THE UNIVERSITY
of EDINBURGH

Deconstructing Mermaid Constructions:
From Control and Raising in Tatar

Anonymous Submission

Master of Science by Research
University of Edinburgh
31 August 2022

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(Anonymous Submission)

Abstract

This thesis provides descriptive and theoretical syntactic analyses of the infinitival constructions in Tatar, particularly with a focus on the ones co-occurring with nominal predicates. Then, based on the analysis in Tatar and evidence from other languages, it deconstructs the category of the so-called Mermaid Constructions and reanalyzes them as part of raising and control. The bulk of the syntactic analyses in this thesis is based on the Minimalist approach of generative grammar, though fundamental evidence shown is descriptive and neutrally applicable to various theories of syntax. Upon the analysis of the Mermaid Constructions at the syntax-semantics interface, we employ Glue semantics to derive semantic composition. The scrutiny of the Tatar infinitival constructions with adjectival and nominal predicates reveals that they consist of raising and control and, in particular, that the ones with a noun predicate are Mermaid Constructions. The In addition, multifaceted evidence from syntax, semantics, and diachronic observations further provides a cross-linguistically generalized claim that the canonical Mermaid Constructions are part of control and raising with a nominal predicate, thereby arguing for their biclausal structure and refuting Tsunoda's (2020a) monoclausalist hypothesis.

Keywords: syntax, Tatar, infinitives, grammaticalization

Contents

List of Figures

5

List of Tables

Introduction

Infinitives in Tatar exhibit several syntactically interesting behaviours. As in many languages that have an infinitive verb form, infinitival complements in Tatar involve classic grammatical phenomena such as control and raising. The contrast between sentences (1 and 2) gives rise to some intriguing issues of control and raising. They are both given the same free translations, and both consist of an infinitival complement (*joqla-rɣa*) and a main predicate (*tiješ/kiräk*) with an AGENT (3SG). A crucial difference is that the AGENT is marked by nominative in (1) whereas it is marked by dative in (2).

(1) *ul   joqla-rɣa tiješ.*
    3SG.NOM sleep-INF obligatory
    'S/he must to sleep.'

(2) *aŋa   joqla-rɣa kiräk.*
    3SG.DAT sleep-INF necessary
    'S/he must sleep.'

    Related to this, infinitival constructions include a modal expression without a main modal predicate as in (1). This has a peculiar structure in the sense that there is no finite predicate that bears tense and person (of the subject) by agreement. In addition, certain nouns can fill the slot of the modal predicates in (1) and (2) as in (2) and (3).

(1) *aŋa   joqla-rɣa*
    3SG.DAT sleep-INF
    'S/he must sleep.'

(2) *aŋa   joqla-rɣa röxsät*
    3SG.DAT sleep-INF permission
    'S/he is allowed to sleep.'

(3) *anïŋ   joqla-rɣa isäb-e*
    3SG.GEN sleep-INF thought-POSS.3
    'S/he is going to sleep.'

    Given these observations, the following closely related questions arise:

1. What makes the difference in case-marking of the agent between (1) and (2)?

2. Why is there no finite predicate in (1)? It seems to be non-compositional.

3. Why is the agent marked by dative in (1)? What assigns the case?

4. Sentences (2) and (3) also seem to lack a finite predicate. What is their structure like?

5. The AGENT is marked by dative in (2), while it is in genitive in (3). What causes this difference?

6. Are these sentences monoclausal or biclausal?

An obstacle to tackling these questions is the lack of theoretical research in Tatar syntax. While there are a number of descriptive grammars of Tatar (Burbiel, 2018; Poppe, 1961; Zäkiev & Xisamova, 2015), none of them tackles the problems enumerated above. Therefore, this master's thesis aims to provide the descriptive report of the raising and control constructions and other deeply related constructions of Tatar. In addition, it will show a theoretical account for the intriguing behaviors of the Tatar infinitival constructions with non-verbal predicates illustrated above. In particular, I will demonstrate the following points.

- *tiješ* is a raising construction predicate, while *kiräk* involves control. The nominative case-marking of *tiješ* comes from the case assignment by the matrix T, and the dative case-marking of *kiräk* is assigned as the internal GOAL of *kiräk* (answer to Question 1).

- The bare infinitive construction involves a null adjectival predicate that is responsible for the modal meaning and dative case-marking of AGENT (answer to Question 2 and 3).

- Infinitive with a noun in (2) and (3) is an instance of mermaid construction, where the noun functions as a predicate (answer to Question 4).

- The difference in case-marking of AGENT between (2) and (3) is due to the difference in the structure of nP, which ultimately originates from grammaticalization (answer to Question 5).

- These constructions are biclausal (answer to Question 6).

Example sentences used in this thesis are, unless citation is not shown, sentences verified as grammatical by a native speaker.

Chapter 1

Overview of the Tatar Language

Tatar (Cyrillic: *татар теле*, Latin: *tatar tele*) is a Turkic language spoken mainly in the Republic of Tatarstan, Russia (cf. the location is illustrated in Figure 1.1), which acknowledges Tatar as one of its two official languages along with Russian. The language is spoken by approximately 7 million speakers [CITATION!!! Wikipedia?]. The speakers' geographic distribution is widespread; Tatar-speaking communities exist in neighbouring territories such as the Republic of Bashkortostan, Mari El, Udmurtia, etc. as well as farther regions and cities in Moscow and Siberia. Furthermore, there are also diaspora communities in other countries such as Kazakhstan, Uzbekistan, Finland, Turkey, the United States, China, etc. (Sakurama-Nakamura, 2018). Most Tatar speakers are bilingual with Russian, and frequent code-switching and code-mixing are observed in daily speech of urban Tatars (Wertheim, 2002).

