cave. Usually, these register II forms appear in conjunction with longer passages prefaced by lexical cues informing the reader that the information is hearsay. Some instances are more subtle, however. Take (1) as an example. Here, the register II form is surrounded by propositions made up of register I forms. The verb in the register II form only scopes over the proposition that describes the symbolism of the seven flags.

The story has several longer narratives that deal with hearsay, such as when Ayni describes supernatural creatures and the lore surrounding them. There are however indications that the villagers themselves did not talk about these creatures using register II forms, at least not in all circumstances. In (2), we see how the villagers speak about these creatures using register I forms. Another example is found in the answer to the question in (30). A possible explanation of the use of register I forms in these instances is that in the minds of the villagers these creatures were no less real than anything else seen in nature. After all, some had seen the “glowing embers” with their own eyes, so naturally the proposition that there are demons and fairies at the graveyard was accepted as common knowledge. When Ayni, as the author, retells their stories, however, he does it consistently using register II forms, sometimes together with lexical means, to emphasise that this is in fact reported speech.

If we compare and contrast the lexical with the grammatical evidentiality, the latter seems to be more independent. This is due to the fact that all lexical cues of reportativity such as “they say that” are followed by either a clearly marked direct quote, such as in (2), or a register II form, such as in (3). On the other hand, grammatical evidentiality is independent of lexical markers, as the register II form can function as the sole marker of a reportative proposition, as in (1).

Moreover, in (9), the indirective propositions are marked as such by both the register II forms that scope over these propositions and the statement “My late mother used to say”, which introduces these reportative statements. In (10), which is the direct continuation of (9), the only indications to the reader that there has been an evidential shift are the register I verb forms of the propositions. In other words, the lexical markers serve an ancillary role in relation to the grammatical marking on verb forms.

5.9.2 Inference, questions and sarcasm

We also find examples of register II forms being used for inferred information. Moreover, we see instances where the register II forms and the Presumptive are used in sarcastic statements. This use is not mentioned by Perry (2005),
but has been attested in Turkish.\footnote{Aksu-Koç and Slobin (1986: 162) list “doubting scorn” as a possible pragmatic extension of the Turkish indirective -mlş.} The register II forms are also used in certain questions. This has been attested in Persian (see page 77 regarding Kasravi’s views on the Persian Perfect) and Uzbek (Straughn 2011: 158).

5.9.3 Epistemic verbs and constructions

An interesting use of the register II forms is its occurrence after lexical evidentials such as “it became apparent” and “it is clear that” and epistemic verbs such as \textit{fahmidan} ‘to understand’ and \textit{šunidan} ‘to hear’. However, as explained by Perry (2005: 228), the “indirect nature of the experience” does not necessarily have to be prefaced by epistemic verbs and phrases.

In \textit{Ahmadi devband}, many instances of this are found in the latter part of the story when Ahmad learns about the natural sciences. Interestingly, in this segment we also find several instances where the register II forms alone mark the proposition as newly incorporated information.

This shows that the register II forms can be used to mark newly incorporated information without being prefaced by an epistemic verb or phrase, as long as it occurs within a longer sequence of such items.

We also have one instance in which the epistemic verb is not followed by a register II form, thus constituting an exception to the example set by the author in previous sentences of the same nature. A reasonable explanation is that a register I form is used in this particular case because the epistemic verb introduces not new knowledge, but integrated knowledge. In other words, only “new” knowledge is marked as such with register II forms, whereas integrated knowledge is conveyed using register I forms.
6. Analysis of *Odamoni jovid*

*Odamoni jovid* (‘Eternal People’) is a novel published in 1949 by Rahim Jalil (1909–1989). The first part of *Odamoni jovid* was originally published in 1941 with the title *Gulrū*, which is the name of the female protagonist (Hitchins 2008). The novel has 283 pages and is divided into 35 chapters. The analysis presented here covers chapters 1–6, comprising a total of 41 pages.

At the beginning of the novel, the narrator describes the rural life of the protagonist Gulrū and her mother and father, Saodatbibi and Rustamamak, who both come from humble backgrounds and now work hard as farmers to make ends meet and provide Gulrū with everything that she needs. The narrator does not specify the age of Gulrū, but from comments made by the other villagers regarding how she has “become beautiful”, and her interactions with a certain Red Army soldier, we can surmise that she is in her teens.

Gulrū and her family live in the village of Tarob. For the last two years, the Basmachis\(^1\) have been using Tarob as a base from which they launch attacks on the Red Army and its supporters, who are called *Fidoiyon* ‘those who sacrifice themselves’. At the start of the novel, however, the Red Army and the Fidoiyon have managed to drive away the Basmachis and capture the village, setting up camp just outside of the house of Gulrū’s family.

Even as the village begins returning to its former condition of peace and quiet, danger looms on the horizon. One day, as Gulrū is filling her jug with water from the river, a Red Army soldier urges her to go home to safety since the Basmachis are approaching. Gulrū and her family are spared in the subsequent recapture of the village and massacre. One of the Basmachi leaders, however, has his eye on Gulrū, and in an attempt to get to her, he summons her father and accuses him of collaborating with the Red Army.

---

\(^1\) The Basmachis (Turkic for “raider, bandit”) were an armed peasant movement resisting Soviet rule in Central Asia. The Basmachis were active to varying degrees between 1918 and 1948, but they were as strongest between 1920 and 1931. One of the hotbeds of the movement was located in Eastern Bukhara, which in 1924 came to be known as Tajikistan (Nourzhanov 2015: 177).

For the greater part of Soviet rule, the politically correct view of the Basmachis was that they were bandits and criminals—mercenaries of the Tsar and local vassals tasked with stifling the revolutionary activities of the peasantry. Since the end of the 1980s, however, revisionist voices have questioned the predominant view of the Basmachis as counter-revolutionary savages. Instead, it was argued, they should be seen as anti-imperialists and patriots (ibid.: 179–180).
6.1 Reportative

In (1), the narrator describes the surroundings of Tarob village. The reported speech is marked both by the introductory *naql me-kun-and* ‘they say that’ and the use of the Perfectoid Imperfect and Perfectoid Pluperfect.

(1) *Naql me-kun-and,* ki yak zamoon-ho qabila-ho=i digar ba narration IMPF-DO.NPST-3PL CLM one time-PL tribe-PL=EZ other to xalq=i on jo hujum me-kard-a=and,* mardum people=EZ DEM.DIST place assault IMPF-DO.PST-PP=COP.NPST.3SG people baro=i mudofia=i xud=ašon in qal’a=ro soxt-a for=EZ defence self=PC.3PL DEM.PROX fortress=DO build.PST-PP bud-a=and.
be.PST-PP=COP.NPST.3SG

They say that at one time there were other tribes that used to attack the inhabitants of that place. In order to defend themselves, the people had built this fortress. (Jalil 1949: 9)

After the fight with the Basmachis, Pūlod Qodirov, the Red Army soldier who had previously warned Gulrū of the imminent attack of the Basmachis, is in the company of two of his fellow soldiers. Before the fight, Pūlod Qodirov had received a letter from one of the villagers, but he had not had time to read it earlier because of the impending battle.

In (2), Pūlod Qodirov reads the letter out loud to the other two soldiers. We see how the author of the letter first uses register I forms (the Preterite and the Imperfect) when describing how the Basmachis fled and how the dead lay in the streets. We then have a shift to the Perfectoid Pluperfect in describing how the village imam gathered and hid the weapons. The author of the letter then switches to the Pluperfect when describing how his son had seen this with his own eyes and reported it to him. Register I forms are thus used for setting the stage, i.e. for describing how the Basmachis fled and how their dead lay in the streets, whereas the reported actions of the village imam are described with register II forms.
When the Basmachis fled our village, their dead were lying in the streets. The imam of the village, Hodixontūra, took three revolvers, four bombs, and all of the rifle ammunition he could from the dead and took them to his home and hid them. My little son saw this with his own eyes and reported it to me. (Jalil 1949:31)

Example (3) follows directly on example (2). The writer continues his letter with more information about the imam and the weapons he took. He writes that his neighbour reported to him that the imam had buried the weapons in the khanqah. The reported information is given using the Perfect, and the perfect participle is used throughout the report for serialisation: pećond-a, gūr kard-a, andoxt-a. It is thus the Perfect mond-a=ast ‘has placed’, the last finite verb of the sentence, which determines the verb form for all of the perfect participles preceding it.

After that, my neighbour Safarboy the cobbler reported that the imam had wrapped these things in a rag and buried them by the mihrab in the mosque’s khanqah and covered the spot with mats and prayer rugs. (ibid.: 31)

2. A khanqah is where Sufis gather to worship, but the term can also be used to denote a little building adjoining the mosque (Nazarzoda et al. 2008b: 448).
Akram, who works for the Basmachi warlords, brings a young man to the Basmachi headquarters. One of the warlords asks Akram who this young man is, and why he has brought him to them. The choice of verb forms in Akram’s reply indicates that the information he is giving to the guards is based on reported speech, presumably information received by Akram from the guards who arrested the young man.

First, Akram uses the Subjunctive together with the word *darkor* ‘must’, saying that the youth “must be a Bolshevik informer”. Later, Akram uses the Perfectoid Imperfect when describing how the youth “(reportedly) left the village and made his way to the city”. Note that there are no lexical markers for the indirective statements.

(4)  
– Janob=i bek, -guft-∅ Akram, -in xabar-kaš=i  
– mister=ez chief say.pst-3sg Akram dem.prox news-draw.npst=ez  
bolševik-ho ∅-boš-ad darkor, ki az qišloq bar omad-a  
Bolshevik-pl subj-cop.npst-3sg must clm from village prev come.pst-pp  
ba taraf-i şahr me-raft-a=ast, qarovul-ho xud=i az  
to direction=ez city impf-go.pst-pp=cop.npst.3sg guard-pl self=ez from  
iferay=direction=ez city  
entrance take.pst.pp come.pst.pp=cop.npst.3pl  
– Mr Bek⁴, said Akram, this must be a Bolshevik informer who (reportedly) left the village and made his way towards the city, the guards (reportedly) arrested him by the city gate. (ibid.: 24)

In (5), Pūlod describes how he ran into Salimboy at the headquarters. Since there is no lexical marker accompanying the Perfect of *budan*, there is no way of determining whether this utterance is reportative or inferential. But since Pūlod is describing how he actually met Salimboy, it is probably safe to assume that the proposition is marked with a register II form due to it being a report by Salimboy himself.

(5)  
– Salimboy ba’d az hodisa=i imrūza […] ba şahr gurext-a raft-a  
– Salimboy, after from incident=ez today […] to city flee.pst-pp go.pst-pp  
ba ştab ham xabar dod-anî bud-a=ast, ki man  
to headquarters also news give.pst-fp be.pst-pp=cop.npst.3sg clm I  
dar in jo ba vay rû ba rû şud-am.  
in dem.prox place to he face to face become.pst-1sg  
– After today’s incident, Salimboy went back to town to (reportedly) also inform⁴ the headquarters, which is where I encountered him. (ibid.: 42)

---

3. The title *bek* / *beg* has historically been used for rulers, but also as a term of respect for noblemen (Nazarzoda et al. 2008a: 173).

4. The form with *-anî* is a future participle used to express volition, intention, and future tense (Perry 2005: 264–266).
6.1.1 Non-uses

One of the Red Army soldiers tells Gulrū to go home because the Basmachis are approaching. Gulrū sees the Red Army troops marching out of sight and then she hears the sound of gunfire.

After returning home, Gulrū’s mother Saodatbibi asks her where the sound of gunfire is coming from. Example (6) is given as an answer to Saodatbibi’s question, and since Gulrū did not actually see the Basmachis, one could expect a register II form here. Instead, the Preterite is used. Perhaps the immediacy and severity of their arrival takes precedence over any indirective signalling. The Preterite can be used for events that are about to happen (Perry 2005:214), and the use in (6) is most likely an example of this.

(6) Ona-jon bosmači-ho omad-and [...] mother-dear Basmachi-PL come.pst-3pl [...] Mother dear, the Basmachis have come (lit. came) [...] (Jalil 1949:14)

Interestingly, a little while after the dialogue in (6) takes place, Gulrū’s father Rustamamak comes home and warns the family by also using the Preterite. For this statement, however, we do not have any back story, so it is possible he actually saw the Basmachis approaching. Most likely, it is the same use of the Preterite as in (6).

(7) - Gulrū=ro rust ⊗-kun-∅, bosmači-ho omad-and!.. - Gulrū=do hidden impv-do.npst-2sg Basmachi-PL come.pst-3pl Hide Gulrū, the Basmachis have come (lit. came)! (ibid.: 14)

6.2 Inference

6.2.1 Non-uses

These suppositions occur with register I forms, even though one could have expected register II forms for inferences and the Presumptive for presumptions.

After hearing the contents of the letter addressed to Pūlod Qodirov, the soldiers discuss whether the weapons are still where the imam had hid them. In (8), the speaker prefaces his proposition with “I think” and “surely”, but uses the Preterite.

(8) – Ba fikr=i man, [...] albatta in yaroq-ho ba dast=i – to thought=ez I [...] surely dem.prox weapon-pl to hand=ez vay=ho aftid-∅ s/he=pl fall.pst-3sg I think that these weapons surely fell into their hands. (ibid.: 31)
Directly after one of the soldiers has uttered his thoughts on the whereabouts of the weapons in (8), Pūlod Qodirov gives his own views in (9). The verb in the subordinate clause (istod-a=and) is in the Perfect. However, since it is a stative-dynamic verb, the Perfect has a present tense meaning in conveying a position, in this case that the things are “standing” under ground. It is thus not possible to discern whether the proposition in (9) is indirective.

(9) Lekin dil=am guvohi dod-a istod-a=ast, ki
    But heart=PC.1SG witness give.PST-PP stand.PST-PP=COP.NPST.3SG CLM
    […] in čiz-ho hol=i hozir dar tag=i xok
    […] DEM.PROX thing-PL state=EZ present in under=EZ soil
    istod-a=and.
    stand.PST-PP=COP.NPST.3PL
    But my heart tells me that these things are beneath the ground right now. (ibid.: 31)

The proposition in (10) is a direct continuation of Pūlod Qodirov’s statement in (9). Just like the speaker in (8), he prefaces the proposition with “surely” and uses the Preterite.

(10) Domullo=i az šodî xud=aš=ro gum kard-a-gî
dem=PC.3SG=DO hidden do.PST-PP-PP₂
    mulla=EZ from happiness self=PC.3SG=DO hidden do.PST-PP-PP₂
    on-ho=ro albatta az xotir=aš bar ovard-∅
    DEM.DIST-PL=DO surely from memory=PC.3SG PREV bring.PST-∅
The mulla, who is beside himself with joy, surely forgot about them. (ibid.: 31)

6.3 Evaluative

6.3.1 Non-uses

The statement in (11) is taken from a dialogue in which a person expresses his admiration for a certain Kucherenko. Since this is an evaluative statement, one could expect a register II form here, yet the Preterite of budan is used. This may be an indication that only evaluative statements with a present tense meaning take the Perfect of budan.

(11) bečora Kučerenko kušt-a šud-∅? Hay afsus! Či
    [...] poor Kucherenko kill.PST-PP become.PST-3SG oh alas what
    odam=i xub=e bud-∅!
    person=EZ good=IND COP.PST-3SG
    [...] was poor Kucherenko killed? For shame! What a good man he was! (ibid.: 35)
6.4 Epistemic verbs and constructions

In (12), the narrator is describing Gulrū’s family. We learn that their white hair and wrinkles testify to the fact that they had aged prematurely.

(12) Ojing=i rūy, mūy-ho=i safed=i sar guvohi me-dod-and ki wrinkle=ez face hair-pl=ez white=ez head witness IMPF-give.pst-3pl CLM mašaqqat-ho=i zindagi=i in xonavoda=i qaššoq peš az vaqt adversity-pl=ez life=ez DEM.PROX family=ez poor before from time pir=aš kard-a=and.

The wrinkles of the face, the white hair of the head testified that the adversities of life had aged this poor family prematurely. (Jalil 1949: 4)

6.4.1 Non-uses

There are three instances where epistemic verbs are used but the subordinate clauses do not contain register II forms. A reason for this may be that the epistemic verbs in these examples—“to see” and “to notice”—are not of the type that normally would be used with register II forms.

The Red Army base is located opposite to Gulrū’s house and Gulrū passes by the soldiers when going to the river to get water. In (13), the author describes how Gulrū notices that one of the soldiers looks at her more than the others. The epistemic verb here is pay burd-∅ ‘she noticed’, but it is not accompanied by a register II form in the subordinate clause.

(13) Vale raft-a – raft-a pay burd-∅, ki ġayr az ū digar-on but go.pst-pp – go.pst-pp trace bring.pst-3sg CLM except from he other-pl ba ū on qadar diqqat na-me-kun-and. to she DEM.DIST degree attention NEG-IMPF-do.npst-3sg

But gradually she noticed that apart from him, the others didn’t pay her so much attention. (ibid.: 11)

In (14), we have the verb “to see”, but it does not trigger the use of a register II form in the subordinate clause. This is possibly due to the fact that the sentence is in the passive voice: “one could see”.

(14) Dar in ravšani did-a me-šud-∅, ki az in DEM.PROX light see.pst-pp IMPF-become.pst-3sg CLM from pešona=i Rustamamak xun me-šorid-∅. forehead=ez Rustamamak blood IMPF-flow.pst-3sg

In this light one could see that there was blood flowing from Rustamamak’s forehead. (ibid.: 23)

The Red Army soldier who has taken a liking to Gulrū watches her filling her jug with water.
The youth saw that the village girl pulled the water-filled jug out of the water with fear. [...](ibid.:11)

6.5 The Presumptive

In (16), Gulrū’s mother Saodatbibi describes how they will marry off their only child. She is using the Non-Past Presumptive (see page 93) to present a hypothetical future scenario in which her daughter makes their house into a home. The verb form in question here is me-šišt-a-gist-∅, the Non-Past Presumptive of šištan ‘to sit’. The emphatic oxir ‘at least’ further accentuates the meaning of the Non-Past Presumptive here.

Let’s pair her up with one of God’s poor servants, she will surely make our dilapidated and decrepit little house into a home and settle down in it after we are gone. (ibid.:6)

In (17) and (18) Salimboy encounters a stray kid goat, and the reader is presented with his thoughts verbatim. If we interpret the Presumptive uses in these sentences according to Perry’s descriptions, the reason for the Presumptive is that Salimboy is making conjectures based on past experience and knowledge, and not inferences based on anything apparent in the present situation. Salimboy’s premises for making the Presumptive statement in (17) can thus be summarised as “Kid goats usually have their owners close by” and (18) as “Rich people own herds of goats”.

He thought to himself “its owner is surely following it from behind”. (ibid.:41)
In (18), Salimboy uses the Presumptive to make assumptions about the kid, this time prefaced with the adverb *albatta* 'surely', further emphasising the presumptive nature of the statement. Note, however, that his thoughts end with *šud-a-gī* used predicatively, which according to Perry (2005: 218) is purely resultative in meaning.

(18) “*Albatta* in *buz-ak* azon=î *yagon* boy surely *DEMPROX* goat-*DIM* belong=*EZ* *rich.person*
*bud-a-gist-∅-* guft-a *Salimboy fikr* kard-∅, – az *yagon*
be.*PST-PP-PRESU-3SG* say.*PST-PP* Salimboy thought do.*PST-3SG* – from one
*podâ judo* šud-a-gī. [...]”
*herd* ‘separate become.*PST-PP-PP₂* [...]’
“Surely, this little goat belongs to a rich person,” Salimboy thought “he has been separated from some herd. [...]” (Jalil 1949: 41)

### 6.6 Sarcasm

In (19), the Presumptive is used for sarcastic effect, just as in *Ahmadi devband*, example (29). The Basmachis are talking and the mulla remarks that he had no bad intent regarding an earlier comment. He affirms that his heart is not “askew”. Another one jokingly remarks that in that case the mulla’s heart must be as straight as a sickle, to which the crowd laughs.

(19) *Rosti=i* dil=aton dos *barin* *bud-a-gist-∅* domullo?
*straightness=EZ* heart=*PC.2PL* sickle like be.*PST-PP-PRESU-3SG* mulla
Your heart must be straight as a sickle then, mulla? (ibid.: 22)

### 6.7 Proverbs and sayings

After quoting a proverb or adage, the characters always append the phrase *guft-a=and* 'they have said' to it.

(20) *Se ėiz* ba *čašm* quvvat me-dih-ad, *guft-a=and*...
three thing to eye power *IMPF-give.NPST-3SG* say.*NPST-PP=COP.NPST.3PL*
They say that three things give power to the eye... (ibid.: 22)

(21) “Avval pursiš, *ba’d* kušiš” *guft-a=and.*
*first* question then killing say.*PST-PP=COP.NPST.3SG*
“First you ask, then you kill”, they say. (ibid.: 23)

(22) *Ya’ne masalan* dar kitob-ho=i *šarif* *guft-a=and*, ki...
i.e. e.g. *in* *book-PL=EZ* venerable say.*PST-PP=COP.* CLM
For example, in the holy books they say that... (ibid.: 23)
(23) Magar na-šunid-a=s, ki nek-on “ganj dar vayrona” but NEG-hear.PST-PP=COP.NPST.2PL CLM good-PL treasure in ruin guft-a=and? say.PST-PP=COP.NPST.3PL

Haven’t you heard that they say good-natured people are a “treasure in the ruin”? (ibid.: 7)

6.8 Šud-a=ast

As explained on page 90, šud-a=ast can be used colloquially to mean “that’s good enough” or “that’s enough”. In this text, however, there are several instances where the characters use šud-a=ast as a way of responding positively to a request. It seems to correspond to the English “will do” or “consider it done”.

In (24), one of the Basmachi leaders asks one of the men present if he knows where Rustamamak lives, and if he can fetch him, to which he responds:

(24) - Me-don-am bek, xub šud-a=ast bek.
- IMPF-know.NPST-1SG chief good become.PST-PP=COP.NPST.3SG chief

I know, chief; it will be done, chief. (ibid.: 22)

In (25), Pūlod Qodirov is commenting on the best time to leave. This example stands out from the others because it is not in response to a request. It still retains its future reference.

(25) To čašm=i rūz=ro kūr na-kun-em, ba roh dar omad-a until eye=EZ day=DO blind NEG-do.NPST-1PL to way PREV come.PST-PP na-me-tavon-em -guft-∅ ǔ.- nisf=i šab ba qišloq dar NEG-IMPF-can.NPST-1PL say.PST-3SG he half=EZ night to village PREV omad-a ∅-rav-em šud-a=st diya, oxir. come.PST-PP SUBJ-go.NPST-1PL become.PST-PP=COP.NPST.3SG EMPH EMPH “We can’t set out until it gets dark,” he said, “leaving the village in the middle of the night is at least doable.” (ibid.: 34)

6.9 Summary

If we disregard the special uses accorded to the Perfect of guftan and šudan, there are not all that many actual instances of the register II forms, and all of them pertain to reported speech.

When the register II forms are used to signal reported speech, the source of information is either mentioned through lexical means before the indirective propositions, as in examples (1) and (3) or is otherwise supplied through context, as in example (2). In example (4), we find no overt source, but the
most probable reading here is that the information comes from the guards who arrested the man and handed him over to the speaker.

The text exhibits some interesting uses of the Presumptive that corroborate Perry’s description of them as expressing “an unsubstantiated conjecture or assumption”. The difference between it and using the register II forms in similar settings seems to be that the Presumptive is used when the supposition is based on prior knowledge of how things usually are, whereas the register II form is reserved for inferences drawn from the scene at hand.

There are also some cases of non-use of the register II forms. These indicate that there is more to the use of the register II and register I forms than just information source, and that reportative, inferential and presumptive statements do not necessarily have to contain the verb forms that are usually associated with them (i.e. the register II forms and the Presumptive).
7. Analysis of *Dar orzui padar*

The novel *Dar orzui padar* (‘Wishing for a Father’) was written by Karomatulloh Mirzo (1942–) in the year 1989. Mirzo was awarded the national Rudaki prize for this work, which is the first part of a trilogy and to date is his most well-known book. The first two parts of the trilogy were made into a very popular TV series that aired on Tajik national television in the 1990s (Mirzo [1989] 2009). *Dar orzui padar* tells the story of a small boy, Mehrubon, and the disintegration of his family due to his father’s unfaithfulness. The analysis presented here covers the first 35 pages of the first part of *Dar orzui padar*.

The book begins with the journalist Nozimi who comes home to his family, looking very wretched and depressed. He is not like his usual self, and does not respond when his son, Mehrubon, greets him. Nozimi is not getting along very well with his wife Musharraf, and they quarrel a lot.

At the beginning of the novel, the story is narrated from the perspective of Mehrubon. On page seven, however, a narrational shift occurs, when Musharraf, in a conversation with Mehrubon, curses the day she first met Shahnoza, Nozimi’s mistress, who was also a close friend of Musharraf. This leads her to have a flashback that spans seven pages and ends with a retrospective conclusion about why Nozimi had “returned that day like that”, referring to the very first paragraph of the book, when Nozimi comes home looking wretched.

From Musharraf’s flashback, the reader learns that her friendship with Shahnoza begins one day at the school where Musharraf works. Musharraf receives news that her son Mehrubon has fallen ill. Musharraf searches desperately for someone to cover her spot, and Shahnoza, who she has never talked to before, agrees to help her. From that point onwards they develop an intimate friendship. It is not long, however, until Musharraf suspects that something is awry between Shahnoza and her husband Nozimi. First, Nozimi openly praises Shahnoza’s beauty in front of Musharraf and Mehrubon. Later, the family is invited over to Shahnoza’s home for a wedding party where Shahnoza dances with Nozimi, the sight of which enrages Musharraf.

Not long after, Musharraf learns that Nozimi has written an enthusiastic and laudatory article about Shahnoza. Enraged, she confronts Shahnoza and an altercation ensues. Shahnoza visits Nozimi at his work and demands that he choose between her and Musharraf. At this point, Musharraf’s flashback ends, and the reader is back in the original timeline with Musharraf visiting Nozimi’s co-workers to make them persuade him not to marry Shahnoza. At
this juncture, Musharraf has a second flashback in which she remembers how she first met Nozimi and how they courted. Since the flashbacks are retold through the eyes of Musharraf, they contain several instances where the register II forms are employed to mark inferences and internal realisations. These retrospective epiphanies occur because “present day Musharraf” knows things that “flashback Musharraf” does not. The register II forms are used to mark “new information” against the narrational background marked with register I forms.

7.1 Reportative

In (1), the narrator describes how Nozimi used to come home late from work. The narrational centre in this segment is Mehrubon; i.e. the information given in this segment is filtered through him. The first two verbs which describe how Nozimi would be on duty and publish newspapers (me-šud-∅; me-bar-ovard-∅) are in the Imperfect, but the third verb is in the Perfectoid Imperfect without any lexical cues. The use of the register II form most likely signals reported information in this example.

(1) Navbatdor me-šud-∅, gazeta me-bar=ovard-∅,
on.duty IMPF-become.PST-3SG newspaper IMPF-PREV=bring.out.PST-3SG yak gazeta az avval to oxir xud=aš one newspaper from beginning until end self=PC.3SG me-xond-a=ast IMPF-read.PST-PP=cop.npst.3sg […]

He would be on duty, he would publish newspapers, he would (reportedly) proofread a newspaper from beginning to end […] (Mirzo [1989] 2009: 5)

In (2), Shahnoza pays Nozimi and Musharraf a visit one evening and invites them to a wedding. The Perfectoid Imperfect is used here to convey second-hand information regarding future events. It signifies that this information was relayed to them through Shahnoza. Note that although there are no lexical markers regarding the source of information, the context makes it clear that the source is Shahnoza.
One night she [=Shahnoza] came to her [=Musharraf’s] house. She invited Musharraf, together with her husband and children, to her home. They were (reportedly) going to have a wedding, her brother’s wedding, they were (reportedly) going to bring a bride. (ibid.: 11)

In (3), we get the inner thoughts of Musharraf when she meets Shahnoza for the first time. The omitted copulas of the first sentences make this section come across as colloquial. The proposition “she taught dancing” is given with the Perfectoid Imperfect to accentuate that it is reportative. Note that there are no lexical markers regarding the source of information; the reader must rely solely on the register II form to come to this conclusion.

She [was] the leader of the musicians’ group at the school. She [was] a dancer, and (reportedly) taught dancing. (ibid.: 8)

Musharraf is at Mavlonzoda’s office. The telephone rings and Mavlonzoda answers. Mavlonzoda hangs up and says that he has to go, because “there’s (apparently) a meeting” he has to attend. In (4), Mavlonzoda is using the register II form to signal that the proposition is either reportative (the person on the telephone said that they are having a meeting) or inferential (the person on the telephone does not mention that they are having a meeting, but Mavlonzoda infers this). Note that his first sentence, “they want me” is not marked as indirective, as he is only conveying the request of the person on the phone.

They want me (to come), there’s (reportedly/apparently) a meeting. (ibid.: 23)

In (5), Musharraf remembers the day she first met Nozimi. Before he came to the school, the students had read an article about him in the newspaper. Although no lexical cues are given, the use of the register II form here is due to
the information being reported, most likely from the mentioned newspaper article.

(5) Vay šogird=i navisanda=i mašhur Qutosi bud-a=ast.
    He (reportedly) was the pupil of the famous Qutosi. (Mirzo [1989] 2009:24)

Musharraf and her girlfriends have asked their teacher Sodiqi questions about Nozimi. In (6), Sodiqi replies that he just spoke with Nozimi, who informed him that a book of his stories will be published. This is an interesting example of a register II form, since it is nested within an account of indirect speech, and it is not clear if Nozimi himself used a register II form or if this is something that Sodiqi adds to his statement.

(6) Guft-∅, ki kitob=i hikoya-ho=(y)aš me-bar
    say.npst-3sg CLM book=ez story-pl=pc.3sg IMPF-PREV
    omad-a=ast
    come.out.pst-pp=cop.npst.3sg [...]
    He said that a book with his stories will (reportedly) be published. (ibid.:26)

In (7), Musharraf is at school and they hear that Nozimi wants to write an article about them. The proposition is in the register II form because it is reportative.

(7) Nozimi dar bora=i on-ho maqola navišt-anī
    Nozimi in about=ez DEM.DIST-PL article wrote.pst-fp
    bud-a=ast.
    be.pst-pp=cop.npst.3sg
    Nozimi (reportedly) wanted to write an article about them. (ibid.:27)

In the paragraph leading up to (8), Musharraf’s father has heard that Nozimi is secretly meeting with Musharraf at her aunt’s house. The Perfectoid Imperfect is used because the father is relaying a repeated past action.

(8) - Tu=ro muxbir suroğ kard-a me-omad-a=ast,
    - you.sg=DO journalist direction do.pst-pp IMPF-come.pst-pp=cop.npst.3sg
guft-∅
    say.npst-3sg father to her
    (I hear) the journalist has been looking for you, her father said to her. (ibid.:38)
7.2 Inference

Arguably, many of the statements that have been listed as inferential in this section can also be labelled as 3rd person miratives. Without lexical cues to the exact source of information, inferential and mirative statements are impossible to separate since it is not clear if the proposition is marked with a register II form on account of it being solely new knowledge / surprise, or if it also is because the new knowledge / surprise came about through an inference.

The passage in (9) comes after Musharraf has told her husband about meeting Shahnoza for the first time. Musharraf’s retrospective realisations are given in register II forms. As marked forms they stand out against the background of register I forms, as if to signal Musharraf’s present day realisations: “now I know that he was actually laughing at me”. The use of dar asl ‘in fact’ further emphasises that these remarks are made in hindsight.

In example (10), we once again have Musharraf injecting “present day” insights into the flashback episode. Hence the register II form is used when describing how Shahnoza “sought the way of friendship and intimacy”, but not when describing how she became “warmer and more loving”. In other words, Musharraf noticed that Shahnoza became warmer and more loving towards her, but it is her “present day” self that is making the connection that Shahnoza did this to be friends with her.

(9) Dar javob šavhar xomûš bud-∅, dar asl ba vay in answer husband silent COP.PST-3SG in origin to her
me-xandid-a=ast u Mušarraf bexabar. Behuda dar IMPF-laugh.PST-PP=COP.NPST.3SG and Musharraf unaware in.vain in ham-on lahza-ho ġamxor na-šud-a EMPH-DEM.DIST moment-PL caring NEG-become.PST-PP bud-a=ast in duxtar ba oča=i Mehrubon be.PST-PP=COP.NPST.3SG this girl to mother=EZ Mehrubon me-donist-a=ast ki ū zan=i Nozimi ast, IMPF-know.PST-PP=COP.NPST.3SG CLM she wife=EZ Nozimi COP.NPST.3SG me-šinoxt-a=ast.
IMPF-recognise.PST-PP=COP.NPST.3SG
He remained silent and gave no answer, but in fact he had been laughing at her and she had been unaware. It was not for no reason that this girl had concerned herself with Mehrubon’s mother (= Musharraf), she had known that she was Nozimi’s wife, she had recognised her. (ibid.: 9)
(10) Šahnoza ham, ki bo vay roh=i nazdiki, düstī, Shahnoza also CLM with her way=EZ closeness friendship me-just-a=ast, muhib=u mehrubon-tar IMPF-search.PST-PP=COP.NPST.3SG loving=and kind-comp gāšt-∅… become.PST-3SG
In turn, Shahnoza, who (it turns out) had sought the way of friendship and intimacy with her [=Musharraf], became warmer and more loving… (Mirzo [1989] 2009:9)

After the narrator has described how Shahnoza visits Nozimi at his office (using register I forms), and tells him that he has to choose between her and Musharraf, we get an explanation (through the mind of Musharraf) in (11) as to why Nozimi had returned home angry at the beginning of the chapter. The register II form is used because it is an inference made in the mind of Musharraf, who is looking back on that day and drawing conclusions.

(11) Baro=i ham-in Nozimī on begoh bo dimoğ=i sūxt-a EMPH-DEM.PROX very Nozimi DEM.DIST night with nose=EZ ba xona bar gāšt-a bud-a=ast […]. burn.PST-PP to house PREV return.PST-PP be.PST-PP=COP.NPST.3SG […] It was for this reason that Nozimi had returned that evening offended […] (ibid.:14)

In (12), Musharraf has gone to Nozimi’s office in order to speak with his co-workers. She is greeted very openly by Shafii, and for this reason she infers that he had received news that she would come. The verb form used here is the Perfectoid Pluperfect.

(12) Šafiī az omad-an=i ū būbar šud-a bud-a=ast. Shafii form come.INF=EZ he aware become.PST-PP be.PST-PP=COP.NPST.3SG Shafii had (apparently) learned of her coming. (ibid.:18)

In (13), Musharraf is visiting one of Nozimi’s co-workers, Mavlonzoda, who acts as if he is searching for a certain sheet of paper. Register II is used because Musharraf is making an inference. There are no lexical cues introducing the inferential proposition. The other propositions employ register I verb forms, even the one that describes how Mavlonzoda feels when he finds what he was looking for.
In (14), Musharraf has visited Mavlonzoda in his office to speak with him, but every time she ventures to speak, someone calls on the telephone and interrupts the conversation. This greatly upsets Musharraf, who infers that Mavlonzoda’s superior has been lying in ambush, waiting to disrupt their conversation. Here, we have an inference preceded by гуя, which in this case means “as if”. It is interesting to note that even though this inference is metaphorical, as emphasised by “as if”, it is still marked with a register II form.

Nozimi is interviewing Musharraf and complimenting her on her beautiful name. He asks her what season she was born in. She answers that she was born in spring. In (16), Nozimi gives his romantic response, which uses the Perfect of будан to mark that the statement is an inference on Nozimi’s part. It could, however, also be argued that this is an evaluative statement.
In the passage leading up to (17), Mehrubon asks Musharraf why she does not wear nice clothes like Shahnoza. Musharraf explains that Nozimi does not buy anything for her to wear. Nozimi enters the discussion, saying that even if Shahnoza had no (nice) clothes, she would still be pretty. This is the first time Musharraf hears her husband say something nice about Shahnoza, and the words throw Musharraf into a rage. Musharraf’s thoughts are supplied verbatim in quotation marks (even though the pronouns show that these words are actually not strictly verbatim, but rather semi-direct). The register II form signals Musharraf’s new realisation.

(17) Bo «Shahnoza xušrū» guft-an=aš Nozimī güyо ba jism u with Shahnoza beautiful say.pst-INF=ez Nozimi as.if to body and jon=i vay otaš rext, sůziš=i dil=aš past soul=ez she fire pour.pst-3sg burning=ez heart=pc.3sg low na-me-šud-∅. «Huš=i šavhar=aš ba in duxtar NEG-IMPF-become.pst.3sg intellect=ez husband=pc.3sg to dem.prox girl raft-a bud-a=ast u vay bexabar...»

In saying “Shahnoza is beautiful”, it was as if Nozimi had set her on fire, the flames of her heart did not quiet down. “The attention of her husband had turned towards this girl, and she (= Musharraf) [had been] unaware...” (ibid.: 11)

In (18), Musharraf’s thoughts about her husband are given in a semi-direct way. At the wedding feast, Shahnoza is dancing and invites Nozimi to dance with her. The use of the register II form signals a new realisation on Musharraf’s part. The word Xudo-zad-a ‘goddamned’ further accentuates that the reader is being presented with her personal thoughts.
In (19), Musharraf reads an article that her journalist husband has written about Shahnoza in the newspaper, and the narrator gives us her inferential (or mirative) thoughts, further accentuated by the exclamation mark. The thoughts are given in a semi-direct way, i.e. not verbatim, since the pronouns have been changed.