This chapter outlines the language in the aspects of the language family, phonology, morphology, syntax, and sociolinguistics. The first section 1.1 overviews the language in the context of the Turkic language family. Next, section 1.2 summarises the characteristics of the Tatar phonology with a special focus on vowel harmonies. Section 1.3 discusses the morphology of Tatar, in particular the agglutinative derivations and inflections by affixation. Then, most importantly, section 1.4 provides a short overview of the Tatar syntax, which we will further discuss in later chapters on raising and control. Readers with good knowledge of Turkic linguistics may skip the following sections.

1.1   The Turkic Language Family

Turkic languages are widely distributed across the Eurasian continent as shown in Figure 1.2. Turkic languages are generally divided into four groups with two language isolates as summarised in Table 1.1. The map in Figure 1.2 is coloured so that languages in the same subgroup are represented by similar colours. Red-based areas including Turkish, Azeri, and Turkmen represent the Oghuz language group, green for Kipchak languages, light blue for Karluk languages, purple for Siberian Turkic languages, and orange for Chuvash, the only living language of the Oghur Turkic language group. The genetic classification is based on morphophonological variations and commonalities (Johanson, 2021); for example, Oghuz languages are distinct from other groups in terms of the loss of

Figure 1.1: The location of the Republic of Tatarstan (in red) in the Russian Federation. The territory of Moscow Oblast is shown in blue for reference.

G in the deverbal marker suffix -GAn (e.g., Azeri *qal-an* vs. Tatar *qal-ɣan*). Kipchak languages have a distinct feature where they drop the final velar consonant of the denominal suffix -LIG (e.g., Azeri *dağ-lıq* vs. Tatar *taw-lı*).

Group	Languages
Oghuz (Southwestern)	Turkish, Azeri, Turkmen
Kipchak (Northwestern)	Kazakh, Tatar, Kyrgyz, Bashkir
Karluk (Southeastern)	Uzbek, Uyghur
Siberian (Northeastern)	Sakha (Yakut), Tuvan, Altai, Khakas
Oghur (isolate)	Chuvash
Arghu (isolate)	Khalaj

Table 1.1: Categorization of Turkic languages. The distribution of the listed languages are illustrated in Figure 1.2.

Tatar is classified in the Kipchak language group of the Turkic language family. According to Johanson (2021), Kipchak languages can be further categorised into four subgroups as listed in Table 1.2. Tatar and Bashkir, both North Kipchak languages, are close enough to each other to share some
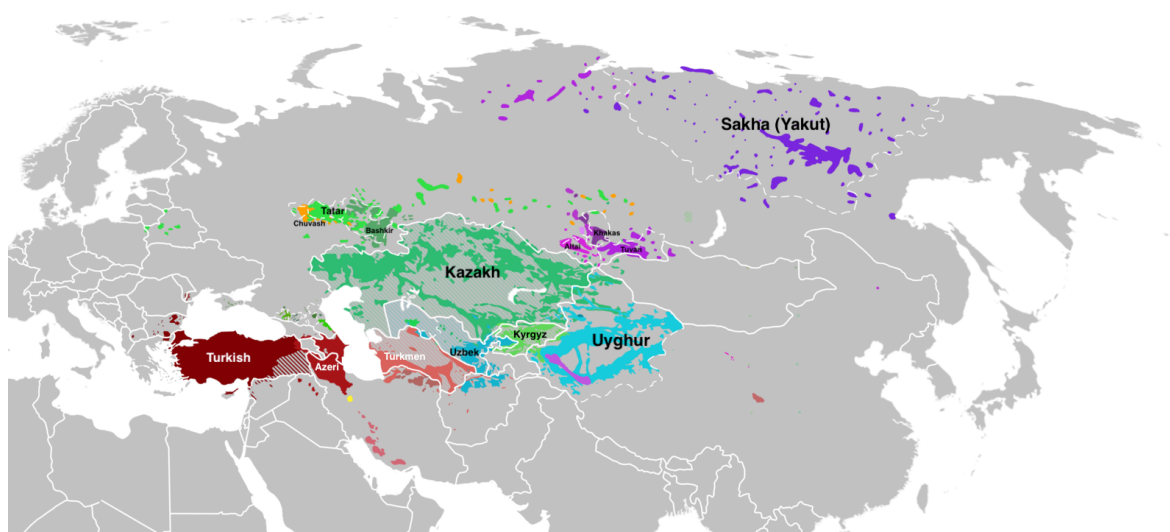
Figure 1.2: The distribution map of Turkic languages. The language names on the map were added by the author. Note that the map only partially names languages, and there are more Turkic languages that are mentioned neither in the figure nor in Table 1.1. The original map, retrieved on 15 May 2022, was originally created by GalaxMaps and is licensed under the Creative Commons Attribution-Share Alike 4.0 International license.

mutual intelligibility [1]. Tatar has three dialects: Kazan Tatar (central), Mishar Tatar (western), and Siberian Tatar (eastern), of which Siberian Tatar is at a controversial status whether it is a dialect of Tatar or a language belonging to a different subgroup.

Subgroup	Languages
West Kipchak	Karachay-Balkar, Kumyk, Crimean Tatar, Karaim
North Kipchak	Tatar, Bashkir
South Kipchak	Nogay, Kazakh, Karakalpak, Kipchak Uzbek
East Kipchak	Kyrgyz, South Altay

Table 1.2: Categorisation of the Kipchak languages based on Johanson (2021, 22).

1.2 Phonology

This section briefly introduces the summary of the Tatar phonology. Tatar vowels consist of 8 or 9 phonemes as listed in Table 1.3. The distribution of Tatar consonants are illustrated in Table 1.4. Russian loanwords are usually pronounced within the Russian phonological system even if the word is considered as a part of the Tatar lexicon. The syllable structure of Tatar is maximally (C)CVCC.