In (20), Musharraf is upset by the above-mentioned article and sets out to find Shahnoza. She runs into her co-worker Khijolat, who asks how Musharraf is doing. The last line conveys Musharraf’s thought or realisation, supplied in an semi-direct way. With the form bud-a=ast, the narrator signals that this is Musharraf’s realisation that Khijolat was aware of the article. This is also confirmed later in the dialogue, when Musharraf asks Khijolat if she knows about the article.
7.2.1 Non-uses

In (21), Musharraf is visiting Nozimi’s office. She is greeted by Shafii and she infers that the reason he comes forward and greets her is because she is alone. The inference is prefaced by an evidential lexical marker, *az aft=aš* ‘from the looks of it’, but the verb form is of the register I type.

(21) *Salom, kelin,* – bo čehra=i kušod-a nazd=i ū dar omad-∅, 
- hi bride – with face=ez open.pst-pp at=ez she prev come.pst-3sg 
*az aft=aš* az tanho bud-an=i ū jur’at paydo kard-∅ from look=pc.3sg from alone be.pst-inf=ez she courage apparent do.pst-∅ [...] 
- Hello ma’am¹, – [he said] approaching her with a beaming face, apparently emboldened by her loneliness [...] (Mirzo [1989] 2009:18)

In (22), Musharraf remembers how she had become close friends with Shahnoza and brought her to her own home, unaware that this might further increase Nozimi’s interest in Shahnoza. The description below is Musharraf’s thoughts about her own behaviour. We have two lexical markers, *gūyo* and *az aftaš*, but no register II forms. The reason for the use of register I forms in (22) is that no real inference or realisation is being made. Musharraf is merely using these lexical markers figuratively to further emphasise her self-deprecating thoughts.

(22) *Gūyo* *az* dil=i mard-ho bexabar ∅-boš-ad, *az aft=aš*, as.if from heart=ez man-pl unaware subj-cop.npst-3sg from look=pc.3sg on-ho=ro ham xud=aš barin tasavvur me-kard-∅ [...] 
DEM.DIST-PL=DO also self=PC.3SG like imagination IMPF-do.pst-3SG [...] 
It was as if she were ignorant of the hearts of men, apparently thinking of them as being like herself. (ibid.:10)

7.3 Evaluative

Musharraf and her girlfriends see Nozimi for the first time when he comes and visits them at their high school. The direct speech is given in quotation marks, and we see how the *bud-a=ast* acts as a marker of evaluative statements.

---

¹ The word *kelin* is of Turkic origin and means “bride; daughter-in-law”. It is here used as a form of address, like “ma’am” in North American English.
Sodiqi talks at length about Nozimi. The sentence in (24) occurs in the middle of a long exposition that Sodiqi makes for Musharraf and the girls after they had asked him about Nozimi.

(24) [...] bebaho bud-a=ast in bača [...] [...] priceless be.PST-PP=COP.NPST.3SG DEM.PROX child [...] this guy is priceless [...] (ibid.: 26)

Nozimi talks profusely with Musharraf’s aunt and wins her over. We get Musharraf’s thoughts regarding this.

(25) Balo bud-a=ast, Xudo ba vay zabon dod-a. calamity be.PST-PP=COP.NPST.3SG God to he tongue give.PST-PP He was smooth, God had given him a tongue. (ibid.: 35)

After the altercation between Musharraf and Nozimi that resulted from the latter’s comment on Shahnoza’s beauty, Musharraf tries to keep Shahnoza at a distance. Musharraf thinks that Shahnoza is an utterly shameless person. The register II form here (the Perfect of budan^2) is prefaced by “in Musharraf’s opinion”, thus making this an example of a register II form being used in an evaluative sense.

(26) Vale ba aqida=i Mušarraf vay benomus=i guzaro but to opinion=ez Musharraf she shameless=ez passing bud-a=st. be.PST-PP=COP.NPST.3SG But in the eyes of Musharraf, she [= Shahnoza] was an utterly shameless person. (ibid.: 11)

7.4 Epistemic verbs and phrases

7.4.1 Non-uses

In (27), we have an epistemic verb “to hear” which does not trigger the register II form.

---

2. The bud-a=st is a typographical error for bud-a=ast.
Musharraf is at Mavlonzoda’s office. She wants to speak with him but they keep getting interrupted by his telephone. When Mavlonzoda answers the phone he becomes pale. In (28), we get Musharraf’s thoughts regarding who the caller is, prefaced by az aft=aš ‘from the looks of it’. The verb, however, is in a register I form.

In (29), the young Nozimi meets with the principal. Musharraf, who is watching them from afar, finds the principal’s humble behaviour towards Nozimi inappropriate, since Nozimi is much younger than the principal.

In (30), the principal introduces Nozimi to Musharraf and her girlfriends in a way that makes it look as if he is giving an economic report. The gūyo here does not introduce an inference, but a simile. The use of a register I form is thus to be expected.
Likewise, in (31), inferences are made from the actions of the principal, but the statement is given with a register I form.

(31) Nigoh=i mudir az aft=aš, az mion=i duxtar-on view=ez principal from look=PC.3SG from middle=EZ girl=PL
on-ho=e=ro me-just-∅, ki ba vay lozim DEM.DIST-PL=IND=DO IMPF-search.PST-3SG CLM to he necessary
bud-and.
COP.PST-3PL
From the looks of it, the eyes of the principal scanned the girls for the ones that he would need. (ibid.: 26)

Nozimi is complimenting Musharraf. The statement “you are a girl who loves literature” is prefaced by *ma’lum ki ‘it’s obvious’. The copula has been omitted, as is common in colloquial speech.

(32) - *Ma’lum, ki šumo duxtar-ak=i adabiyyot-düst, […]
- obvious CLM YOU.PL girl-DIM=EZ literature-friend, […]
It’s obvious that you [are] a girl who loves literature, […] (ibid.: 28)

### 7.5 The Presumptive

It is springtime and Nozimi has picked a flower for Musharraf. In (33), we have the words Musharraf speaks after ceremoniously accepting the flower. It is not entirely clear why this statement is uttered in the non-past Presumptive. It could be that Musharraf sees receiving the flower as an auspicious sign of how the future will unfold. If this is the case, it is difficult to argue that the Presumptive statement is “an unsubstantiated conjecture or assumption” (Perry 2005: 243), because the Presumptive statement is based on receiving the flower. One could still argue, however, that the statement is unsubstantiated, in so far as it merely expresses a general (inexplicable) feeling about how things will unfold (triggered by the flower), not a logical inference based on in situ observation.

(33) […] gul=i avval=i bahor=ro az dast=i navisanda girift-em,
[...] flower=ez first=ez spring=DO from hand=ez author take.PST-PC.1PL
bo-fayz, bo-barakat me-šud-a-gist-∅ bahor…
with-grace with-blessing IMPF-become.PST-PP-PRESU-3SG spring
[...] I took the first flower of spring from the hands of the author (= Nozimi); spring will (surely) be full of grace, blessed… (Mirzo [1989] 2009: 34)

In (34), Musharraf is thinking to herself. She is wondering how long she and Nozimi will keep meeting in secret. The inner dialogue is given in a semi-direct way. She thinks that they should get engaged and make it official.
It had to be enough, they had become well acquainted with one another. (Mirzo [1989] 2009:39)

7.6 Hypothetical questions
In (35), Musharraf is waiting for Nozimi and asking herself when he will come. The verb form here is the Subjunctive Perfectoid Imperfect.

(35) «Kay me-omad-a ∅-boš-ad?» az xud suol
    when IMPF-come.PST-PP SUBJ-COP.NPST-3SG from self question
    me-kard-∅.
    IMPF-do.PST-3SG
    “When would he come?” she asked herself. (ibid.:32)

7.7 Summary
In the first 35 pages of *Dar orzuyi padar*, we find abundant use of register II forms, the majority of which are used for marking inferences and mirative / evaluative statements. This is mainly due to the narrational style of the author. Two major segments of the text take the form of flashbacks experienced by one of the main characters, Musharraf. Interspersed within these flashbacks, however, are conclusions and comments made by “the main timeline Musharraf”. Moreover, the author favours a narrational style where many of Musharraf’s inner monologues are presented. This type of storytelling generates plentiful instances where the register II forms are employed to mark realisations based on inferences.

The text does, however, also contain several instances where register I forms are used after epistemic verbs and phrases, especially after *az aftaš* ‘by the looks of it’. By virtue of the inference being made in these cases, one would have expected register II forms.
8. Analysis of *Se rūzi yak bahor*

Sattor Tursun was born in 1946 in the Surkhondaryo region of Uzbekistan. The novel *Se rūzi yak bahor* (‘Three Days of One Spring’) was first published in the magazine *Sadoi šarq*, and subsequently printed in 1988. It is considered one of his best works, having a simple language that builds on the tradition of the fathers of Modern Literary Tajik (Luqmon 2018).

The novel depicts the early years of the 20th century and the conflict between revolutionaries and the Basmachi. The protagonist of the novel, Anvar, is the leader of a small band of Tajiks who have aligned themselves with the Bolshevik cause to fight Usmon Aziz, a Basmachi leader.

Unlike earlier works dealing with this historical period, like Rahim Jalil’s *Odamoni jovid*, in Tursun’s novel the Basmachis are not depicted as thoroughly evil or bad people. Its nuanced depiction of the “enemies of the Revolution” initially led to the work not being approved for printing. Tursun refused to alter his work to meet the standards of the time, and instead published the novel at the end of the perestroika years, when Soviet society was beginning to open up (Maxsumzod 2013: 10–11).

For this section, the first 70 pages of the novel were analysed.

8.1 Reportative

Usmon Aziz admonishes his friend Halimboy not to be too attached to the fortunes and riches of the world. To illustrate the vanity of seeking riches, he tells the story of Alexander the Great.¹ The story can be divided into two parts. The first, which provides the background information about Alexander has register II forms (the Perfect and the Perfectoid Imperfect). The second part, which contains the main content of the story, is told using register I forms (the Present tense).

In (1), we find the beginning of Usmon Aziz’s account of Alexander. It is introduced with a lexical marker *dar zamon=aš* ‘once upon a time’ and the use of the Perfect of *budan*.

---

¹ Persian Alexander legends typically identify him with *Ḏu al-Qarnayn* ‘the two-horned one’, a righteous ruler mentioned in the Qur’an (Hanaway 1998).
There once was a world-conquering king named Iskandar Zulqarnayn (= Alexander the Great). (Tursun [1988] 2013:41)

In (2) and (3), the story continues in the Perfect. The sentences below have been truncated to only show the last finite verb in a chain of perfect participles.

(2) [...] čun buzurg-tar-in šoh=i olam šūhrat yoft-a=ast.
 [...] as big-COMP-SUP king=ez world renown find.pst-pp=cop.npst.3sg [...] he was known as the greatest king in the world. (ibid.:41-42)

(3) [...] xud=aš tamom=i umr dar ayš u nūš [...] self=pc.3sg whole=ez lifetime in feasting and drinking
bud-a=ast.
be.pst-pp=cop.npst.3sg [...] he spent his life feasting and drinking. (ibid.:42)

In (4), the sentence is introduced with the lexical marker “they say” and uses the Perfectoid Imperfect to mark a reportative past continuous action.

(4) Me-gūy-and, ki az ša’n u šavkat=i Iskandar hatto
IMPF-say.npst-3pl clm from grandeur and splendour=ez Alexander even
sitora-ho=i osmon xira me-sud-a=and.
star-pl=ez sky dim IMPF-become.pst-pp=cop.npst.3pl
They say that even the stars were dimmed by his grandeur and splendour. (ibid.:42)

As the story progresses and the main plot of the story is told by Usmon Aziz, there is a shift to the Present tense, as can be seen in (5). In other words, the background of the story is provided using register II forms, while the main plot is told in the Present tense.

(5) [...] vay peš az marg muqarrab-on=i vafodor=aš=ro ba nazdi= [...]
 [...] he before from death esteemed-pl=ez trusted=pc.3sg=do to at=ez
xud xond-a, vasiyat me-kun-ad...
self call.pst-pp will IMPF-do.npst-3sg
 [...] before [his] death he summons his trusted friends and makes his
will known... (ibid.: 42)

Anvar’s friend Murod recounts how as a child he asked his father why a certain mountain was called Chilchiroq (lit. “forty lamps”). It is interesting to note that even though his father begins his story with zamon=e ‘once’, he uses the Present tense when describing the enemy’s attack on the Gardon valley. However, when he is describing how the enemy appeared to the people, he
uses the Perfect. In other words, backgrounding is marked with register II, whereas the main event uses a register I form (the Present).

(6) [...] dušman-on hujum me-or-and. On-ho ba monand=i [...] enemy-PL attack IMPF-bring.NPST-3PL DEM.DIST-PL to like=EZ mūr=u malax hisob na-došt-a=and; ba misl=i gurg ant=and locust reckoning NEG-have.PST-PP=3PL to like=EZ wold berahm=u čun zoḡ=u zaḡan noseram bud-a=and. merciless=and like raven=and kite rapacious be.PST-PP=COP.NPST.3PL [...] the enemies attack. They were as innumerable as ants and locusts, as merciless as wolves, and as rapacious as ravens and kites. (ibid.:66)

We see the same kind of interplay between the Present tense and register II forms in (7), which follows directly after (6). The main storyline, which describes how the people of Gardon valley took up arms against the attackers, is told in the Present tense. On the other hand, relative clauses which describe both the Gardon valley people in (7) and their assailants in (8) employ register II forms.

(7) Hama mard-on=i vodi=i Gardon, ki ajjod ba ajjod az all man-PL=EZ valley=EZ Gardon CLM ancestor.PL to ancestor.PL from zahmat=i dehqoni non me-xūrd-a=and, [...] toil=EZ agriculture bread IMPF-eat.PST-PP=3PL [...] All the men in the Gardon valley, who had been earning their livelihood through agriculture from generation to generation, [...] (ibid.:66)

(8) [...] ba on sarboz-on=i yakkačin, ki umr=ašon ba jang u jidol [...] to DEM.DIST soldier-PL=EZ superior CLM life=PC.3PL to war and strife me-guzašt-a=ast, čil šab=u čil rūz dalerona IMPF-pass.PST-PP=COP.NPST.3SG forty night=and forty day courageously korzor me-kun-and. battle IMPF-do.NPST-3PL [...] (they) courageously fight, for forty days and nights, those superior soldiers, who had spent their lives warring and fighting. (ibid.:66)

After (8) and (7), the story continues in the Present tense, describing how some of the Gardon valley inhabitants seek shelter in the mountains. There is, however, a shift to the Perfectoid Imperfect in (9) and (10) when describing how each group would light a fire to inform the remaining valley dwellers of their existence. In these examples, the main clauses are in the Perfectoid Imperfect, but the subordinate clauses are in the Present tense.
During the night, every group would light a bonfire at their home, informing their compatriots who remained in the valley that they are still alive [...]. 

Every night, from forty locations on the mountain, the bonfires of the refugees would say “You come, too” to the valley people [...].

Usmon Aziz remembers what people in his village used to say about the Soviets. The rumours about the rule of the Soviets are introduced with *me-guft-and* ‘they said’, and they are given in the form of a list, each separated with a semicolon and a new line. A total of seven propositions are made, all in register II forms, as subordinate clauses to the introductory *me-guft-and*. Only the first statement is in the Perfect, while the rest are in the Perfectoid Imperfect, which in effect is tense neutral. But the scaremongering quality of these statements makes a future reading most reasonable. Only in certain cases is the reader aided in deciding the correct tense with temporal markers such as *minba’d* ‘from now on’ and *aknun* ‘now’.

As with other instances where other people’s speech is conveyed indirectly, there is an ambiguity regarding the actual usage of the indirective form. The indirective forms could be used either because the villagers themselves used this form when conveying these rumours or because Usmon Aziz decided to use them to further emphasise their reportative nature.

They used to say that the Soviet regime has given men and women equal rights; (ibid.: 84)

Personal ownership will collapse. (ibid.: 84)
Religion and faith will be trampled on. (ibid.: 84)

From now on, the poor Tajiks will only toil for the infidels. (ibid.: 84)

Whoever wants to will freely make love with the woman of another man. (ibid.: 84)

Now, old and young, women and men... will eat from one pot like wandering dervishes. (ibid.: 84)

Islam will be obliterated... (ibid.: 84)

They say that a lot of people in the valley [...] have given all of their belongings to the kolkhoz.

In (19), Amonov is describing the surrounding area. He introduces his statement with güyö ‘it is said that’, but uses a register I form for the proposition.

It is said that no one is checking those borders (lit. those borders are ownerless). (ibid.: 23)
8.2 Inference

Anvar is telling his mother about all of the good things that the Soviet system will bring about. As he tells her how the Soviet cause just recently has gained momentum, and how day by day it is increasing in strength, he is interrupted by his mother who says it is apparent that the Soviet cause will spread without him.

The use of the lexical marker *ma’lum* ‘it is apparent that’ in (20) coupled with the Perfectoid Imperfect signals that the proposition is based on inference. Although future meaning has been chosen in the translation, the inherently ambiguous nature of the verb form makes it possible to interpret it as referring to multiple tenses.

(20)  
– *Ma’lum*, ki be tu ham kor=aš rivoj  
– apparent CLM without you also work=PC.3SG progress 
  me-yoft-a=ast.  
  IMPF-find.PST-PP=COP.NPST.3SG  
– It is apparent that it will succeed without you. (Tursun [1988] 2013: 21)

In (21), one of the characters uses the Subjunctive Perfectoid Imperfect in a conditional clause. Since the verb *alam kardan* ‘to cause trouble’ had been uttered just before the sentence in (21), the Subjunctive Perfectoid Imperfect is probably used sarcastically here.

(21)  
– *Alam na-me-kard-a∅-boš-ad*, dam na-zad-a fur  
– pain NEG-IMPF-do.PST-PP SUBJ-COP.NPST.3SG breath hit.PST-PP PREV  
  ∅-or-∅  
  ∅-boš-ed,  
  subj-cop.npst.3sg  
  xürjin=at=r0,  
  ...  
  ∅-or-∅  
  ∅-bring.npst-2sg saddle=PC.2SG=DO ...  
– If isn’t too much trouble, take off your saddle without saying a word [...] (ibid.: 32)

In (22), the Subjunctive Perfectoid Imperfect is used to state an inference nested in a conditional clause (without *agar* ‘if’). As in (21), the verb form does not seem to carry any strong evidential meaning. Instead, it seems to be used here for rhetorical purposes; this is further emphasised by the use of *masalan* ‘for example’.