---

[1]A good level of mutual intelligibility is demonstrated by the fact that Bashkir-speaking clips are not subbed in Tatar television channels, as can be seen in the video https://youtu.be/RUGT2GcdB_Q, from 7:36 to 7:47.

Possible onset clusters are typically /st/, as in *qa.zaq.**stan*** ('Kazakhstan'), though onset clusters are not found very often in words native to Tatar. When a coda consists of two consonants, the first consonant is typically a sonorant excluding nasals, as in ***art**.qan* ('increased'), ***qïlt*** ('suddenly'), and ***qajt**.qan* ('returned').

		[+front]		[-front]	
		[+round]	[-round]	[+round]	[-round]
High		/ü/	/i/	/u/	(/ïj/)
Mid		/ö/	/e/	/o/	/ï/
Low		/ä/		/a/	

Table 1.3: The vowel inventory of Tatar. The parenthesised phoneme /ïj/ is controversial.

		Labial	Dental	Postalveolar	Palatal	Velar	Uvular	Glottal
Nasal		/m/	/n/			/ŋ/		
Stop	[-voice]	/p/	/t/			/k/	/q/	
	[+voice]	/b/	/d/			/g/		
Affricate	[-voice]			*/t͡s/	*/t͡ɕ/			
	[+voice]							
Fricative	[-voice]	/f/	/s/	/ʂ/	/ɕ/		/χ/	/h/
	[+voice]	*/v/	/z/	/ʐ/	/ʑ/		/ʁ/	
Trill			/r/					
Approximant		/w/	/l/		/j/			

Table 1.4: The inventory of Tatar consonants. Labiodental fricatives /f/, /v/ and labiovelar approximant /w/ are merged into the Labial column. Phonemes with an asterisk are used only in Russian-origin words. The voiced uvular fricative /ʁ/ can also be realised as a voiced velar fricative [ɣ].

As is common in Turkic languages, Tatar has fronting and rounding harmonies, of which the latter is not represented in the orthography. The fronting harmony of Tatar is a forward propagation of the [+front] feature from the last syllable of a stem. Example (3) contains no front vowel, because the stem morpheme *bar-* ('to go') lacks the [+front] feature, and therefore [+front] is never found in the suffixes and enclitic attached to the stem, {-DI}, {-IGIz}, and {=mI}. On the contrary, in (4), the syllable of the verbal stem *kil-* contains the vowel /i/, whose [+front] feature is propagated to the following suffixes and enclitic, as illustrated by the arrows. Words that are code-switched to Russian never have [+front] feature, cf. (5), unless they end with either -ия (*-ijä*) (6), -ие (*-ije*) (7), or a palatalisation sign -ь (-ʲ), also known in Russian as мягкий знак 'soft sign' (8).

(3) *bar-dï-ɣïz=mï?*
    go-PST-2PL=Q
    'Did you go?'

    [+front]
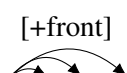
(4) *kil-de-gez=me?*
come-PST-2PL=Q
'Did you come?'

(5) *injener-nïŋ / *injener-neŋ*
engineer-GEN
'engineer's'

(6) **pandemijä-nïŋ / pandemijä-neŋ*
pandemic-GEN
'pandemic's'

(7) **učreždenije-nïŋ / učreždenije-neŋ*
establishment-GEN
'establishment's'

(8) **apreľ-nïŋ / apreľ-neŋ*
april-GEN
'April's'

The rounding harmony, or roundedness coarticulation, is a propagation of [+round] to succeeding mid vowels, as exemplified in (9).[2] However, the assimilation is blocked when non-mid vowels intervene; in (10), for instance, the low vowel /a/ in the suffix *qan* blocks the harmony, therefore the assimilation stops before *qan*. The rounding harmony is not expressed in the orthography, and in the transcription here either, because the rounding harmony is not crucial to the identity of a phonological word. The existence of the rounding harmony in Tatar is in fact controversial, and some researchers disagree with it (Conklin, 2015; Henry, 2018).

[+round]

(9) *onït-tï-ɣïz=mï?* realised as: /o.not.to.ʁoz.mo/
forget-PST-2PL=Q
'Did you forget?'

[+round]

(10) *onït-qan-sïz=mï?* realised as: /o.not.qan.sïz.mï/
forget-PST.PTCP-2PL=Q
'Had you forgotten?'

---

[2]Note that the actual phonetic impression of /o/ and /ö/ is not rounded. Therefore, what is assimilated in the rounding harmony of Tatar is an abstract phonological feature that is tentatively tagged as [±round].

1.3   Morphology

Tatar morphology is epitomised by the agglutinative derivation, declension and inflection by adding suffixes.

1.3.1   Nouns

There are six morphological case-marking in Tatar: nominative -Ø, genitive {-nIŋ}, dative {-GA}, accusative {-nI}, locative {-DA}, and ablative {-DAn}.

1.3.2   Verbs

Finite verbs inflect by tense (present and past), person (first, second, and third) and number (singular and plural). Non-finite verb forms include participles (definite future {-AčAG}, indefinite future {-(I)r}, agentive present {-UčI}, and past {-GAn}), converbs (sequential {-(I)p}, posterior 'after ...' {-GАč}, prior 'before' {-GAnčI}, simultaneous 'while ...' {-A}), conditional form {-sA}, verbal noun ({-U}), and infinitive ({-(I)rGA}, or {-mAsGA} in negative). These non-finite verb forms do not inflect by tense, person, or number, except for the conditional form that does agree with person and number. Verbs may also inflect by voice (passive {-(I)l}~{-(I)n}, middle {-(I)n}, causative {-t}~{-tIr}~{-Ar}, and reciprocal {-(I)š}), and polarity (negative {-mA}). Forms of conditional moods are inflected by person and number of their corresponding subject, but not by tense. Below are examples of verbal inflection.

(11) *čïɣ-ar-ïl-dï-q*
      go_out-CAUS-PASS-PST-1PL
      'We were forced to go out'

(12) *tanï-š-tïr-ïrɣa*
      be_acquainted-REC-CAUS-INF
      'to introduce'

(13) *söjlä-š-kän-dä*
      speak-REC-PST.PTCP-LOC
      'when (we, you, they) talked'

1.4   Syntax

As with many Turkic languages, Tatar is a nominative-accusative language with the canonical word order of the Subject-Object-Verb type, and a modifier in a noun phrase precedes its head. An example sentence with an intransitive verb is shown in (14), and one with a transitive verb in (15). Assuming that an external argument in these verbal sentences is introduced as the specifier of a functional

phrase vP (Kratzer, 1996), a simplest representation of their syntactic form is given in (1.1) and (1.2), respectively. Here we are also accepting the close-knit relationship of a phrase structure and semantic arguments, as formalized in the Uniformity of Theta-Assignment Hypothesis (UTAH) (Adger, 2003; Baker, 1988), where the AGENT argument of a verbal predicate is a direct daughter of vP and the THEME a direct daughter of VP.

(14) *min   jöz-äm.*
     1SG.NOM swim-1SG
     'I swim.'

(15) *min   kitap-nï   uqï-dï-m.*
     1SG.NOM book-ACC read-PST-1SG
     'I read the book.'

(1.1)

```
              TP
           /      \
         min       T'
                 /    \
               vP    [Pres]
              /  \
          ⟨min⟩   v̄
                 /  \
               VP    v
                |
              jozäm
```

(1.2)

```
              TP
           /      \
         min       T'
                 /    \
               vP    [Past]
              /  \
          ⟨min⟩   v'
                 /  \
               VP    v
              /  \
        kitapnï   V'
                  |
                uqïdïm
```
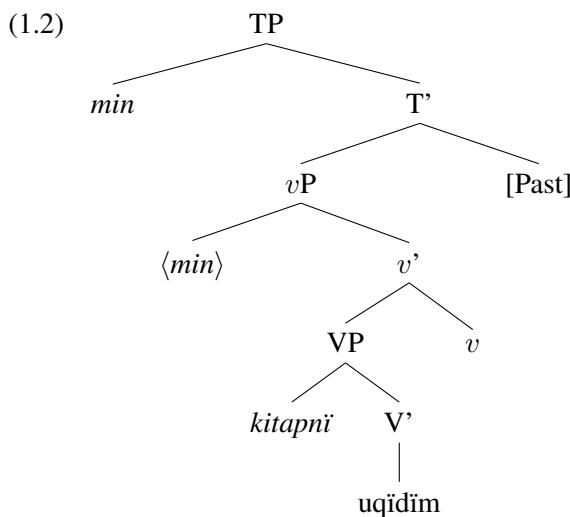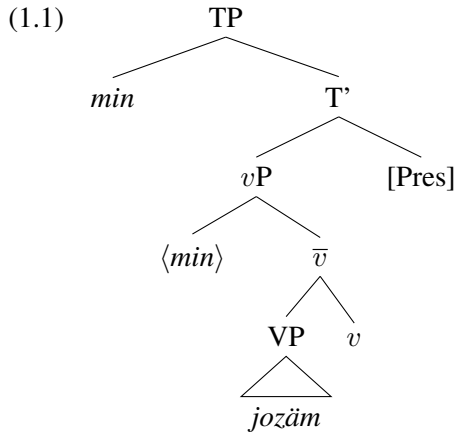
Adnominal modifiers, adjunct modifiers and adverbial clauses also precede the head, as in (16), (17), and (18), respectively. Finite verbs obligatorily agrees with the subject's person and number, though plural agreement in third person is optional and is often marked by the singular form.

(16) *aq   külmäk*
     white shirt
     'white shirt'

(17) *bügen tatarstan-da   unsigez gradus-qa   qädär salqïn bul-ačaq*
     today Tatarstan-LOC eighteen degree-DAT up_to cold   be-DEF.FUT.PTCP
     'Today, it will be as cold as eighteen degrees in Tatarstan.' (Saykhunov et al., 2021)

(18) *čišmä suw-ï   eč-ep,   bit-lär-ebez-ne   juw-ɣač,   räxät bul-ïp kit-te.*
     spring water-POSS.3 drink-CVB face-PL-POSS.1PL-ACC wash-CVB.PF happy be-CVB leave-PST.3
     'After (he/she/they) drank spring water and washed our face, (he/she/they) left happily.'
     (Saykhunov et al., 2021)

Since this thesis puts a particular focus on infinitival constructions, syntactic properties and se-
mantic functions of infinitives in Tatar are described here. Infinitives do not inflect by tense, person,
or number. A common role of infinitive is to form an adverbial clause with a purposive or modal
meaning as in (19). It can also be an argument of certain predicates as shown in (20). Several modal
predicates, such as *kiräk* 'necessary', *mömkin* 'possible', *tiješ* 'obligatory', *jara-* 'be allowed', *turï
kil-* 'have to', *žïjïn-* 'be intended', etc. require infinitive complement as in (21). Infinitives may also
modify a noun when the noun phrase is the subject of a main predicate, typically an existential predi-
cate *bar* 'exist' or *juq* 'not exist.' A few nominal predicates allow direct modification as in (23), which
we will claim in later chapters to be mermaid construction. Though infinitives are typically used to
form an embedded phrase, a bare infinitive form with no main finite predicate is also used to express
a deontic meaning as shown in example (24). The syntactic clause structure of the latter two types is
the chief interest in this thesis, and will be further analysed later in Chapters 2 and 3.