(22)  
– Peš-tar, masalan, ba čupon=aton dar yak sol yak gūsfand before-COMP e.g. to shepherd=PC.2PL in one year one sheep  
  me-dod-a∅-boš-ed, holo ehtimol du gūsfand  
  IMPF-give.PST-PP SUBJ-COP.NPST-2PL now possible two sheep  
  ∅-dih-ed.  
  ∅-give.npst-2pl  
– For example, if you gave one sheep a year to your shepherd before, you probably give two sheep now. (ibid.: 39)
In (23), the reader is presented with the experiences and thoughts of a she-wolf that lives in a cave. The first sentence describes how the cave was dark. The second sentence is introduced with *ma’lum* ‘obvious’ and is in the Perfect. In this context, the first sentence (that the cave was dark) gives the reason for the inference in the second sentence (that night had fallen).

(23) Darun=i ğor torik bud-∅. Ma’lum, ki šab faro insideez cave dark COP.PST-3SG obvious CLM night PREV rasid-a=ast. 
arrive.PST-PP=COP.NPST.3SG  
It was dark inside the cave. It was obvious that night had fallen. (ibid.: 77)

8.3 Evaluative

Usmon Aziz is thinking about his friend Halimboy, about his past and present situation. Among these thoughts we have two cases of evaluative utterances, as can be seen in (24) and (25).

(24) Badbaxt bud-a=ast vay, Halimboy... 
unfortunate be.PST-PP=COP.NPST.3SG he Halimboy 
He, Halimboy, is unfortunate. (ibid.: 35)

(25) Xom bud-a=ast, xom! 
uncooked be.PST-PP=COP.NPST.3SG uncooked 
He is naive, naive! (ibid.: 35)

In (26), Aziz is talking about the Russians and mentions some reasons why they are a strong power in the region.

(26) Dono bud-a=and rus-ho. 
knowledgeable be.PST-PP=COP.NPST.3PL Russian-PL 
They are knowledgeable, the Russians. (ibid.: 40)

In (27), Usmon Aziz is scolding one of his companions. The evaluative nature of the expression triggers the Perfect of *budan*.

(27) - Tu sirf ġarča bud-a=i! 
- you.sg solely fool be.PST-PP=COP.NPST.2SG 
- You are a complete fool! (ibid.: 80)

In a longer passage, the reader is presented with the thoughts of a she-wolf who makes two evaluative statements using the Perfect of *budan*:

(28) Či iloj, taqdir čunin bud-a=ast. 
what remedy fate such be.PST-PP=COP.NPST.3SG 
What can be done? Such is fate. (ibid.: 70)
8.3.1 Non-uses

One of the characters is thinking about a woman and we are given his thoughts verbatim. In (30), he expresses what a kind woman she, using no copula. Usually, these type of statements take the Perfect of budan.

(30) Či zan=i mehrubon!

What a kind woman!

8.4 Presumptive

In the middle of the night, someone knocks on the gate of Anvar’s house. Anvar goes to the gate and someone tells him that he should immediately make his way to the Raykom\(^2\) office, because the head of the militia is expecting him. When Anvar returns to his house, he is greeted by his mother, who was also awakened by the knocking at the gate. She asks him who it was that knocked at this hour. Anvar first tells her that it was the person on duty from the Raykom office, but after noticing that this upsets her, he adds that she should not worry, and that an urgent matter must have arisen, like always. The addition of this last part, “like always”, further corroborates the theory that the Presumptive is used for suppositions based on past experience or previous knowledge.

(31) - Šumo beqaror na-šav-ed, oča, yagon kor=i ta’jili bar omad-a-gist-∅. Harvaqta barin…

- Don’t worry, mother, an urgent matter must have arisen. Like always… (ibid.:17)

When Anvar comes back badly bruised and injured after travelling to a remote village to start up a kolkhoz there, his mother scolds him for having gone, saying that there has to be someone older, and thus more suitable for such a mission, in the vicinity.

(32) Dar in muzofot ġayr az tu ham kalon-tar in dem.prox vicinity except from you also big-comp bud-a-gist-∅, oča, yagon kor=i ta’jili bar omad-a-gist-∅, neg-become.npst-2pl mother one work=ez urgent prev come.pst-pp-presu-3sg always like be.pst-pp-presu-3sg emph

---

2. From Russian rajonnyj komitet ‘district committee’.
There has to be someone older than you around here! (ibid.: 20)

Usmon Aziz asks his friend Halimboy how long they have known each other. In (33), Halimboy answers, using the Presumptive, that they must have known each other for 20–25 years.

(33) - Yagon... bist-bist u panj sol šud-a-gist-∅.
- one twenty-twenty and five year become.PST-PP-PRESU-3SG
- It must be some... twenty-twentyfive years. (ibid.: 36)

In (34), Halimboy is pleading with Usmon Aziz to take him along when he leaves, saying that he has some money saved, including coins that Usmon Aziz has given him.

(34) Kam-tar pul dor-am, az in tanga-ho=i xud=aton little-COMP money have.NPST-1SG from DEM.PROX coin-PL=ez self=PC.2PL lutf kard-a-gi ham dor-am. Ba yak kampir=am mercy do.PST-PP-PP, also have.NPST-1SG to one old.woman=pc.1sg rüz=amon me-guzašt-a-gist-∅.
day=PC.1PL IMPF-pass.PST-PP-PRESU-3SG
I have some money, I also have these coins you have graciously given me. With one, my wife [and I] can surely get by. (ibid.: 43)

Usmon Aziz, who is leaving, asks Halimboy where he should take him. In (35), Halimboy says that Usmon Aziz must certainly (albatta) be crossing the river, since they say that he resides in Peshawar. The river mentioned here is possibly the Amu Darya. In other words, we learn that the premises for Halimboy’s conclusion that Usmon Aziz “must be crossing the river” is based on the fact that Usmon Aziz lives in Peshawar.

Note that this example also presents us with an instance of reported speech. However, in the colloquial style in which this is uttered, the subordinate clause introduced by “they say” lacks the copula.

in Peshawar.
Certainly, you must be crossing the river. They said that you [live] in Peshawar. (ibid.: 43)

Anvar and his friends are going from village to village, warning the villagers of the Basmachis. After visiting a certain village, Anvar thinks of the different ways in which the villagers can defend themselves against the Basmachis. In (36), he lists the tools they can use as weapons: shovels, pick-axes, and axes. He also makes the conjecture that a hunting bow can probably even be found.
Anvar’s friend Murod is recounting a conversation he had with his father as a child. His father asks him what one calls people who speak Tajik. Murod answers that they are called Tajiks. His father then asks him “What are you?” In (37), Murod answers that he “must be a Tajik”, using the Presumptive. The Presumptive is here used for logical conclusions based on previous knowledge: “if those people are Tajik on account of their speaking the Tajik language, then that must make me a Tajik, too.”

(37) «Tojik bud-a-gist-am.»
Tajik be.pst-pp-presu-1sg
“I must be a Tajik.” (ibid.: 64)

8.5 Hypothetical questions
In (38), the Subjunctive Perfectoid Imperfect is used to express a hypothetical question expressed by one of the characters of the novel.

(38) To kujo me-gurext-a ∅-boš-em?
to where impf-flee.pst-pp subj-cop.npst-1pl
Where could we flee? (ibid.: 30)

Tamanno, a girl who is in love with Anvar, wonders if he also thinks about her. In (39), we are given her thoughts in a semi-direct way.

(39) Anvar čī — niz ū=ro ba yod me-ovard-a
Anvar what — also she=do to memory impf-bring.pst-pp
∅-boš-ad?
subj-cop.npst-3sg
And what about Anvar—is he also thinking about her?

In (40), Tamanno is wondering when Anvar will come.

(40) «Kay me-omad-a ∅-boš-ad? Kay?..»
when impf-come.pst-pp subj-cop.npst-3sg when
«When will he come? When?» (ibid.: 57)

8.5.1 Non-uses
There are also two instances where Tamanno is wondering when Anvar will come, but the Present Indicative is used instead of the Subjunctive Perfectoid Imperfect.

166
8.6 Proverbs and sayings

We find a number of examples where the Perfect of guftan / farmudan ‘to say’ is used after proverbs and sayings.

(43) – Zabon dar dahon posbon=i sar ast
tongue in mouth guardian=EZ head COP.NPST.3SG

The tongue in the mouth is the guardian of the head, they say (lit. have said). (ibid.: 32)

The sentences in (44) and (45) are uttered in close proximity to one another by one of the characters.

(44) Dar kalom=i šarif ham farmud-a=and, ki banda=i
speech=EZ noble also say.pst-pp=COP.NPST.3PL CLM slave=EZ
musulmon ba šikam=aš boyad sang=i qanoat
Muslim to stomach=PC.3SG must stone=EZ contentment
∅-band-ad.

In the hadith (lit. venerable speech) it is said (lit. they have said) that a Muslim (lit. a Muslim slave) must tie a ‘contentment stone’ to his belly.3 (ibid.: 38–39)

(45) Behuda na-guft-a=and:
čašm=i tang=i
in.vain NEG-say.pst-pp=COP.NPST.3PL eye=EZ narrow=EZ
dunyo-dor=ro yo qanoat pur ∅-kun-ad, yo xok=i
world-have.NPST=DO or contentment full SUBJ-do.NPST-3SG or dust=EZ
gür.

It is not in vain that they have said: “The miserly eye of a worldly person will be filled with either contentment or the dust of the grave”.4 (ibid.: 39)

---

3. According to the hadiths, the Prophet and the early Muslims used to tie stones to their bellies to keep hunger pangs at bay.
4. This saying is from the Golestān of Sa‘di (chapter 3, story 21).
In (46), a character is citing a proverb and appends \textit{guft-a=and} ‘they have said’ to it. Note that the same phrase is repeated in the following sentence.

(46) \begin{align*}
- \text{Ba umed=izudo na-šav-∅,} & \quad \text{butta=ro} \\
- \text{to hope=EZ god} & \quad \text{NEG.IMPV-become.NPST-2SG, shrub=DO} \\
\text{∅-dor-∅} & \quad \text{gutf-a=and.} \\
\text{IMPV-hold.NPST-2SG say.PST-PP=COP.NPST.3PL wonder good} & \quad \text{Ajab nāgz} \\
\text{gutf-a=and.} & \quad \text{say.PST-PP=COP.NPST.3PL} \\
\text{Don’t set your hopes on God’s grace, keep hold of the shrub (i.e.} & \quad \text{God helps those who help themselves), they say. How well they have} \text{said [it].} \quad \text{(Tursun [1988] 2013:42)}
\end{align*}

In (47), we also have a saying with a \textit{guf-t-a=and} appended to it.

(47) \begin{align*}
\text{Zūr=i behuda miyón me-šikan-ad} & \quad \text{guf-t-a=and...} \\
\text{force=EZ in.vain middle IMPF-break.NPST-3SG say.PST-PP=COP.NPST.3PL} & \quad \text{Excessive force is detrimental, they say. (ibid.:45)}
\end{align*}

In (48), no \textit{guf-t-a=ast} or \textit{farmud-a=ast} is appended to the saying. Instead, the adage itself is in the Perfect.

(48) \begin{align*}
\text{Ustuxon=i odamizod az rūz=i azal ba šir=i xom šax} & \\
\text{bone=EZ mankind from dayEZ eternity to milk=EZ raw hard} & \quad \text{šud-a=ast.} \\
\text{become.PST-PP=COP.NPST.3SG} & \quad \text{Since time immemorial the bones of mankind have hardened through} \text{the drinking of raw milk (i.e. to err is human). (ibid.:45)}
\end{align*}

8.7 Summary

The analysis of the first 70 pages of this novel shows us that the use of register II forms is very much in conformity with Ayni’s use of them in \textit{Ahmadi devband}.

As for the reportative use, one of the segments, which concerns a legend about Alexander the Great, is especially interesting because it displays how the introduction of a narrative is done with register II forms (Perfect and Perfectoid Imperfect) whereas the main plot of the story uses a register I form (the Present). The same type of interplay between register I and II forms is also seen in a story about the valley of Gardon. In one section of the text, the reader is presented with several rumours about the future implications of the rise of the Soviet state. All except one of these employ the Perfectoid Imperfect to convey a present-future tense.
Two inferential statements are made with the Subjunctive Perfectoid Imperfect, but the way in which they are uttered indicate that they are used more for rhetorical purposes than evidential.

The frequent use of the Presumptive further corroborates the theory that these are used for suppositions based on habit or how things usually are. In particular, three examples stand out. In (31), when Anvar tries to ease his mother’s worry by saying that “an urgent matter must have arisen”, using the Presumptive and adding “like always...” to this statement, as if to further emphasise that the presumptive statement is based on prior knowledge and experience.

The same kind of reasoning based on prior knowledge, not in situ observations, can be seen in (35), when Halimboy makes suppositions regarding where Usman Aziz is heading. Halimboy presumes that Usman Aziz is crossing the river, since they say that he lives in Peshawar. Lastly, in (37), the speaker draws a conclusion based on previously received information.

The text also presents several evaluative statements, which characteristically use the Perfect of budan to make a judgement about a person or group’s character.

There are also a number of proverbs and sayings which have the Perfect of guftan / farmudan ‘to say’ appended to them. One of the proverbs, however, is without such an appendage. Instead, the last verb of the proverb is in the Perfect.
9. Analysis of Bozgašti padar

The 112 page story Bozgašti padar (‘Father’s Return’) written by Umar Safar in 2014, tells the story of a Tajik farmer Faromurz who receives a letter from Maria, a Russian woman he met while they were both serving in the army during the Soviet–Afghan war (1979–1989). The reason Maria is contacting him is that they have a daughter together, Shahnoza, who now wants to get married and obtain the consent of her father, Faromurz, to get married. The story describes the day when Faromurz receives the letter, and in flashbacks and through Maria’s letter, we get glimpses of what life was like in Afghanistan when they were in the army. The story was analysed in its entirety.

9.1 Reported speech

Faromurz and Maria are discussing the Second World War, and its impact on Soviet families. Maria describes how her grandfather fought and died in the war. Naturally, she did not witness these events herself. The indirect forms she uses while talking about her grandfather signal that this is reported speech. The sentence sarboz bud-a can be interpreted as either a serial verb construction or a finite form with the copula omitted. Either way, its value is that of the Perfect.

(1) 
Az xonadon=i mo bobo=(y)am dar in jang=i xanumonsüz from family=ez we grandfather=pc.1sg in dem.prox war=ez ruinous jon boxt-a=ast. Sarboz bud-a va dar jang-ho=i
life lose.pst-pp=cop.npst.3sg soldier be.pst-pp and in war-pl=ez Stalingrad ištirok kard-a=ast.

From our family, my grandfather lost (lit. has lost) his life in this disastrous war. He was (lit. has been) a soldier and participated (lit. has participated) in the Battle of Stalingrad. (Safar 2014: 29)

Another example of how the Perfect is used for second-hand information is found in (2), where Shahnoza, who has been told by her mother that her father went missing in the war, gives her reasons for being interested in Afghanistan. We can therefore conclude that everything she says in this paragraph is second-hand information.
My father, who went missing in this war, is the reason for my special interest in Afghanistan. He was a high-ranking officer in the military. (ibid.: 69)

9.1.1 Non-uses

If we take Ayni’s Ahmadi devband as a model for when to use the register II forms, there are some instances in Bozgaštī padar where we would have expected them to be used on account of the second-hand information that is being relayed.

In (3), Maria is describing the fate of Ivan Fedrovich, one of her fellow soldiers in Afghanistan. She writes in her letter to Faromurz that he was brutally killed by the Mujahideen. Since Maria introduces her account with the phrase “from what I’ve heard”, we know that her account is based on hearsay. However, the choice of verb form does not mirror the reported nature of the passage, since the Present tense is used for the “main action” whereas the backgrounding is conveyed through the Pluperfect.

Example (4) comes directly after (3), and is the continuation of Maria’s written account of what happened to Ivan Fedrovich and his driver. Here, there is a shift from the Present tense of the previous sentence to the Preterite.
Qaribi-ho=i šom jasad-ho=i on-ho=ro yoft-and.
closeness-pl=ez evening corpse-pl=ez dem.dist-pl=do find.pst-3pl
Qotil-on mošin=i xizmati=i on-ho=ro niz bo xud
murderer-pl car=ez service=ez dem.dist-pl=do also with self
burd-and.
bring.pst-3pl
Around evening, they found their bodies. The killers had also taken
their car with them. (Safar 2014: 61)

In (5), the narrator describes how Faromurz remembers his childhood and a
certain man who was known as a healer among the people. Since this account
is introduced with me-guft-and ‘they used to say that’, we would expect the
verbs to be in register II forms but instead they are in the Present tense.

9.2 Evaluative

The backstory leading up to (6) is that Sharif’s hens have been killed by a fox.
He did not see the attack himself, but his little son reported to him that the
fox was the culprit (ibid.: 47). Sharif ultimately manages to catch the fox and
he wants to kill it, but his friend Faromurz objects to this act, which he deems
to be cruel. In (6), Sharif questions Faromurz’s non-violent stance, saying “it
is a vermin” using the Perfect of budan. Given the context in which this is
uttered, it is very unlikely that Sharif has some kind of in situ realisation that
this animal is a vermin. After all, his reason for catching the fox was to put
an end to the attacks on his hens. The register II form rather denotes that
this is his evaluation, and the Perfect of budan is used for emphasis. It is also
interesting to contrast the use of bud-a=ast with the unmarked negation nest
‘is not’ that comes right after it. The proposition that the fox is a vermin is
marked as being Sharif’s personal view, but the proposition that that the fox
is not human is given as a neutral fact.
(6)  - Eh, aka=i Faromurz, yak gap-ho=i me-gū-i ki?  
- hey older.brother=ez Faromurz one talk-pl=ez ind-say.npst-2sg emph

Dunyo čappa na-šud-a=ast ku?  
world upside.down neg-become.pst-pp=cop.npst.3sg emph one

hayvon=i muẓir bud-a=ast,   
animal=ez harmful be.pst-pp=cop.npst.3sg human cop.neg.npst-3sg

ka, - ranjid-a-tab’ guft-∅ Šarif.  
EMPH - suffer.pst-pp-temperament say.pst-3sg Sharif

“Hey, brother Faromurz, what are you talking about? Is the world upside down, or what? It’s a vermin (after all), not a human being, you know,” Sharif said exasperated. (ibid.: 78)

9.3 Epistemic words and phrases

There are many instances where epistemic phrases such as “it was clear that” is followed by a verb in the Perfect. In all of the cases, however, the use of the Perfect seem to be more a matter of Tajik using the tense of the direct experience rather than something triggered by the epistemic phrase. In (7), we see how the Perfect is used. In English we would have the pluperfect: “It was apparent that he had drunk his fill of sleep.”

(7)  Ravan bud-∅, ki az xob ser šud-a, xub dam   
bright be.pst-3sg that from sleep satiated become.pst-pp good rest

girift-a=ast.  
take.pst-pp=cop.npst.3sg

It was obvious that he had drunk his fill of sleep, he had rested well. (ibid.: 19)

In (8), Faromurz remembers his deployment in Afghanistan and a particular airplane that flew with.

(8)  Sol=i tavlid=i havopaymo ma’lum na-bud-∅, vale az   
year=ez production=ez airplane clear neg-be.pst-3sg but from

zohir va holat=i darun=aš, havopaymo=e bud-∅, ki az   
exterior and state=ez inside=pc.3sg plane=ind cop.pst-3sg clm from

umr=aš sol-ho=i ziod=e guzašt-a=ast.  
life=pc.3sg year-pl=ez many=ind pass.pst-pp=cop.npst.3sg

The production year of the airplane was not apparent, but from [the appearance of] its exterior and its inner state, it was an airplane that had seen many years. (ibid.: 22)

9.3.1 Non-uses

In (9), we have an epistemic verb, natiya giriftan ‘to gather, understand’, which is not used together with the Perfect.
Faromurz from manner=ez speaking=and look=ez Sharif result
girift-∅ ki ü qodir ast ba a’mol=i az in ham take.PST-3SG CLM he capable COP.NPST.3SG to act.PL=ez from DEM.PROX also faje’-tar dast ∅-zan-ad. heinous-comp hand SUBJ-hit.NPST-3SG
By Sharif’s speech and demeanour, Faromurz gathered that he is capable of committing acts even more heinous than this. (Safar 2014: 78)

9.4 Hypothetical questions

In (10), we are given Faromurz’s thoughts as he encounters a man he has never met before. The verb form is the Perfectoid Imperfect in the subjunctive.

In mard=i noošno [...] az ü či me-xost-a
DEM.PROX man=ez unknown [...] from him what IMPF-want.PST-PP
∅-boš-ad
SUBJ-COP.NPST-3SG
What could this stranger want from him? (ibid.: 16)

In (11), Shahnoza is wondering when she will see her father again. The rhetorical nature of this question is further emphasised by it being addressed to God.

Xudo=yo kay in rūz=i nek me-rasid-a
God=voc when DEM.PROX day=ez good IMPF-reach.PST-PP
∅-boš-ad?
SUBJ-COP.NPST-3SG
O God, when will this day blessed day arrive? (ibid.: 66)

In (12), Faromurz is surprised that Sharif wants to kill the fox. Faromurz’s inability to understand where such a cruelty could come from is conveyed through the Perfectoid Imperfect in the subjunctive. The statement is technically not a question, but one can argue that it is a nested question.

But Faromurz could not at all understand where savagery and cruelty of such a magnitude could have its source in a village youth.
9.5 Summary

The use of the register II forms in this text is very sparse compared to other texts, such as Ayni’s *Ahmadi devband* and Mirzo’s *Dar orzui padar*.

We only have two cases where reported information is marked as such with the Perfect tense. The first is when Maria is talking about her grandfather’s participation in the Second World War, and the second is when Shahnnoza talks about her father’s participation in the Soviet–Afghan war. From the context, we know that the information about these two men is not first-hand.

In contrast to this, we also have two cases where lexical cues indicate that the information is second-hand, but there is no marking on the verb. The first case appears in a letter from Maria where she writes about the fate of her colleague Ivan Fedrovich. The account is introduced with the phrase “from what I’ve heard”, and initially the verbs are in the Present tense, but later they switch to the Preterite. The second case is when Faromurz remembers a certain man from his childhood. The proposition is prefaced by “they used to say” and the verb is in the Present tense.

A comparison between the above-mentioned reportative register II accounts with those that use register I shows that the difference between them has to do with temporal distance. Maria’s reportative accounts detailing her grandfather’s participation in the Second World War and her father’s participation in the Soviet–Afghan war lie further in the past than the fate of her colleague Ivan Fedrovich. The same can be said about Faromurz’s memories of the healer. In other words, reportative accounts which refer to actions in the distant past are more likely to have register II forms than accounts referring to events that have taken place more recently.

Perhaps some level of epistemic intimacy is also at play here. One could argue that the reason Maria uses register II forms when describing her grandfather and father is because she never knew them. But she did meet and know Ivan Fedrovich, just as Faramurz knew the healer, and hence register I forms are used with these accounts.
10. Analysis of *Zamini modaron*

The novel *Zamini modaron* (‘Land of Mothers’) was written by Shahzoda Nazarova and published in 2015 in both the Perso-Arabic and the Cyrillic scripts. For the purposes of this dissertation, the Cyrillic version was used, and the book was analysed in its entirety (148 pages).

Nazarova was born in the city of Samarkand, Uzbekistan, in 1975. For the last two decades she has been living in Holland (Nazarova 2015b). Nazarova is a proponent of a unified international Persian language. She foresees a future where citizens of Iran, Afghanistan, Tajikistan and Uzbekistan can communicate with one another without loss of meaning and misunderstandings. Moreover, she views the language in her books as not being a local variety, such as “Samarkandi” or “Tehrani”, but as plain Persian (ibid.).

Nazarova’s wish for a unified Persian language is reflected in the novel *Zamini modaron*. Whereas the lexicon used is decidedly Tajik in nature, the grammar displays a nearly non-existent use of the many perfectoid verb forms that characterise Modern Literary Tajik. Moreover, Modern Literary Tajik prefers not to mark the subjunctive with the prefix *bi*- . However, since this prefix is used heavily in the novel, the style is closer to that of Afghanistan and Iran.

The book is narrated through the eyes of a girl from Samarkand who has suffered amnesia after a period of illness. She is trying to piece her life back together, and to this end she returns to Moscow where she had received in-house treatment.

10.1 Reportative

Mahtob, the protagonist of the story, asks hospital staff if she can retrieve the recorded material from the CC-TV in her room. They inform her that they have never given anyone access to the recorded material before, except for one time when someone had stolen things from the hospital. The reportative propositions in (1) are introduced with *guft-and* ‘they said’, and the verb used is the Perfect. The last sentence, however, uses the Perfectoid Pluperfect.
They said that they had not done such a thing, and they had never opened the archives for anyone. Except for one time, and that was to find a person who had stolen things from the hospital. (Nazarova 2015a:53)

In (2), Mahtob describes how there was so little regard for books during the Soviet era that people made egg cartons out of volumes of Classical Persian poetry. The reportative propositions are prefaced by “some people said” and use the Perfect and the Perfectoid Imperfect.

Some people said they had seen how they made egg cartons out of the poetry compilations of Hafez and Khayyam in the carton factory. (ibid.:87)

10.1.1 Non-uses

At the beginning of the novel, the protagonist relates the manner in which she was born, as told by her mother. The source of information for the proposition, i.e. the mother, is indicated lexically, but there is no use of register II forms for the reported proposition.

But my mother used to say that I pulled myself out without anyone’s help, without a lot of pain and effort [...]. (ibid.:13)
In (4), Mahtob describes a burial ground where the local gardener is buried. It is said that he comes out of his grave at night to take care of the trees growing on the burial mound. The proposition is introduced with “they say”, but a register I form is used.

It is interesting to compare this example with that of example (4) in Ayni’s text, as the subject matter is very similar (supernatural beings coming out at night), and both of the examples are marked lexically as hearsay with either “they say” or “it is said”. Ayni uses register II forms, however, while Nazarova uses register I.

(4) [...] me-gůy-and šab-ho az gůr dar me-o-yad va [...] IMPF-say.NPST-3PL night-PL from grave PREV IMPF-come.NPST-3SG and ba daraxt-on rasidāgī me-kun-ad. to tree-PL tending IMPF-do.NPST-3SG [...] they say that he comes out of his grave at night and tends to the trees. (Nazarova 2015a: 80)

10.2 Inference

Mahtob is visited by Natasha and a man introduced as her boyfriend. But when Mahtob studies his face more closely, she recognises him as an old acquaintance and makes two inferences marked with the Perfectoid Imperfect and introduces them with pas ‘so; in that case’. On the other hand, this example can also be analysed as being mirative.

(5) Pas in hama muddat ba unvon=i düstpisar=i Natasha didan=i so DEM.PROX all time to title=EZ boyfriend=EZ Natasha see.INF=EZ man me-omad-a. Pas dar in hama muddat me-donist-a, I IMPF-come.PST-PP so in DEM.PROX all time IMPF-knew.PST-PP ki man kujo=(y)am. CLM I where=CO.P.NPST.1SG
So, all this time, he had come to visit me under the pretext of being Natasha’s boyfriend. So, all this time, he had known where I had been.

10.2.1 Non-uses

The protagonist is watching her mother, and from the look on her mother’s face, she infers what she is thinking. The register I form is used for the inferred proposition.
Mahtob is worried that Natasha is ill, and she expresses this to the physician. When she mentions that Natasha is crying a lot lately, we learn that the proposition that “Natasha is ill” is an inference based on Natasha’s crying.

“Doctor, I have a feeling that Natasha is also ill; she’s been crying a lot as of late.” (ibid.: 73)

10.3 Evaluation

10.3.1 Non-uses

In (8), a character is describing the protagonist’s mother, and in (9) a character is describing a girl. As we have seen in previous texts, such evaluative statements that refer to the present and are based on a personal impression of the speaker’s impression usually take the Perfect of budan in Tajik (cf. section 8.3). Nevertheless, the Present tense is employed in (8), which lends it an air of generality and makes it less personal.

“Your mother is a wise woman.” (ibid.: 51)

In (9), a character is making a statement about a dead person using the Preterite, just like example (11) on page 137 in the analysis of Se rūzi yab bahor.

“Duxtar=i xub=e good=IND COP.PST-3SG
She was a good girl.” (ibid.: 49)
10.4 Epistemic verbs and phrases

10.4.1 Non-uses

The protagonist describes how her mother used to receive letters but did not appear to read them with any interest. The epistemic verb here is *diqqat kardan* ‘to notice’. The verb in the subordinate clause is in the Preterite.

(10) Goh=e *diqqat me-kard-am* dar xond-an=i on-ho raġbat=e na-došt-∅.
instance=IND notice IMPF-do.pst-1sg in read.pst-inf=ez dem.dist-pl interest=IND neg-have.pst-3sg
I sometimes noticed that she read them without interest. (Nazarova 2015a: 49)

10.5 Summary

The number of register II forms found in this book is rather low considering that it spans 148 pages. This is probably due to the fact that Nazarova prefers to write in a way that is an international Persian that does away with Tajik’s many idiosyncrasies, one of which is its elaborate use of register II forms. This being the case, we do not find any typically Tajik hypothetical questions (i.e. Perfectoid Imperfect in the subjunctive) or the Presumptive.

Likewise, with some types of statements, such as those appearing after epistemic verbs and phrases and evaluative statements, register I forms are used in all the instances we have. The only use-cases for the register II forms are found in reportative and inferential statements—categories that also exist in the other national varieties of Persian and in Classical Persian.
Part IV: Presentation and analysis of the oral sources

This part of the thesis is devoted to analysing audio recordings that were gathered in Dushanbe, Tajikistan, in the years 2016 and 2017. The participants in these recordings were all university students between the ages of 18 and 23. All of them were either from Dushanbe or its surrounding villages.

Compared to the recordings from 2017, the 2016 recordings are freer and less structured in nature, because free conversations were conducted with the participants, either in group sessions or individually. The main purpose of conducting the 2016 sessions was to elicit reportative statements.

Table 10.1. An overview of the two types of recording sessions.

<table>
<thead>
<tr>
<th></th>
<th>Participants</th>
<th>Sessions</th>
<th>Hours</th>
<th>Main elicitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>20</td>
<td>6</td>
<td>~ 5</td>
<td>Reportative</td>
</tr>
<tr>
<td>2017</td>
<td>8</td>
<td>4</td>
<td>~ 4.5</td>
<td>Inferential &amp; presumptive</td>
</tr>
</tbody>
</table>

As for the 2017 sessions, these are of a structured kind, closely following the instructions set out in the *Family Problems Picture Task*, a type of collaborative storytelling exercise where the two participants are tasked with constructing possible storylines from a set of pictures that are given to them in sequence. The exercise was developed as a method for investigating grammatical categories related to social cognition (San Roque et al. 2012: 135).

In their entirety, the pictures can generate multiple storylines, but the pictures are deliberately presented in a predetermined non-chronological order in order to elicit suppositions, doubts, and inferences. At the end of the exercise, the participants are asked to decide on a final storyline by rearranging the 16 picture cards in their preferred order. They are then asked to retell
the story in the third person, as if they were telling it through the eyes of an objective onlooker, and then again in the first person through the eyes of one of the characters in the story (San Roque et al. 2012:135).

Ideally, according to the enclosed instructions, the session should be video-recorded. Moreover, for the last part of the exercise, when the participants narrate their stories, a new person who has not participated in the previous part of the exercise should be introduced in order to partake of their story.

For practical reasons, I did not adhere to the above-mentioned recommendations, even though they might have been conducive to further eliciting evidential statements. First, video-recording the participants could add an extra element of deterrence for people who might already be hesitant to participate in an exercise where their voice is being recorded. Moreover, the participants’ body language was not germane to the study at hand, which is why I opted for only recording the audio. Secondly, as regards introducing a new person for the last part of the exercise, this would have been ideal, but for practical reasons and because of a scarcity of participants, this was not feasible.
11. Recordings from 2016

This section is based on six sessions of recordings made with students in Dushanbe between May 25 and June 1, 2016. The total length of the material is about five hours. Four of the sessions were conducted in groups, and two were done with only me and the student. A total of 17 men and three women between the ages of 18 and 24 participated.

The main purpose of recording the students was to see how the they related reported information regarding events and people, in the past and present. I tried as much as possible to let the students discuss different topics among themselves, without interfering.

11.1 Contemporary events and people

The students generally only use register I forms when speaking about present events and people, even though most of the utterances are decidedly reportative. In (1), one of the participants talks about how his cousin recently visited several different countries around the world. Even though he clearly had not taken part in these events himself, he uses the Preterite with no lexical marking of information source when describing them.

(1) Švetsiya raft-∅, [...] Turtsiya raft-∅.
Sweden go.pst-3sg, [...] Turkey go.pst-3sg
He went to Sweden, [...] he went to Turkey.

In (2), when talking about the contemporary singer Manizha’s sudden withdrawal from the public eye, one of the students use the Preterite with no overt lexical marking of information source. It is not clear if the student makes these statements based on personal observation (i.e. that she herself noticed that Manizha had stopped performing live) or if she is going by other people’s statements.

(2) pēš-tar, to du hazor u nû, to du hazor u before-comp until two thousand and nine until two thousand and hašt=a, saroyanda=i xub bud-∅, mašhur bud-∅. či eight=DAT singer=ez good cop.pst-3sg famous cop.pst-3sg what sabab=e omad-∅ ki, az sahna birun omad-∅. reason=ez come.pst-3sg clm from stage out come.pst-3sg.
Before, until 2009, until 2008, she was a good singer, she was famous. Something happened that made her leave the stage.
When asked why she had stopped performing live, one of the participants says that it perhaps is due to her getting married. This sentence is expressed using the epistemic modal adverb šoyad ‘maybe’ with the Perfect. It is not clear if this statement is based on other people’s reports or the student’s own conjecture.

(3) sabab=aš šoyad šavhar kard-a-gi=ay
reason=PC.3SG maybe husband do.pst-PP-PP₂=cop.npst.3SG
Maybe the reason [is that she] has married.

The students are talking about the contemporary singer Shabnami Surayo. In (4), the Perfect is used without any overt lexical marking of information source, and possibly with a resultative rather than indirective meaning. It is not clear if the statements are based on personal experience (i.e. the student having seen what kind of sponsors she has and concluding that this is the reason for her fame) or reports (i.e. the statements being based on other people’s reports).

(4) hozir xeyli mašhur šud-a=ast čun ki sponsor-ho=i kalon paydo kard-a-gi=ay sponsor-pl=EZ big find do.pst-PP-PP₂=cop.npst.3SG
Now she has become very famous because she has found big sponsors.

One of the students continues describing Surayo, but this time focusing on her current actions. The speaker employs the Present in (5), with no overt lexical marking of information source.

(5) Germanya me-rav-ad u me-saro-yad.
Germany IMPF-go.NPST-3SG and IMPF-sing.NPST-3SG
She goes to Germany and sings.

The students are asked about Agha Khan, the leader of the Ismaili community. The proposition in (6) lacks a copula, but is prefaced by a lexical cue to the information source: “according to the Pamiris”. We are thus presented with an overtly reportative statement that is without a copula.

(6) Oğo Xon ba guft=i pomiri-ho yak xeš u tabor=i Agha Khan to say.PST=EZ Pamiri-pl one self and ancestry=EZ hazrat=i Ali.
excellency=EZ Ali
According to the Pamiris, Agha Khan [is] related to Ali.

Still describing Agha Khan in (7), the Present is employed with no overt lexical marking of information source. The context in which this sentence is uttered, however, makes it clear that it is reportative.
Agha Khan helps the people of Badakhshan a lot.

11.2 Past events and people

Generally, when the students talk about events and people that are not too distant from them temporally, such as life under Soviet rule and the civil war, one can see a propensity for using register I forms prefaced with lexical cues such as “they say that”.

Conversely, when talking about things and events that are more removed from them, such as notable people of the 20th century and events and people in the ancient past, we see more instances of the Perfect and less use of lexical marking of information source.

11.2.1 Soviet times and civil war

In (8), life under Soviet rule is expressed with register I forms prefaced by a lexical cue to the source of information: “the teachers say that”.

(8) muallim-o me-g-an ki zamoni=xub bud-∅.
teacher-PL IMPF-say.NPST-3PL CLM time=EZ good COP.PST-3SG
The teachers say that it was a good time. (continuation also in Imperfect)

In an example from another recording session, we find a very similar statement. In (9), one of the participants says her grandparents say that it was the best of times. In this utterance, we find a lexical marker (they say), and the verb is in the Preterite.

(9) me-guy-and davra=i beh-tar-in bud-∅
IMPF-say.NPST-3PL period=EZ good-COMP-SUP COP.PST-3SG
They say it was the best of times.

Another person chimes in, saying that his grandfather has said that the social conditions were better under Soviet rule. For this proposition he uses the Preterite.

(10) naql me-kard-and ki da zamon=i shrav [...]. shyaroi=i
narration IMPF-do.PST-3PL CLM in time=EZ Soviet [...]. condition.PL=EZ
ijtimoi=i mardum xub bud-∅.
social=EZ people good COP.PST-3SG
They used to say that the social conditions were better during the Soviet times.
In (11), they speak about the civil war using the Imperfect prefaced by “they used to say”. Before making this statement the students said they have no recollections of the civil war.

(11)  me-guft-and  bandit-ho me-omad-∅,  zuri  me-kard-∅.
IMPF-say.PST-3PL  bandit-PL  IMPF-come.PST-3SG  force  IMPF-do.PST-3SG
They used to say that the bandits came, who were violent. (continuation also in the Imperfect)

In (12), when talking about the civil war, register I forms are used.