(19) *kafe-ɣa dust-ïbïz   belän **očra-š-ïrɣa**   bar-dï-q*
     cafe-DAT friend-POSS.1PL with   **meet-REC-INF** go-PST-1PL
     'We went to the cafe to meet with our friend' (adverbial modification)

(20) *min   dus-lar belän **arala-š-ïrɣa**   jarat-am*
     1SG.NOM friend-PL with   **hang_out-REC-INF** like-1SG
     'I like hanging out with friends' (control)

(21) *sez-gä   **kit-ärgä** mömkin*
     2PL-DAT **leave-INF** possible
     'You can leave' (raising)

(22) *ozaq söjlä-š-ep   **tor-ïrɣa** waqït-ï   juq   ide*
     long speak-REC-CVB **stand-INF** time-POSS.3SG not_exist COP.PST.3
     '(he/she) had no time to talk for long' (Burbiel, 2018) (adnominal modification)

(23) *išle* *ɣailä* *bala-lar-ïn-a* *šähär-gä* **kit-ärgä** *röxsät*
populous family child-PL-POSS.3-DAT city-DAT leave-INF permission
'Children of a large family are allowed to leave for the city' (Ibrahim, 2008)

(24) *tïrïš-ïp* **uqï-rɣa**, **jalqawlan-masqa**
try-CVB **study-INF be_lazy-NEG.INF**
'Study hard, don't be lazy'

Chapter 2

Control and Raising

In textbooks of linguistic theories, control and raising are often introduced together (Adger, 2003; Carnie, 2013; Dalrymple et al., 2019). This convention comes from the strikingly similar appearance of control and raising constructions. At one glance, an example of control in (2.1) merely differs from that of raising in (2.2) in the matrix verb.

(2.1) Alice tries to hit Bill. (*control*)

(2.2) Alice seems to hit Bill. (*raising*)

Despite their similarity, various tests by linguists have shown that these predicates have syntactically different structures. First of all, raising predicates do not assign a semantic role to the external argument (i.e., the matrix subject) while control predicate does. This can be confirmed by using *it*-expletive, though it is a somewhat English-centric diagnostic. In (2.3a), paraphrasing with *it*-expletive fails because the control predicate *try* obligatorily assigns an agentive semantic role to the external argument. Sentence (2.3b), in contrast, is a grammatical sentence, because *seem* does not assign any semantic role.

(2.3)   (a) * It tries that Alice hits Bill. (*control*)

       (b) It seems that Alice hits Bill. (*raising*)

Another piece of evidence that predicates like *seem* do not assign a semantic role to the matrix subject is the difference in semantic role assignment of the passivized counterparts. Passivized (2.4a) gives us impression that Bill is deliberately making some effort because he wants to be hit by Alice; that is, the control predicate *try* assigns a semantic role to Bill. In (2.4b), however, there seems to be no change in semantic roles from (2.2). In both instances, the hitter (the agent of *hit*) is Alice, and the hittee (the patient of *hit*) is Bill. The patientive semantic role of Bill is solely given by the subordinate predicate *hit*. It is also intuitively supported by the unnaturalness of *seemer* in contrast to the acceptability of *trier* (the one who tries).

(2.4) (a)   Bill tries to be hit by Alice. (*control*)

(b)   Bill seems to be hit by Alice. (*raising*)

In addition, the fact that raising predicates do not assign any semantic role to the external argument enables them to be combined with idiom chunks retaining their idiomatic meaning. The English idiom in (2.5a) means that a secret has been revealed, and the combination of a raising predicate *seem* (2.5b) just adds an evidential perspective and does not cause any change in the original idiomatic meaning. However, when used with a control predicate *try* (2.5c), the sentence loses the idiomatic meaning and literally denotes a situation where the cat attempts to get out of the bag.

(2.5) (a)   The cat is out of the bag.

(b)   The cat seems to be out of the bag. (*raising*)

(c)   # The cat tries to be out of the bag. (*control*)

Now, if the control predicates assign a semantic role (typically agent role) to their external argument, does it mean that the matrix subject bears two semantic roles together with the one given by the predicate in the subordinate clause? If so, it violates the classic notion of theta criterion that states 'each arguments bears one and only one θ-role, and each θ-role is assigned to one and only one argument' (Chomsky, 2010). Assuming that this statement is valid, the double semantic role assignment of control predicates is solved by supposing an independent pronoun PRO that is phonetically unpronounced. That is, we assume a covert subject pronoun, coreferential with the matrix subject, in the embedded clause (25). In raising, on the other hand, given that the semantic role of the matrix subject is assigned by the predicate in the embedded clause, the subject is interpreted as 'raised' from the embedded clause, leaving the trace in its hypothetical original position (26)[1].

(25) Alice$_i$ tries [PRO$_i$ to hit Bill]. (*control*)

(26) Alice$_i$ seems [t$_i$ to hit Bill]. (*raising*)

The phonetically null but syntactically present argument PRO is known to be a fairly theory-neutral concept (Landau, 2013). Landau (2013: 69–78) enumerates pieces of empirical crosslinguistic evidence for the existence of PRO. First, if we assume an infinitival phrase constitutes a clause, it is inevitable to have some syntactic element, phonetically overt or covert, as the subject of the clause.[2] This assumption seems to hold true in Tatar, because a bare infinitive phrase can be an independent sentence, as will be discussed in detail in Section 2.4. Indeed, a number of languages express obligatory control with a finite verb, for example in Serbo-Croatian (2.6)-(2.7), where the clausal analysis is obviously valid.

---

[1]Note that LFG, which do not assume any such movements, treats raising as functional control, and treats control as anaphoric control.

[2]This assumption excludes the constructions of restructuring verbs, such as long-distance passive in German (Wurmbrand, 2001).

(2.6) Serbo-Croatian (Mihalicek, 2012)

Ana	*hoće*	[*da*	*Marko*	*ode*]
Ana.NOM	want.PRS.3SG	comp	Marko.NOM	leave.PRS.3SG

"Ana wants Marko to leave.'"

(2.7) Serbo-Croatian (Mihalicek, 2012)

Ana	*hoće*	[*da*	*ode*]
Ana.NOM	want.PRS.3SG	COMP	leave.PRS.3SG

"Ana wants to leave.'"

Second, various diagnostics support that PRO is a syntactic substance that is able to occupy a subject position and can anaphorically bind an element or can be bound, just like an independent DP. The chief evidence is that secondary predicates require an overt DP outside as their argument, as the contrast in (2.8) shows (examples from Landau (2013: 72)). In this respect, PRO in control infinitive can serve as the subject of the secondary predicate (2.9). (2.10) shows PRO's ability of anaphoric binding. In addition, sentence (2.11)'s peculiar discrepancy in case-marking (accusative and dative) can be accounted for by assuming a PRO bearing dative as the default case by an infinitive verb. Landau (2013) himself mentions more supporting arguments for the existence of PRO and concludes that it seems to be infeasible to establish a grammatical framework without PRO.

(2.8) (a) He served dinner [angry at the guests]

(b) * Dinner was served [angry at the guests]

(2.9) [PRO to serve dinner angry] is not good.

(2.10) [PRO$_i$ to behave oneself$_i$] is important.

(2.11) Russian

Ona	*poprosi-l-a*	*ego*	[PRO	*ne*	*ezdi-t'*	*tuda*	*odn-omu*]
3SG.F.NOM	ask-PST-SG.F	3SG.M.**ACC**		NEG	go-INF	to.there	alone-**DAT**

"She asked him not to go there alone'

This chapter provides a theoretical analysis of control and raising by non-verbal (i.e., adjectival and nominal) predicates in Tatar. Section 2.1 enumerates more or less grammaticalized infinitival constructions of Tatar, and outlines a general descriptive observation. Based on the constructions described in Section 2.1, Sections 2.2 and 2.3 theoretically analyze the raising and control, respectively, in the framework of the Minimalist Approach. Section 2.2 argues that raising predicates in Tatar do not have CP/TP but involve a lower (smaller) phrase vP. Section 2.3, on the other hand, proposes that control predicates in Tatar involve a CP or TP.

2.1   Tatar Infinitives

As described in sections 1.3 and 1.4, Tatar infinitives take the form of the suffix -IrGA and chiefly function as adverbial modifier to signify purposive events. When used with certain predicates or

without any predicate (i.e., bare infinitive), however, it may also express several kinds of modal or aspectual meanings by grammaticalization. Types of such constructions are limited, and they are listed below.

2.1.1 Verbal predicates

1. {-IrGA} *kereš-*: inchoative aspect 'start to ...'. The subject must be the AGENT of the event (cf. (2.12) and (2.13)).

   (2.12) *Marat   mäktäp-kä   bar-ïrɣa kereš-te*
          Marat.NOM school-DAT go-INF   start-PST.3
          'Marat started going to school.'

   (2.13) **Šul xat   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   kereš-te*
          that letter Marat.NOM side-POSS.3-ABL write-PASS-INF start-PST.3
          'The letter began to be written by Marat.'

2. {-IrGA} *totïn-*: inchoative aspect, "start to ...". The subject must be the AGENT of the event (cf. (2.14) and (2.15)). The verb *totïn-* lexically means "to cling" but means "to start to" in this infinitival construction.

   (2.14) *Marat   mäktäp-kä   bar-ïrɣa totïn-dï*
          Marat.NOM school-DAT go-INF   cling-PST.3
          'Marat started going to school.'

   (2.15) **šul xat   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   totïn-dï*
          that letter Marat.NOM side-POSS.3-ABL write-PASS-INF cling-PST.3
          'The letter began to be written by Marat.'

3. {-IrGA} *źijïn-*: dynamic modality (plan), "plan to ..."; The subject is typically AGENT of the event (cf. (2.16) and (2.17), but THEME can also sometimes be the subject when *źijin-* means epistemic modality, cf. (2.18). The verb *źijïn-* lexically means "to get together" but means "to plan to" in this infinitival construction.

   (2.16) *Marat   mäktäp-kä   bar-ïrɣa źijïn-a*
          Marat.NOM school-DAT go-INF   get.together-PRS.3
          'Marat is going to go to school.'

   (2.17) **šul xat   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   źijïn-a*
          that letter Marat.NOM side-POSS.3-ABL write-PASS-INF get.together-PRS.3
          'The letter is going to be written by Marat.'

   (2.18) *jañ ɣïr   jaw-arɣa źijïn-a*
          rain.NOM fall-INF   get.together-PRS.3
          'It is going to rain.'

4. {-IrGA} *tor-*: prospective aspect, 'be about to ...'. The subject can be either AGENT (2.19) or THEME (2.20) marked with nominative. This construction allows passivization of the embedded clause (2.21). The verb *tor-* lexically means "to stand" but means "to be about to" in this infinitival construction.

    (2.19) *Marat   šul   xat-nï   jaz-arɣa   tor-a*
            Marat.NOM   this   letter-ACC   write-INF   stand-PRS.3
            'Marat is about to write the letter.'