(12)  barodar bo  barodar me-jangid-∅.
brother  with  brother  IMPF-war.PST-3SG
Brother would fight brother.

11.2.2 Notable people of the 20ᵗʰ century

When the students talk about notable people of the 20ᵗʰ century, there is a greater propensity for them to use the Perfect and not supply any lexical cues to the information source. In two cases we see a shift from the Perfect to the Present, possibly to highlight a key event.

When speaking about Ayni in (13), one of the students begins his account by using the Perfect without any lexical marking of the information source. It should be noted, however, that using the Perfect with the verb tavallud šudan ‘to be born’, as seen in (13), is fairly common in Tajik, even for the first person singular: Man dar sol=i 1992 tavallud šud-a=am ‘I was born in 1992’. However, there is no reason to interpret the Perfect of kalon šudan ‘to grow up’ as being anything other than reportative.

(13)  dar qişloq=i  Giţiuvon tavallud šud-a-gi=ay,
in  village=ez  Gizhduvon  birth  become.PST-PP-PP₂=COP.NPST.3SG
yatim  kalon  šud-a-gi=ay  ham-i  Ayni.
orphan  big  become.PST-PP-PP₂=COP.NPST.3SG  EMPH-DEM.PROX  Ayni
He was born in the village of Gizhduvon, he grew up an orphan, this Ayni.

In (14), another speaker chimes in to give some clarifications using the Present tense:

(14)  dar xurdi  az  padar=aš  mahrum  me-šav-ad.
in  smallness  from  father=PC.3SG  bereaved  IMPF-become.NPST-3SG
He loses his father as a child.

---

1. Note that the subject bandit-ho ‘the bandits’ is in the plural, whereas the verbs are in the singular.
But then there is another shift back to the Perfect in (15), which occurs twice with two different speakers.

(15) maktab=i kūhna=da xond-a-gi=ay
school=ez old=LOC read.pst-pp-pp₂=cop.npst.3sg
He went to the old(-fashioned) school.2

Similarly, when the same people are talking about another well-known author, Bobojon Ghafurov (1908–1977), they use the Perfect with no lexical marking of information source.

(16) Bobojon Ġafurov baro=i baland bar došt-an=i ham-i
Bobojon Ghafurov for=ez high PREV have.pst-inf=ez EMPH-DEM.PROX
tojik-on muarrifi kard-an=i ham-in ki tojik-on maqsad
Tajik-PL introduction do.pst-inf=ez EMPH-DEM.PROX CLM Tajik-PL intent
millat hast-an, kitob=i Tojik-on=ro navišt-a-gi=ay
people cop.npst-3pl book=ez Tajik-PL=DO write.pst-pp-pp₂=cop.npst.3sg
Bобojon Ghafurov wrote (lit. has written) the book Tojikon (‘The Tajiks’) for the purpose of elevating these very Tajiks, to introduce the Tajiks as an ethnic group.

When speaking about the Afghan singer Ahmad Zahir (1946–1979), one of the students introduces the topic by using the Perfect for different types of backgrounding information, as can be seen in (17) and (18).

(17) ham-un qadar=eš me-don-am ki dar Angliya vay
EMPH-DEM.DIST amount=PC.3SG IMPF-know.npst-1SG CLM in England he
xond-a-gi=ay, litse
read.pst-pp-pp₂=cop.npst.3sg upper-secondary.school
xond-a-gi=ay vay.
read.pst-pp-pp₂=cop.npst.3sg he
I only know that he studied (lit. has studied) in the UK, he went (lit. has gone) to upper-secondary school [there].

(18) Padar=aš ham vazir bud-a=y, puldor
father=PC.3SG also minister be.pst-pp=cop.npst.3SG rich
bud-a=y.
be.pst-pp=cop.npst.3SG
As for his father, he was (lit. has been) a minister, he was (lit. has been) rich.

However, when the same speaker as in the last two examples begins to describe Zahir’s untimely death, he switches to a register I form (the Present).

2. Maktabi kūhna is also the name of an autobiographical story by Ayny printed in 1935. The word kūhna ‘old’ is used to describe the school is to contrast it with the new type of school system that the Jadidis, including Ayny himself, proposed in order to implement a new educational system in Central Asia, as described in section 2.7 (Aïnî et al. 1998: 222).
We can thus see how backgrounding of a cohesive narrative is done with register II forms, while the main events are told using register I forms.

(19) dar si u se solagī me-kuš-an=eš in thirty and three at.the.age.of IMPF-kill.NPST-3PL=PC.3SG They kill him at the age of 33.

When talking about Ahmad Shah Masud in (20), the speaker uses the Preterite to describe him, but switches to the Perfect when describing how he defeated the Soviet Union.

(20) vazir=i mudofe’a vay bud-∅, afgoniston. vay Šer=i Panjšer minister=ez defence he COP.PST-3SG Afghanistan he lion=ez Panjshir me-g-an. Šūravi=ra ham-u šikast IMPF-say.NPST-3PL Soviet=DO EMPH-DEM.DIST defeat dođ-a-gi=ay give.PST-PP-PP₂=COP.NPST.3SG He was the minister of defence [in] Afghanistan. They call him the Lion of Panjshir. He was the one who defeated (lit. has defeated) the Soviet Union.

The literature major
One student, who I will refer to as ‘the literature major, since he was majoring in Tajik literature, stands out when speaks about these matters as he uses the Present when speaking about the qualities of the 20ᵗʰ century Tajik poet Loiq (1941–2000). Another thing that sets him apart from the other students is the literary register of his speech and the lack of colloquial verb forms. All of the subsequent examples in this section are from him.

(21) šoir=i muboriza=s, šoir=e ast ki poet=ez endeaevouring=COP.NPST.3SG poet=IND COP.NPST.3SG CLM donanda=i luğat ast, bisyor luğat=i zabon=i tojiki=ro naïz knower=ez word COP.NPST.3SG very word=ez language=ez Tajik=DO good me-don-ad. IMPF-know.NPST-3SG He’s an endeaevouring poet, he’s a poet who knows words, he knows many words of the Tajik language.

When describing Ayni, the literature major also uses the Present.

(22) u xele ravšan-fikr ast, bedor-fikr ast. he very bright-thought COP.NPST.3SG awake-thought COP.NPST.3SG He is very intellectual, he is enlightened.

Later, he uses the Preterite when describing how Loiq was a historian. The Perfect form in the proposition regarding Loiq’s support for the Tajik language is probably used to accentuate the relevance of the proposition rather
than to indicate that it is indirective. The sentences following the ones in (23) are in the Preterite.

(23) ta’rixšinos bud-∅. donanda=i xub=i sath=i zindagi=i mardum historian cop.pst-3sg knower=ez good=ez level=ez life=ez people bud-∅. u az zabon=i tojiki xele puštibonī cop.pst-3sg he from language=ez Tajik very support kard-a=ast. do.pst-pp=cop.npst.3sg He was a historian. He knew the living conditions of the people well. He supported the Tajik language a lot.

In (24), the speaker refers to an article by Loiq in which he describes how the people in Dushanbe speak a mixed language, switching between Russian and Tajik. The speaker introduces his statement with “he said that”, but the proposition uses a Present verb form.

(24) dar yak maqola=šon guft-and ki […] bo lahja=i omext-a in one article=pc.3pl say.pst-3pl clm […] with dialect=ez mix.pst-pp gap me-zan-and. speech impf-hit.npst-3pl In one of his articles, he said that they speak with mixed dialects.

Describing Loiq’s style of poetry, the speaker uses the Perfect when pointing out that he relied heavily on neologisms.

(25) dar še’r=aš ham ustod ba kalimasozi xele muvazzaf in poetry=pc.3sg also professor to neologism very charged gašt-a-gi=ast. become.pst-pp-pp₂=cop.npst.3sg Loiq also used neologisms a lot in his poetry.

When describing the beginning of Loiq’s life, the speaker uses the Perfect with the verb tavallud šudan ‘to be born’. This is expected, however, as this verb is usually in the Perfect when used with a past meaning. The subsequent verbs are in the Present.

(26) dar ham-on rusto tavallud šud-a-gi=ast. in emph-dem.dist village birth become.pst-pp-pp₂=cop.npst.3sg He was born in that very village.

11.2.3 Ancient history

When talking about ancient history and the origins of the Persians and the Tajiks, one of the students introduces the topic by using the Preterite with no lexical cue to the information source.
In that ancient time, the Persians, the Tajiks were like this, little by little did they become separated.

But then one of the participants interrupts the speaker in (27) and uses the Perfect for propositions about the past, introduced with a lexical cue to the information source: “we read in history that...”. In the first two sentences, the Perfect is used without a copula. In the third, the Perfect of budan is used.

In history we read that at first there was the Indo-European, but they split […] in the beginning it was the Indo-European family.

When speaking about the Achaemenid empire, one of the students uses the Perfect of budan, then the Preterite, but when describing how they were defeated by the Greeks, he uses the Present.

The hofiz
One of the students that I interviewed privately prided himself on being a Tajik of Arab ancestry and a hofiz, i.e. someone who has committed the whole of the Qur’an to memory. His speech is of a higher register than colloquial speech, but the verb stems are of the colloquial kind, i.e. me-g-an rather than me-gūy-and. Since I asked him about Islamic history, he would often recite verses from the Qur’an in Arabic to illustrate a certain point.
In (30), he is retelling a story about the Prophet Muhammad. He prefaces the story with “they say” and uses the Preterite for the propositions.

(30) me-g-an ki payombar vaqt=e ki nubuvvatı omad-a IMPF-say.NPST-3PL CLM messenger time=IND CLM prophethood come.PST-PP bud-∅, ba ziyorat=i yak, bi-g-em ki yak a’mo, kur, COP.NPST-3SG to visitation=EZ one subj-say.NPST-1PL CLM one blind blind me-raft-an. IMPF-go.PST-3PL

They say that the Messenger, after he had reached the state of prophethood, used to go and visit a blind man.

The word me-g-an ‘they say’ is repeated throughout the story, whereas the verb form remains in the Imperfect. Since the acts in (31) are repeated, it makes sense to express them in the Imperfect. Later in the story, the speaker uses the Present and the Imperfect.

(31) gadoī me-kard-∅ va payombar me-g-an har ruz begging IMPF-do.PST-∅ and messenger IMPF-say.NPST-3PL every day ba’d=i bomdod ba ziyorat=i ham-on a’mo me-raft-an. after=EZ morning to visitation=EMPH-DEM.DIST blind IMPF-go.PST-3PL

[The blind man] used to beg and the Messenger, they say, used to go and visit that very blind man after the break of dawn.

In (32), he is describing what happened after the death of Muhammad. The speaker uses the Preterite:

(32) Hazrat=i Abu Bakr az Oiša pursid-and ki […]

his.holiness=EZ Abu Bakr from Aisha ask.PST-3PL CLM

Abu Bakr asked Aisha…

In (33) the speaker uses the Preterite when describing how the Prophet Muhammad was known for his moral uprightness and honesty before he began to receive revelations.

(33) payombar peš az nobovvatı ba axloq=i xud va basiddiqi=i xud messenger before from prophethood to moral=EZ self and honesty self mashur bud-∅.

famous be.PST-3SG

Before receiving revelation, the Messenger was renowned for his moral uprightness and honesty.

In (34), the speaker narrates the Hijra in the Present.

---

3. The speaker first uses a word of Arabic origin, a’mo, and then the native Tajik word kūr.
4. The Hijra, also spelled Hegira, is the emigration of Muhammad from Mecca to Medina in 622 (Watt 1971: 366).
11.3 Narrative jokes and mirative expressions

While I was socialising with an acquaintance and his two friends, I noticed that one of them told a joke using only register II forms. Since I had my recording device with me I asked if he could retell the joke so that I could record it, a request he accepted. At the time of recording he was 20 years old. Although fluent in Tajik, he was of Uzbek origin and spoke Uzbek with his family. He grew up in the city of Shahritus in southwestern Tajikistan.

In (35), we have the introduction to a narrative joke, which is told using register II forms throughout. Only the first few sentences are given here. Note that the only time we find the Perfectoid Imperfect used in this chapter is in this joke.

(35) yak boy **bud-a=ast** [...] duxtär=aš one rich be.PST-PP=COP.NPST.3SG [...] daughter=PC.3SG
    **bud-a=ast** [...] bad vay šavhar be.PST-PP=COP.NPST.3SG [...] later he husband
    **me-dod-a=ast**.
    IMPF-give.PST-PP=COP.NPST.3SG
There was a rich person [...] He had a daughter [...] Then he married her off.

The same person tells a story about how he went fishing with a friend. The speaker thought that he got a fish on the hook. The verb form of the proposition is in the Present.

(36) fiKR kard-am mohi ast.
    thought do.PST=1SG fish COP.NPST.3SG
I thought it was a fish.

But then his friend points out that it actually is a snake. The friend’s realisation, being a mirative statement, uses the Perfect of **budan**, a register II form.

(37) “mohi qapid-am!” guft-am. “ne, un ki mor fish catch.PST-1SG say.PST-1SG no DEM.DIST EMPH snake
    **bud-a=ast!”**
    be.PST-PP=COP.NPST.3SG
“I caught a fish!” I said. “No, that’s a snake!”
11.4 Summary

The participants in these recordings tend to use only register I verb forms when describing contemporary people, even when they have not witnessed the events associated with these people themselves. They also do so when describing what other people have told them about Soviet times and the civil war. There is, however, a key difference between the utterances made for these two time periods; there is namely a higher instance of lexical markers, such as "they say", when the students talk about Soviet times and the civil war than about contemporary times.

On the other hand, we see a higher frequency of the Perfect\(^5\) when they talk about notable people from the 20\(^{th}\) century than when they talk about Soviet rule or the civil war. It is difficult to draw any definite conclusions regarding the reason why they use the Perfect in these instances. First, in and of itself the Perfect is charged with a weak evidential value. Even though the Perfect of *budan* has somewhat greater evidential markedness, one must not draw overly far-reaching conclusions about this. Secondly, the use of the Perfect in these examples is not consistent, as it is not uncommon for the speakers to shift verb forms many times within just a few seconds.

As we have seen above, the Tajik speakers display a propensity to use the register I forms of the past tense, i.e. the Preterite and Imperfect, for contemporary and not-too-distant events. It seems as if the more removed the speakers are from the subject in time and space, the more likely it is for them to choose the Perfect. The Perfect occurs most frequently when they speak about temporally distant subjects, such as the 20\(^{th}\) century author Ayni or the origins of the Indo-Iranian languages. We find an exception to this rule in the recordings of “the Tajik major” and “the hofiz”, who almost exclusively used register I forms, probably because they were talking about people and events with which they were intimately familiar. If this is the case, then the degree of certainty, or epistemic intimacy, may be another factor that plays into the choice of verb form.

\(^5\) This is also the case for the colloquial Perfect in -a-gī. This speaks in favour of it being no different than the Perfect in -a as regards its evidential status.
12. Recordings from 2017

This chapter is devoted to the analysis of four recordings made in Dushanbe, Tajikistan, at the end of May in 2017. Each session included two male participants between the ages of 19 and 25. Roughly four and a half hours of recorded material were analysed for this chapter.

The Family Problems Picture Task was used for these recordings. As outlined in section 1.2 and in the introduction to part IV, the participants are tasked with describing a set of 16 pictures. After receiving all of the pictures, they are asked to rearrange them into a story as they see fit. They are then asked to narrate the story, first in a third person perspective, and then in a first person perspective through the eyes of any character of their choosing. In all but one of the recordings, the participants chose the main protagonist as their narrative locus for the first person narrative. In one recording, however, the participant chose to narrate the story through the eyes of the protagonist’s drinking friend.

12.1 General findings

In the overwhelming majority of cases where the participants make suppositions or inferences regarding the pictures, they use register I verb forms, usually the Present. The participants frequently introduce their suppositions with “I think that”, as in (1).

(1) dar in aks, yak, ba fikr=i man yak oila=e
     in DEM.PROX picture one to thought=ez I one family=IND
     hast [...]
     cop.npst.3sg [...] In this picture, a..., I think there is a family.

When the participants are in the process of describing what is transpiring in the pictures, it is not uncommon for them to hedge their statements using “maybe”, with a verb in either the Present Indicative, as in (2), or the Present Subjunctive, as in (3).

(2) šoyad barodar=i kalon=i u ast
    maybe brother=ez big=ez he cop.npst.3sg [...] Maybe this is his older brother.
Another common phrase which the speakers use to introduce a suppositional statement is “as can be seen in this picture”, which is used irrespective of the verb form. We see the use of this phrase in (4), with a “maybe” inserted after a proposition in the Present.

(4) boz dar in aks čunin did-a me-šav-ad bo again in DEM.PROX picture thus see.PST-PP IMPF-become.NPST-3SG with in-ho irtibot=e na-dor-ad šoyad DEM.PROX-PL relationship=IND NEG-have.NPST-3SG maybe
Again, we see in this picture that he doesn’t have any relationship to them, maybe.

When the pictures are described as parts of a larger narrative, there are often several tense shifts. In (5), selected passages from such a narrative are presented to give an idea of what these tense shifts can look like. The participant is going through the pictures and pointing at them while commenting. He starts out in the Preterite and switches to the Present when describing how the man was sent to court. He finishes the story by saying that the man was sent to prison, using the Perfect. The use of the Perfect here seems not to have an inferential meaning but rather a resultative one: he has gone to prison, and therefore he is in prison.

(5) Narrative. Words in between the examples have been removed for a more lucid presentation.

a. in=ro afsar-on=i pulis či kard-and, burd-and
The police officers, what did they do, they took him.

b. dar in-ja muhokima me-š-a
He is tried here.

c. ana in čunin fikr-ho=ro dar nazar dor-ad va
And here, he has these thoughts in his mind, and he’s in prison (lit. he has gone to prison).
12.2 The resultative Perfect

The Perfect is commonly used when making suppositions and inferences. This form can function as a marker for inferences and suppositions, especially if the inferences or suppositions together make up a series of verbs that are all in the Perfect. But we also see the Perfect being used for past actions which are still relevant in the present.

We find an example of this in (6), where the participant is describing picture 5, which shows a bruised woman in a court-like setting talking about being hit. Opposite her sits a man with a ball and chain around his ankle. He is holding his head in his hands, as if in remorse.

![Figure 12.1. Picture 5, titled Court room.](image-url)
Another example of this use of the Perfect can be seen in (7), where the participant is describing picture 9, which shows a man being dragged away by the police, leaving behind a bruised woman and a child.
12.3 The inferential Perfect

In some cases, the use of the Perfect is used in a way that stands out, since it does not seem to be emphasising the current relevance of a past action. In (8), the participant is describing picture 10, which depicts a man in chains with two thought balloons above his head. One shows him being hit by the police, the other one shows him looking through a barred window.