    (2.20) *jañɣïr   jaw-arɣa   tor-a*
            rain.NOM   fall-INF   stand-PRS.3
            'It is about to rain.'

    (2.21) *šul   xat   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   tor-a*
            that   letter   Marat.NOM   side-POSS.3-ABL   write-PASS-INF   stand-PRS.3
            'The letter is about to be written by Marat.'

5. {-IrGA} *jör-*: dynamic modality, 'be prepared to ...'. This construction allows only AGENT as its subject (2.22). Subject of THEME and passivization of the embedded verb are ungrammatical (2.23), (2.24). The verb *jör-* lexically means "to walk" but means "to be prepared to" in this infinitival construction.

    (2.22) *Marat   šul   xat-nï   jaz-arɣa   jöri*
            Marat.NOM   this   letter-ACC   write-INF   walk.PRS.3
            'Marat is prepared to write this letter.'

    (2.23) **jañɣïr   jaw-arɣa   jöri*
            rain.NOM   fall-INF   walk.PRS.3
            'It is prepared to rain.'

    (2.24) **šul   xat   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   jöri*
            that   letter   Marat.NOM   side-POSS.3-ABL   write-PASS-INF   walk.PRS.3
            'The letter is prepared to be written by Marat.'

6. {-IrGA} *bul-*: attitude, 'decide to ...'. This construction allows only AGENT as its subject (2.25) which is marked in nominative. Subject of THEME and passivization of the embedded verb are unacceptable (2.26), (2.27). The verb *bul-* lexically means "to be" or "to become", but means "to decide to" in this infinitival construction with a nominative AGENT.

    (2.25) *Marat   šul   xat-nï   jaz-arɣa   bul-dï*
            Marat.NOM   this   letter-ACC   write-INF   become-PST.3
            'Marat decided to write this letter.'

    (2.26) **jañɣïr   jaw-arɣa   bul-dï*
            rain.NOM   fall-INF   become-PST.3
            'It decided to rain.'

(2.27) *šul xat (Marat taraf-ïn-nan) jaz-ïl-ïrɣa bul-dï
      that letter Marat.NOM side-POSS.3-ABL write-PASS-INF become-PST.3
      'The letter decided to be written by Marat.'

7. {-IrGA} *bul-*: modality (possibility), 'be possible to ..., be allowed to ...' This construction requires AGENT with dative case-marking.

(2.28) *Marat-qa šul xat-nï jaz-arɣa bul-a*
      Marat-DAT this letter-ACC write-INF become-PRS.3
      'Marat is allowed to write the letter.'

(2.29) ??*jaŋɣïr-Ga jaw-arɣa bul-a*
      rain-DAT fall-INF become-PRS.3
      '(*lit.*) Rain is allowed to fall.'

(2.30) ??*šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa bul-a*
      that letter-DAT Marat side-POSS.3-ABL write-PASS-INF become-PRS.3
      '(*lit.*) The letter is allowed to be written by Marat.'

8. {-IrGA} *jarat-*: attitude, 'like to ...' As is expected, this construction allows only AGENT as its subject (2.22). Subject of THEME and passivization of the embedded verb are unacceptable (2.23), (2.24).

(2.31) *Marat šul xat-nï jaz-arɣa jarat-a*
      Marat.NOM this letter-ACC write-INF like-PRS.3
      'Marat likes to write this letter.'

(2.32) *jañɣïr jaw-arɣa jarat-a*
      rain.NOM fall-INF become-PRS.3
      'It likes to rain.'

(2.33) *šul xat (Marat taraf-ïn-nan) jaz-ïl-ïrɣa jarat-a*
      that letter Marat.NOM side-POSS.3-ABL write-PASS-INF become-PRS.3
      'The letter likes to be written by Marat.'

9. {-IrGA} *telä-*: desire, 'want to ...' As is expected, this construction allows only AGENT as its subject (2.34). Subject of THEME and passivization of the embedded verb are unacceptable (2.35), (2.36).

(2.34) *Marat šul xat-nï jaz-arɣa teli*
      Marat.NOM this letter-ACC write-INF want.PRS.3
      'Marat wants to write this letter.'

(2.35) *jañɣïr jaw-arɣa teli*
      rain.NOM fall-INF want.PRS.3
      'It wants to rain.'

(2.36) *šul xat (Marat taraf-ïn-nan) jaz-ïl-ïrɣa teli
   that letter Marat.NOM side-POSS.3-ABL write-PASS-INF want.PRS.3
   'The letter likes to be written by Marat.'

10. {-IrGA} *turï kil-*: modality (deontology), 'have to ...'. In this construction, the AGENT is marked with dative. Subject of THEME and passivization of the embedded verb are unacceptable (2.60), (2.61).

 (2.37) *Marat-qa šul xat-nï jaz-arɣa turï kil-de*
   Marat.DAT this letter-ACC write-INF direct come-PST.3
   'Marat had to write the letter.'

 (2.38) **jañɣïr-ɣa jaw-arɣa turï kil-de*
   rain-DAT fall-INF direct come-PST.3

 (2.39) **šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa turï*
   that letter-DAT Marat.NOM side-POSS.3-ABL write-PASS-INF direct
   kil-de
   come-PST.3

11. {-IrGA} *qal-*: 'have no choice but to ...' This construction expresses deontic modality. The AGENT is marked by dative (2.40). THEME in dative is not usually accepted (2.41), but some speakers allow passivization (2.42).

 (2.40) ?*Marat-qa mäktäp-kä bar-ïrɣa qal-a*
   Marat-DAT school-DAT go-INF remain-PRS.3
   'Marat has no choice to go to school.'