Figure 12.3. Picture 10, titled Thinking of jail.
Later, the same speaker as in (8) again uses the Perfect when speaking about the man’s thoughts. The use of the Perfect of guzoštan ‘to place’ is not out of the ordinary, however, because this is a dynamic-stative verb the Perfect of which denotes a present state. Moreover, use of the third person plural when referring to a third person singular is not uncommon in Tajik, as it is considered more polite.

In (10), however, the same speaker as in (8) and (9) continues to describe the picture in and once again states that the man is thinking, this time using the Present tense.

As soon as the speaker in (10) has uttered this sentence, he is interrupted by his co-participant, who uses the Present with this verb, as can be seen in (11).

The Perfect of the verb fikr kardan ‘to think’, also occurs in another passage where the participants talk about picture 16, which depicts a man in prison thinking about coming home to a happy family. In (12), the participant de-
scribes the picture and uses the Present and later the Perfect prefaced by šoyad ‘maybe’ to express that the man is thinking.

Figure 12.4. Picture 16, titled Thinking of home.

(12) fikr me-kun-ad, andeša me-kun-ad, dar in-jo thought IMPF-do.NPST-3SG thought IMPF-do.NPST-3SG in DEM.PROX-place hast, ya’ne xayol-ot=aš ast. šoyad fikr COP.NPST.3SG i.e. imagining-PL=PC.3SG COP.NPST.3SG maybe thought kard-a=ast man az mahbas bar ∅-o-(y)am ki do.PST-PP=COP.NPST.3SG I from prison PREV SUBJ-come.NPST-1SG CLM yak zindagi=i xub došt-a ∅-boš-am ya’ne fikr one life=ez good have.PST-PP SUBJ=COP.NPST-1SG i.e. thought kard-a=ast.
do.PST-PP=COP.NPST.3SG

He is thinking, he is thinking, it’s here, i.e. his thoughts are [here]. Maybe he’s thinking (lit. has thought) “if I get out of prison I will have a good life”, i.e. he is thinking (lit. has thought).
12.4 The inferential Perfectoid Imperfect

One participant is talking about picture 16, which depicts the man in prison thinking about coming home to his family. In one of his statements, he uses the Perfectoid Imperfect. Considering the fact that this is the only Perfectoid Imperfect found in the recordings, and that the finite verb forms prior to it are in the Preterite, this instance of the Perfectoid Imperfect might be a slip of the tongue.

(13) šoyad in da yagon jo=i dur bud-∅, omad-∅,
maybe DEM.PROX in one place=ez distant COP.PST-3SG come.PST-3SG
omad-∅,
šoyad ay in yagon one jo=i place=ez
dur distant
bud-∅,
cop.pst-3sg
dumad-∅,
come.pst-3sg
i
šoyad in dem.prox
maybe in yagon one jo=i place=ez
dur distant
bud-∅,
cop.pst-3sg
dumad-∅,
come.pst-3sg
i
xel david-a pešvoz=iš girift-an, šoyad ay
DEM.PROX manner run.PST-PP reception=PC.3SG take.PST-3PL maybe from
kor me-omad-a bega
work IMPF-come.PST-PP evening.

Maybe he was in a far away place, he came, and they received him running like this, maybe he was coming home from work in the evening.

12.5 Hypothetical narratives in the Perfect

In one of the recordings, we find several coherent passages that serve as hypothetical narratives in which the Perfect is the preferred form for presenting the hypothetical storyline.

Picture 2 shows a man in uniform handing over a folded pile of clothes to the main character of the story.

One of the participants discussing picture 2 uses a series of Perfects to create one of these coherent hypothetical narratives. The only verb forms that break the series of Perfects are those found in the direct quotes, as uttered by the characters in the story. The participant begins his story with the sentence in (13), using mumkin ‘(it is) possible’. The verbs in this statement are in the Perfect, and a resultative reading makes sense, since the participant is giving reasons for why the protagonist is visiting the police (as he sees it).

(14) in šaxs mumkin yagon pisar=aš=ro, yagon barodar=aš=ro
DEM.PROX person possible one son=PC.3SG=DO one brother=PC.3SG=DO
az az dast dod-a=ast, gum kard-a=ast va....
from hand give.PST.PP=COP.NPST.3SG hidden do.PST.PP=COP.NPST.3SG and
It’s possible that this person has lost a son or a brother.

As the story progresses in (15), the speaker continues his explanation of why the man has come to the police, still using the Perfect except when quoting direct speech.
Figure 12.5. Picture 2, titled Receiving clothes.

(15) dar pulis murojiat kard-a=ast ki “barodar=am in police returning do.PST-PP=COP.NPST.3SG CLM brother=PC.1SG nest-∅” yo “pisar=am nest-∅”. 
    NEG.COP.NPST-3SG or son=PC.1SG NEG.COP.NPST-3SG
He has gone to the police, [saying] “My brother is missing” or “My son is missing”.

The hypothetical narrative continues in the Perfect, with the participant suggesting that the policeman has the body of the relative and informs the man that he is dead. Note that the word mumkin is repeated again here, as if to emphasise the hypothetical nature of this scenario.

(16) in jasad=aš=ro yoft-a, inja nišon
    DEM.PROX corpse=PC.3SG=DO find.PST-PP here sign
    na-dod-a=ast ammo libos-ho=i ham-on
    NEG-give.PST-PP=COP.NPST.3SG but clothes-PL=EMPH-DEM.DIST
    šaxs=ro nišon, ya’ne in patinka nišon dod-a=ast.
    person=DO sign, i.e. DEM.PROX shoe sign give.PST-PP=COP.NPST.3SG mumkin guft-a O-boš-ad “barodar=at vafot kard-∅” possible say.PST-PP SUBJ-COP.NPST-3SG brother=PC.2SG death do.PST-3SG yo “pisar=at vafot kard-∅” or son=PC.2SG death do.PST-3SG
He has found his body, he has not shown it here but he’s shown the clothes of that person, that is, he has shown the shoes. It’s possible that he has said “Your brother died” or “Your son died”.

In (17), we have another hypothetical narrative in the Perfect. The participant is discussing picture 4, which depicts two men drinking alcohol on a bench. From the speech balloon, we understand that the one to the left is telling the one to the right about a woman and a man talking. The man to the right is visibly upset. In (17), the participant presents his hypothesis, which he introduces with “maybe” and expresses using the Perfect.

(17) šoyad i ham duxtar=e=ro, i ham dust
maybe dem.prox also girl=ind=do dem.prox also friend

došt-a-gi=ast [...] in mard duxtar=e ham dust
have.pst-pp-pp₂=cop.npst.3sg [...] dem.prox man girl=ind also friend

došt-a-gi=ast
have.pst-pp-pp₂=cop.npst.3sg

Maybe this one has also liked a girl [...] this man has also liked a girl.

The participant then goes into a longer hypothetical narrative about the man and the girl he liked. He says they cannot have children, and this is the reason they divorced.
When they separate (lit. have separated) this one becomes (lit. has become) a vagabond.

The participant in (18) is still talking about picture 4. In (19), the participant presents a hypothetical scenario introduced with mumkin ‘(it is) possible’, using the Perfect throughout.

There is an interesting shift in verb forms when the participants are presented with picture 12, which depicts how the main character is beating a woman. In the background stands a shocked onlooker. The first account that describes the picture is given within seconds after seeing it. As can be seen in (20), the participant is using the Present to give a straightforward account of what the picture shows.
When he hits, when he drinks alcohol and comes and beats his wife, her child falls from her hands, the eye of the child becomes blind.

Directly after (20), the participant points to the man in the background and says he is the one who reported the beating to the police. For this proposition he uses the Perfect.
This one, this one has reported it to the police.

The participant then gives a hypothetical account in the Perfect, prefaced by *mumkin*. The gist of the account is that while the man was in prison, his wife talked to someone else, and his friends informed him about it upon his release from prison. The man drinks alcohol, loses his self-control and beats his wife. Notice that the segment contains a quotation of direct speech in the Preterite. The segment continues with more verbs in the Perfect, but for brevity’s sake only the beginning is given here.

It’s possible that this was (lit. has been) his spouse [...] his friend had said (lit. has said) “she did this, she did that” and this one drank (lit. has drunk), i.e., alcohol, i.e. wine, he lost (lit. himself), he came (lit. has come), i.e. he hit (lit. has hit) her.

### 12.6 Hypothetical questions

Hypothetical questions only occur when the participant speaks from the point of view of the protagonist in the story. The majority of such questions are formed with the Subjunctive Perfectoid Imperfect:

(23) oyo kudak=i man čunin holat=ro *me-did-a* ⊕-boš-ad?
    q    child=ez  such  state=DO IMPF-see.PST-PP SUBJ-COP.NPST-3SG
Will my child see such a state?

(24) ki bar=i u in guna mehruboni *me-kard-a* ⊕-boš-ad?
    who  for=ez he  DEM.PROX manner  kindness  IMPF-do.PST-PP
    SUBJ-COP.NPST-3SG
Who will show such kindness to him?
(25) oyo agar man zindoni ə-šav-am, zer-i şikanja=i q if I prisoner subj-cop.npst-1sg under=ez torture=ez kormand-on=i pulis qaror me-girift-a ə-boš-am? 
officer-pl=ez police position impf-take.pst-pp subj-cop.npst-1sg
If I get incarcerated, will I be subjected to torture by the police officers?

(26) farzand-on=i man [...] hamsar=i man, ma-ra ba xursandi=i ki maqṣad child-pl=ez I [...] spouse=ez I I=do to joy=ez clm purpose ma me-xoh-um pešvoz me-girift-a ə-boš-and?
I impf-want.npst-1sg reception impf-take.pst-pp subj-cop.npst-3pl
Will my children [...] and my wife receive me with, like, the kind of joy that I want?

(27) fikr kard-a istod-a=ast ki maqṣad “Ma-ra I=do ham-in xel xursand pešvoz me-girift-a emph-dem.prox manner happy reception impf-take.pst-pp ə-boš-and?” subj-cop.npst-3pl
He is thinking, like, “Will they receive me with joy like this?”

(28) padar=am u zan=am u kudak=am man ham-in father=pc.1sg and wife=pc.1sg and child=pc.1sg I emph-dem.prox xel pešvoz me-girift-a ə-boš-an yo ki manner reception impf-take.pst-pp subj-cop.npst-3pl or clm na-me-girift-a ə-boš-an.
neq-impf-take.pst-pp subj-cop.npst-3pl
Will my father, my wife and my child receive me in this way or will they not?

Interestingly, we also find the Subjunctive Perfect used with the same intent. In (29), the participant first uses the Perfectoid Imperfect in the Subjunctive, but for the hypothetical questions that follow, he uses the Subjunctive Perfect.

He was thinking “If I go home, how will they receive me? [...] Will my son remember me? Will my spouse remember me?”
12.7 Gnomic statements

During the version of the story told in the first person, the speaker is describing how he starved in prison and was very hungry. A general statement about how one acts when one is hungry is then added, using the Perfectoid Imperfect and the Perfect of budan.

(30) gurusna me-mond-am,  bisyor azob me-kašid-am,  dar hungry IMPF-remain.PST-1SG very torture IMPF-draw.PST-1SG in vaqt=e in sond gurusna me-mond-a=ast har time=IND CLM human hungry IMPF-remain.PST-PP=COP.NPST.3SG every čiz=ro xurd-an=aš mumkin bud-a=ast. thing=DO eat.PST-INF=PC.3SG possible be.PST-PP=COP.NPST.3SG

I went hungry, I was suffering greatly. When one is hungry, one can eat anything.

12.8 Perfect after epistemic verbs

In one instance, when the speaker is telling the story from a first person perspective, we find the verb “to see” with a following proposition in the Perfect.

(31) yak ruz az roh me-guzašt-am va did-am ki on duxtar, one day from road IMPF-pass.PST-1SG and see.PST-1SG CLM DEM.DIST girl ba čid-an=i meva-jot-ho şuğl varzid-a=ast. to pickPST-INF=EZ fruit-PL-PL occupation cultivate.PST-PP=COP.NPST.3SG

I was walking down the road one day, and I saw that the girl was (lit. has been) busy picking fruit.

Directly after the sentence in (31), the speaker adds the following explanation in the Present, introduced with ya’ne ‘that is’.

(32) ya’ne meva-jot-ho=ro jam’gir me-kun-a. i.e fruit-PL-PL=DO collecting IMPF-do.NPST-3SG

That is, she’s gathering the fruit.

12.9 Indirectives in the first person version of the story

We find two interesting cases of a shift to the Perfect in the part of the task where the participants are asked to retell the story from a first person perspective. Given that these chains of Perfects only occur when the locus of narration has not witnessed the events, it is very probable that they are being used in an indirective sense, i.e. that the propositions are based on reports or inferences. The nature of this indirectivity is difficult to tease out, as no lexical cues are given in the examples.
In (33), the participant is telling the story from the perspective of the man in the drama. He uses the pictures in the order that he and his partner have agreed is the best. In (33), he is telling the story using picture 12 (see p. 205). The participant uses the Preterite when describing the abuse.

(33)  u=ro  zer=i  šikanja  va  lat  u  kub  qaror  dod-am
    he=DO under=EZ torture and blow and strike position give.pst-1SG
    I tormented and beat her.

However, after the sentence in (33), the participant switches to the Perfect and describes how an old man saw the abuse and reported him to the police. This shift probably occurs because the character who the speaker is portraying did not see the man do this, but heard it from others or made an inference to this effect. The events in (33) were personally witnessed by the speaker and are thus in register I forms, whereas those in (34) are indirective.

(34)  ba’dan  piramard=e  dar  pas=i  mo  bud-a=ast  va  mo=ro
    later  old.man=IND in  behind=EZ we  be.pst-pp=cop.npst.3sg and we=DO
    did-a=ast  ba  kormand-on=i  pulis  xabar
    see.pst-pp=cop.npst.3sg to  officer-pl=ez police news
    dod-a=ast.
    give.pst-pp=cop.npst.3sg

Then, there (allegedly / apparently) was an old man behind us, who (allegedly / apparently) saw us and (allegedly / apparently) reported [me] to the police.

Directly after the sentence in (34), the speaker switches back to the Preterite when narrating what happened to him personally. The participant thus starts out using the Preterite to describe how he hit the woman, changes to the Perfect when recounting how the old man saw it and reported him to the police, and then switches back to the Preterite to continue his story.

In (35), we find another use of the Perfect when retelling the story in the first person. The speaker is using the Preterite to describe how he fell ill in prison. But when describing how his brother found out about it and went to his wife to plead, he uses the Perfect, probably because he did not personally witness this. However, he switches back to the Preterite when narrating how his wife did not grant his brother’s plea.
bemor gašt-am, in xabar ba guš=i barodar=am sick become.pst-1sg dem.prox news to ear=ez brother=pc.1sg rasid-a=ast, barodar=am ba nazd=i xonum=i man reach.pst-pp=cop.npst.1sg brother=pc.1sg to at=ez wife=ez I omad-a=ast az u xohiš kard-a=ast ki come.pst-pp=cop.npst.3sg from she wish do.pst-pp=cop.npst.3sg clm “barodar=i man bisyor bemor ast, boyad u=ro az mahbas brother=ez I very sick cop.npst.3sg must he=do from prison ozod ∅-kun-em” ammo u ki bisyor azob kašid-a free subj-do.npst-1pl but she clm very torment draw.pst-pp bud-∅ va ranj=i bešumor az man did-a cop.npst-3sg and suffering=ez uncountable from I see.pst-pp bud-∅, qabul na-kard-∅. cop.npst-3sg acceptance neg-do.pst-3sg I fell ill. This news reached my brother. My brother went to my wife and pleaded, saying “My brother is very ill, we have to get him out of prison.” But she, who had suffered a lot, and suffered tremendously at my hands, did not agree.

12.10 Summary

With the vast majority of suppositions and inferences, the participants rely almost exclusively on lexical cues. We have some rare cases where they use the Perfect in isolation, but these are most likely used with a resultative meaning, and there is no reason to believe that they have any indirective meaning.

On the other hand, when the Perfect is not used in isolation, but in a chain of consecutive events in a hypothetical narrative, the use of the Perfect\(^1\) does not lend itself well to a resultative reading. Rather, it seems to be used throughout the narrative in order to mark it as a coherent hypothetical turn of events.

In the first person narrative part of the exercise, we find two parts of the story where the Perfect is used with an indirective meaning. When the participant who has assumed the character of the man in the story gives his perspective on the unfolding of events, he repeatedly uses the Perfect for events that were not witnessed by him. The first time this happens is when he is describing the scene in which he is beating his wife. His own acts of domestic violence are conveyed with the verb in the Preterite. However, as soon as talks about the old man standing behind him, who he suspects reported him to the police, he uses the Perfect. Likewise, later in the first person narrative, when the participant is describing his days in prison, he switches

---

1. The colloquial Perfect in \(^{-a-gī}\) is also used for these type of hypothetical narratives. This further corroborates the theory that they function with the same kind of evidential meanings as the Perfect in \(^{-a}\).
from the Preterite to the Perfect to describe how his brother got news of his illness and went to his wife to plead on his behalf. The likelihood that there is an indirective quality to these Perfects in the first person narrative is very high, considering that they occur successively with different verbs in a series of events which the main character did not witness himself. The Perfect does not occur like this anywhere else in the first person narrative. However, since no lexical cues are given with these narratives, it is not possible to ascertain whether they are inferential or reportative. In any case, they are no doubt indirective in nature. Moreover, they show that first person narratives can be used to successfully elicit indirective accounts.

We also find a few cases attesting that forms we previously only have seen in the written language also exist in the oral language, namely, the use of the Perfect and the Perfectoid verb forms for hypothetical questions and gnomic statements.

Hypothetical questions only occur when the participant speaks from the point of view of the protagonist in the story. The majority of these questions are formed with the Subjunctive Perfectoid Imperfect, but in one example the participant also uses the Subjunctive Perfect to the same effect.

We also have one case of the Perfect occurring after the epistemic verb “to see”, but since this is the only such instance in the recordings, we cannot draw any definitive conclusions from it.

The Perfectoid Pluperfect appears only once, and not in a setting with other indirective verb forms. In any case, the scarcity of this verb form in the material is to be expected, since it is rather rare to begin with.

Interestingly, we find no uses of the Presumptive in these exercises. An explanation of this can be found in the fact that the inferences and suppositions made in the Family Problems Picture Task come from referring to the pictures themselves, not prior knowledge or “how things usually are”, which is the alleged basis for statements using the Presumptive.
Part V:
Concluding discussion
The literary sources consist of texts written by six different authors who published their work during three different periods: the 1920s—1940s, the 1980s, and the 2010s. The older texts were written by Ayni ([1928] 1948) and Jalil (1949). The middle period is represented by Mirzo ([1989] 2009) and Tursun ([1988] 2013). The most recent period is represented by two stories that were analysed in their entirety, by Safar (2014) and Nazarova (2015a).

A total of 424 pages were analysed for this part of the study, but as can be seen in table 13.1, these pages are not evenly distributed across the authors. In fact, the more recent the works become, the more pages that have been analysed. This is because of the scarcity of register II forms in the more recent texts, which in part is due to the subject matter of the novels and the narrative techniques of the newer authors, but is also because there seems to be a general shift towards less use of register II forms.

Table 13.1. A cursory summary of all of the texts in part II, showing the instances of the Presumptive and the non-resultative use of the Perfect and the Perfectoids, the total number of analysed pages, and the instances per page ratio. Note that not all of the instances in Ayni’s text were given in chapter 5, but they have been included in this table.

<table>
<thead>
<tr>
<th>Author</th>
<th>Instances</th>
<th>Pages</th>
<th>Ratio (instances per page)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayni</td>
<td>67</td>
<td>28</td>
<td>2.39</td>
</tr>
<tr>
<td>Jalil</td>
<td>12</td>
<td>41</td>
<td>.29</td>
</tr>
<tr>
<td>Mirzo</td>
<td>31</td>
<td>35</td>
<td>.89</td>
</tr>
<tr>
<td>Tursun</td>
<td>45</td>
<td>70</td>
<td>.64</td>
</tr>
<tr>
<td>Safar</td>
<td>9</td>
<td>112</td>
<td>.08</td>
</tr>
<tr>
<td>Nazarova</td>
<td>7</td>
<td>148</td>
<td>.05</td>
</tr>
</tbody>
</table>

Naturally, there is no reason to believe that all of the possible non-resultative functions in which the register II forms are used (such as reports, inferences, miratives, evaluatives, gnomic statements etc.) will occur with the same frequency in all of the texts. As a rule, there are more indirective uses of the register II forms in the older texts than in the newer texts. For example, the register II forms for reportative statements are so abundant in Ayni’s text that the analysis in chapter 5 would have been too repetitive if they had all been included. On the other hand, if the same number of pages from the more recent texts had been analysed as in the older texts, not enough cases of the register II forms would have been found to get a comprehensive view of their usage.
The use of register II forms to signal reportative propositions is found in all of the literary sources, and they are especially prevalent in texts from the older and middle periods. For example, in texts from the older and middle periods we find many cases where the Perfectoid Imperfect is used with reportative meaning, whereas we only have one such case in the newer texts. Instead of relying on register II forms to express reportative propositions, there is a tendency in the newer texts to use lexical markers in conjunction with register I forms.

We find the most consistent use of register II for reportative propositions in the work of Ayni. As a rule, the reportative use of register II forms can occur without any lexical markers of reportativity. But in Ayni’s Ahmadi devband, the inverse is not true. In other words, we do not find any cases where lexical markers of reportativity are used in conjunction with register I forms (except when someone’s speech is quoted verbatim). In the newer texts, however, we find several examples of lexical markers of reportativity occurring in conjunction with register I forms.

Register II forms used for propositions based on inferences can be found in all works except those of Jalil (1949) and Safar (2014). In the majority of cases where register II forms are used, there is no lexical cue. Instead, the reader has to rely on context to determine that the statement is inferential rather than reportative. Cases where there are lexical cues, such as “it is clear that...”, have been analysed as a separate category, but they could in fact be analysed as inferential statements with lexical cues. Nearly all of the cases where register II forms follow epistemic verbs and phrases can be found in Ayni’s text.

Ayni’s text is the only one that consistently marks the acquisition of new information (i.e. classic mirativity) with register II forms, albeit in a so-called “third person” mirative sense.

The use of register II forms in evaluative statements can be found in all of the texts except those of Jalil (1949) and Nazarova (2015a). The latter contains evaluative statements, but uses register I forms with them.

The pragmatic extension of the register II verb forms, as found in the evidential question, the hypothetical question, sarcasm, and gnomic statements, are for natural reasons, namely the restricted settings in which they occur, not as prevalent, as for example, the register II forms used for reportative purposes. Only one case of the evidential question was found, namely in Ayni’s text in the form of the Perfect of budan. The hypothetical question, however, is more common. This is a type of open-ended question, taking the form of the Subjunctive Perfectoid Imperfect, which is used for rhetorical effect. No instances of this were found in three of the texts, namely those of Ayni ([1928] 1948), Jalil (1949), and Nazarova (2015a). Sarcastic use of the Perfect is found in Ayni, and both Ayni and Jalil display sarcastic use of the Presumptive. The only text to display register II forms for gnomic statements is that by Ayni, which has one such case with the Perfect of budan. It could
be argued that the use of *guft-a-and* in Jalil’s text is a variant of the Perfect for gnomic statement. The difference is that *guft-a-and* is used when the gnomic statement already has a finite register I verb, or where the copula is omitted.

The Presumptive is consistently found in all of the written sources for Tajik, except in the texts of Safar and Nazarova, both of which belong to the newer period. The reason for this may be that *boyad* ‘must’ has supplanted the Presumptive in expressing conjectures, thus making the Tajik verbal system, as used in these stories, closer to that of Classical, Afghan, and Iranian Persian.
The participants in the recordings from 2016 tend to use only register I verb forms when describing the actions of contemporary persons. The register I verb forms are also used when they describe what other people have told them about the Soviet period and the civil war. One difference between their descriptions of contemporary people and of the events of the Soviet period and civil war is that there is a greater presence of lexical markers for the information source in the latter, with speakers usually introducing their propositions with “they say” or similar constructions.

We see a higher frequency of the Perfect\(^1\) when participants talk about notable people from the 20\(^{th}\) century than when they talk about Soviet rule or the civil war. It is difficult to draw any definite conclusions regarding the reason why they use the Perfect in these instances. First, the Perfect in and of itself carries very little evidential value. Even though the Perfect of \textit{budan} has somewhat more evidential markedness, one cannot draw far-reaching conclusions regarding this matter. Secondly, the Perfect is not used consistently in these examples as it is not uncommon for the speakers to shift verb forms many times within a span of seconds.

In other words, the Tajik speakers display a propensity to use the past tense register I forms, i.e. the Preterite and the Imperfect, when talking about contemporary and not-too-distant events, even if they have not witnessed these events themselves. However, the more removed in time and space the speakers are from the subject, the more likely it is for them to choose the Perfect. However, we find an exception to this rule in the recordings of “the Tajik major” and “the hofiz”, who almost exclusively used the register I forms, probably because they were talking about people and events with which they were intimately familiar, due to a personal interest in their respective topics.

The above findings regarding temporal distance being a trigger for register II forms very closely mirror observations made in Uzbek, which is also part of the “Great Evidential Belt” and is a language that has influenced Tajik greatly. Speakers of Uzbek prefer to use the simple past tense -\textit{di} when referring to “events that their grandparents could have witnessed”, whereas the Uzbek Perfect in -\textit{gan}, which, like the Tajik Perfect, can carry an indirective meaning, is used in “the vast majority of references to historical figures and events” (Straughn 2011: 67). According to Straughn, the reason for this is that temporality is but one of many factors that inform the choice of past tense in Uzbek.

\(^1\) Including the colloquial Perfect in -\textit{a-gi}. 
We also have two cases displaying the indirective use of the Perfect in the 2017 recordings. These occur when one of the participants assumes the role of the man in the story and tells the story from his perspective. The likelihood that there is an indirective quality to these Perfects is very high, considering that they occur consecutively with different verbs to form a series of events which the main character did not see. This use of the Perfect is especially remarkable considering that the Perfect is not used anywhere else in the first person narration segment of the recordings.

The vast majority of suppositions and inferences about the pictures in the *Family Problems Picture Task* are made with register I forms. However, in one of the recordings we find several consecutive passages that serve as hypothetical narratives where the Perfect\(^2\) is used in a consistent manner throughout the segment. The Perfect here seems to be the preferred form for binding together a hypothetical storyline that spans across several sentences. Since these hypothetical storylines are based on cues gleaned from the pictures, they may very well be based on inference.

Only one mirative statement, in the form of the Perfect of *budan*, could be found, when a participant tells a personal story about how he thought he had caught a fish, but it turned out to be a snake.

Hypothetical questions, which usually take the form of the Subjunctive Perfectoid Imperfect, appear many times during the first person narrative of the 2017 recordings. They are used to express hopes and fears about the future.

By the very nature of the *Family Problems Picture Task*, which is mainly centred around eliciting suppositions and story-telling techniques, some lesser used instances of the register II forms are not elicited at all. Among these are the register II forms as used in “evidential” questions, sarcasm, and after epistemic verbs and phrases.

We also find a few cases which attest that forms we previously have only seen in the written language also exist in the oral language, namely the use of the Perfect and the Perfectoid verb forms for hypothetical questions and gnomic statements.

As regards the Presumptive, no instances were found in the recordings. This could be interpreted as further corroborating the hypothesis that the Presumptive is used when making inferences based not on sensory input from the situation at hand, but rather on past experiences and “general knowledge”, which act as premises for reaching a conclusion. Naturally, all of the conclusions made about the pictures are based on sensory input (vision). There are thus no conducive parameters for eliciting the Presumptive in this particular exercise, and one will have to rely on other exercises to gauge how prevalent the Presumptive is in spoken Tajik.

\[^2\] Including the colloquial Perfect in -a-gi.
15. General conclusions—revising the Tajik verb system

Modern Literary Tajik as we know it today is a language that has been re-created many times in the last hundred years. Before the Russian conquest, its predecessor was the literary language of a bilingual elite in Bukhara who commanded Tajik and Uzbek equally well. Through Soviet policy Modern Literary Tajik became a vehicle for building a nation state with a new capital, Dushanbe, as its epicentre in the newly formed Republic of Tajikistan. Since the new language was based on the Bukhara variety of Tajik, it was replete with grammatical constructions and calques of Turkic origin, which set it apart from not only the Classical Persian from which it had sprung, but also its sister varieties of Persian, as written in Iran and Afghanistan. Two of these distinctly Turkic features of Tajik were codified into the new language and have been the centre of study for this dissertation. These are the augmented evidential uses of the register II forms and the Presumptive.

The aim of this dissertation has been to study the evidential uses of the register II forms and the Presumptive as they occur in written and spoken Tajik. The novels and stories that serve as the written material can be divided into three periods, an old (1920s–1940s), a middle (1980s), and a new period (2010s). Generally, texts from the older and middle periods display greater use of the register II forms and the Presumptive than the more recent texts.

There are several possible explanations for this. One is that the text by Ayni, which is the oldest text, was obviously written with the intent and purpose of being a model text of the new language, and by a native speaker of the variety of Tajik that came to be the normative national language.

At the other end of the spectrum we have Nazarova, a Tajik who, like Ayni, was born in Uzbekistan—outside the territory of present-day Tajikistan. Nazarova’s personal conviction that the Persian language should be uniform is reflected in her text, which is stripped of the many Turkic features that make literary Tajik language stand out in comparison with the Persian of Afghanistan and Iran.

The conservative attitude of the authors of the older and middle periods with regard to the use of the register II forms and the Presumptive, possibly

1. As we have seen in section 4.1.2, Iranian and Afghan Persian also display evidential uses of the register II forms. The difference is that the forms are confined to the past in these varieties, whereas the Tajik Perfectoid Imperfect has taken on tense-neutral reportative and inferential meanings. Moreover, the mirative and evaluative functions of the register II forms are exclusive to Tajik Persian.
has a political explanation, as the Soviet society in which these authors lived did not tolerate any deviation from the standard. Moreover, there was little contact with Persian speakers from Iran and Afghanistan, except for some increased cultural contact with communist Afghanistan in the 1970s, and later, the deployment of Tajik military personnel and journalists to Afghanistan in conjunction with the Soviet invasion and subsequent war.

Another possible explanation of why we see less of register II and the Presumptive in the newer texts is that the Tajik civil war (1992–1997) led to a significant change in the demographics of Dushanbe, with large numbers of people from the southern part of the country settling in the city. As far as lexicon is concerned, we know that the southern dialects of Tajik have considerably fewer words of Turkic origin than the northern dialects. It is not unreasonable to suppose that this low degree of influence from Turkic languages also prevented the verbal system of the southern dialects from becoming as elaborate as the northern ones, and that the vernacular speech of the new citizens of Dushanbe also affected the literary language.

The same political and demographic changes can possibly also help explain the low frequency of evidential use of the register II forms in the oral material. The speakers display a clear preference for register I forms and lexical markers when making reportative statements about current and not too temporally distant events. There is, however, a tendency for the speakers to use the Perfect without lexical markers when describing events that took place further back in time. Since similar uses can be seen in spoken Uzbek, it is not unreasonable to think that this is an areal phenomenon of the oral variety of the languages, while the written norm, as exemplified by Ayni’s text and continued by other Soviet authors, displays a degree of conservatism in its approach to matters of “seen and unseen”. In other words, we may have two different norms for how the register II forms function depending on whether the language is spoken or written, with the trend being that the written language is becoming more similar to the spoken language, with decreasing reliance on register II verb forms and more use of lexical markers to express source of information.

As regards inferences in the spoken language, the speakers prefer to employ lexical markers together with register I forms. However, the Perfect is employed to create a coherent “hypothetical narrative”, where the speaker narrates a hypothetical series of events. This way of telling a story, using the register II forms used from beginning to end to form a complete story, is very similar to how narrative jokes are told in Tajik and Turkic languages, where the use of the indirective throughout the joke is very common.
15.1 The Presumptive

The findings of the Presumptive in the written sources suggest that this verb form is in fact evidential. It is used for an “unsubstantiated conjecture or assumption” (Perry 2005:243). Perry goes on to describe the verb form as being “inferential by implication”, incorporating within itself a meaning of *boyad* ‘must’, and often lacking an overt or alluded source of information (“evidence”, in Perry’s terminology).

Judging from the analysed texts, all of Perry’s descriptions of the Presumptive ring true. We see that the Presumptive is often used with either an overt or covert reference to the speaker’s experience or a logical conclusion. For example, one of the characters in *Se rūzi yak bahor* admonishes his mother not to worry about his being summoned to a meeting late at night, as “an urgent matter must’ve arisen. Like always...” The addition of “like always” in this example gives us a hint that these summonings have happened before, and that the reasons have previously been of the sort mentioned.

In another example from the same novel, we see how the Presumptive is used for conclusions based on previous knowledge. Using the terminology of logic, the past experiences or previous knowledge that a speaker relies upon to make a statement with the verb form in the Presumptive are called *premises*. The *conclusion* based on such premises is expressed in the Presumptive. Example (35) overtly expresses one of those premises—that the other person lives in Peshawar. The other premise is alluded to—that anyone who wishes to travel to Peshawar must cross the river. The conclusion based on these two premises are thus in the Presumptive, coupled with the lexical marker *albatta* ‘surely’. The premises and the conclusion can be presented as follows, where \( P = \text{premise} \) and \( C = \text{conclusion} \): \( P_1 \) You want to go to Peshawar. \( P_2 \) One has to cross the river when going to Peshawar. \( C \) You will cross the river.

In yet another example from the same novel, we see how a little boy is questioned by his father about his ethnicity. After first giving the boy the premise that those who speak the Tajik language are called Tajiks, the father asks the son what that makes him. The son answers “I must be a Tajik”, using the Presumptive. The premises and the subsequent conclusion can be presented as follows where \( P = \text{premise} \) and \( C = \text{conclusion} \): \( P_1 \) Those who speak the Tajik language are Tajiks. \( P_2 \) I speak Tajik. \( C \) I am Tajik.

Since the Presumptive seems to be used for propositions that essentially are conclusions based on either past experiences or previous knowledge, I argue that this particular verb form should be classified as evidential. Going back to the different types of evidentials that exist, as outlined in section 3.3, the Presumptive fits the criteria of the “presumptive” category, which is re-

---

2. “Unsubstantiated” is a rather infelicitous term in this regard. What Perry probably meant is that it is not substantiated by sensory evidence, but by past experience and knowledge, which serve as premises for reaching a conclusion.
served for propositions based on the speaker’s knowledge about the world and their capacity to draw logical conclusions. In other words, the Presumptive type of evidentiality is indirect and mental, being based on premises that are not observed in situ, as opposed to the inferences we find marked with the register II forms.

15.2 The evaluative and the mirative

As we have pointed out on page 88, the Tajik “mirative” is essentially of two kinds. One is for marking new information (i.e. classical mirative), and the other is for marking personal opinion (evaluative). The reason for separating the two notions is because the latter is functionally different from the classical definition of mirativity in that its core function is to convey the speaker’s direct personal impression. For this reason, the term evaluative has been used in this dissertation in an attempt to better describe its usage. Moreover, since evaluative and mirative utterances with register II forms are always based on direct experience, I argue that they are evidential. Tajik thus has two strategies for referring to directly experienced sources of information. The majority of direct experiences are conveyed through register I forms, as this is the unmarked or neutral way of formulating such statements. However, should the speaker wish to emphasise the personal impression made by a direct experience, or the fact that it is new knowledge, s/he can use a register II form (usually the Perfect of budan) to this effect.

15.3 Evidential categories in Tajik

With the knowledge of how evaluative, mirative, and presumptive statements function in Tajik, figure 3.3 on page 60 can be updated as follows:

![Evidential categories in Tajik](image-url)

*Figure 15.1. A representation of the evidential categories belonging to the direct / indirect dichotomy, adjusted for Tajik. e/m = evaluative / mirative. Adapted from Willett (1988) and Plungian (2010), with some terminological adjustments.*
In this classification, the Presumptive is analysed as an evidential verb form. This is because it scopes over propositions that are inferences based on the speaker’s past experiences or knowledge, rather than sensory, in situ sources of information. This non-sensory quality of the Presumptive could also explain why no instances of the Presumptive were found in the recorded material. Most if not all of the inferences made during the *Family Problems Picture Task* are based on the pictures, and thus are sensory in nature. In order to elicit the Presumptive, one would need to devise an exercise that forces the participants to make inferences based on previous experiences or knowledge, rather than sensory input.

Evaluative and mirative statements made with register II forms are also classified as evidential in nature. The reason for this is that the use of register II forms in statements with present tense meaning is a marked way of conveying personal impressions or new information gained by direct access. For the majority of direct experiences, which are not tinged with a personal evaluation or the acquisition of new information, register I forms are used.


5. ASHK DAHLÉN, Deciphering the meaning of revealed law. The Surūšian paradigm in Shi’i epistemology. 2001.
27. PADIDEH PAKPOUR, Identity Construction. The Case of Young Women in Rasht. 2015.
34. ERIK ANONBY & ASHRAF ASADI, Bakhtiari Studies II: Orthography. 2018.
38. NOZHAT AHMADI (SARRAF) & MOHSEN MOHAMMADI FESHARAKI (Eds.), A Versified History of Shah Abbas the Great [Futūḥāt-i Gītīsitān]. 2021.
41. ALEXANDER NILSSON, Evidentiality in Tajik. 2022.