 (2.41) **jañɣïr-ɣa jaw-arɣa qal-a*
   rain-DAT fall-INF remain-PRS.3

 (2.42) ?*šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa qal-a*
   that letter-DAT Marat side-POSS.3-ABL write-PASS-INF remain-PRS.3
   'The letter must be written by Marat.'

12. {-IrGA} *jara-*: permission, 'be allowed to ...' This construction requires AGENT with dative case-marking.

 (2.43) *Marat-qa šul xat-nï jaz-arɣa jarïj*
   Marat-DAT this letter-ACC write-INF be.suitable.PRS.3
   'Marat is allowed to write the letter.'

 (2.44) ??*jaŋɣïr-Ga jaw-arɣa jarïj*
   rain-DAT fall-INF be.suitable.PRS.3
   '(*lit.*) Rain is allowed to fall.'

(2.45) ??*šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa jarïj*
    that letter-DAT Marat side-POSS.3-ABL write-PASS-INF be.suitable
    '(*lit.*) The letter is allowed to be written by Marat.'

13. {-IrGA} *it-*: attempt, 'try to ...' This construction, with a nominative subject, is reported to be grammatical in Burbiel (2018), but some native speakers do not allow it.

(2.46) ?*Marat mäktäp-kä bar-ïrɣa it-te*
    Marat.NOM school-DAT go-INF do-PST.3
    'Marat tried to go to school.'

2.1.2 Adjectival Predicates

1. {-IrGA} *kiräk*: modality (deontology), necessity, 'need to ...'. In this construction, the AGENT is marked with dative. Subject of THEME and passivization of the embedded verb are unacceptable (2.60), (2.61).

(2.47) *Marat-qa šul xat-nï jaz-arɣa kiräk*
    Marat-DAT this letter-ACC write-INF necessary
    'Marat needs to write the letter.'

(2.48) **jañɣïr-ɣa jaw-arɣa kiräk*
    rain-DAT fall-INF necessary

(2.49) **šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa kiräk*
    that letter-DAT Marat.NOM side-POSS.3-ABL write-PASS-INF necessary

2. {-IrGA} *tiješ*: deontic/epistemic modality, obligation, 'must ...' In this construction, the subject is marked in nominative. It allows both a THEME subject and passivization unlike other constructions.

(2.50) *Marat šul xat-nï jaz-arɣa tiješ*
    Marat-DAT this letter-ACC write-INF obligatory
    'Marat must write the letter.'

(2.51) *jañɣïr jaw-arɣa tiješ*
    rain-DAT fall-INF obligatory
    'It must rain.'

(2.52) *šul xat (Marat taraf-ïn-nan) jaz-ïl-ïrɣa tiješ*
    that letter-DAT Marat.NOM side-POSS.3-ABL write-PASS-INF obligatory
    'The letter must be written by Marat.'

3. {-IrGA} *mömkin*: deontic modality (permission) or dynamic modality (ability), 'be possible to ...'. *mömkin* is an adjective meaning 'possible'. The AGENT of this construction is marked with dative (2.53). THEME with dative (2.54) and passivization (2.55) is unacceptable.

(2.53) *Marat-qa   mäktäp-kä   bar-ïrɣa mömkin*
       Marat-DAT   school-DAT   go-INF   possible
       'Marat can go to school.'

(2.54) **jaŋɣïr-Ga jaw-arɣa mömkin*
       rain-DAT   fall-INF   possible

(2.55) **šul xat-qa   (Marat taraf-ïn-nan)   jaz-ïl-ïrɣa   mömkin*
       that letter-DAT Marat side-POSS.3-ABL write-PASS-INF possible

4. {-IrGA} *mömkin*: epistemic modality "might". Unlike the construction above, this construction takes a nominative subject, and also allows a THEME subject and passivization.

(2.56) *Marat   mäktäp-kä   bar-ïrɣa mömkin*
       Marat.NOM   school-DAT   go-INF   possible
       'Marat might go to school.'

(2.57) *jaŋɣïr   jaw-arɣa mömkin*
       rain.NOM   fall-INF   possible
       'It might rain.'

(2.58) **šul xat   (Marat taraf-ïn-nan)   jaz-ïl-ïrɣa   mömkin*
       that letter.NOM Marat side-POSS.3-ABL write-PASS-INF possible
       'That letter might be written by Marat.'

2.1.3 Others

1. {-IrGA} *ide*: desire, suggestion, 'should, must ...' *ide* is an auxiliary that functions either as a copula marking the past tense or as a particle expressing the speaker's conditional, desiderative, or deontic attitude to the event. In this construction, the AGENT is marked with dative (2.59). Subject of THEME and passivization of the embedded verb are unacceptable (2.60), (2.61).

(2.59) *Marat-qa   mäktäp-kä   bar-ïrɣa ide*
       Marat-DAT   school-DAT   go-INF   AUX.3
       'Marat should go to school.'

(2.60) **jañɣïr-ɣa jaw-arɣa ide*
       rain-DAT   fall-INF   AUX.3
       'It should rain.'

(2.61) **šul xat-qa   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   ide*
       that letter-DAT Marat.NOM   side-POSS.3-ABL write-PASS-INF AUX.3
       'The letter should be written by Marat.'

2. {-IrGA}: deontic modality, 'must ...'. This construction expresses deontic modality. The AGENT (the entity bearing the obligation) is marked in dative (2.62), but its usage is mostly limited to deontic statement to the general audience (2.63). Theme with the dative (2.64) and passivization (2.65) are unacceptable.

(2.62) *?Marat-qa mäktäp-kä bar-ïrɣa*
      Marat-DAT school-DAT go-INF
    'Marat must go to school.'

(2.63) *söjäl-mäskä*
      lean-NEG.INF
    'Don't lean (e.g., on a train's doors)'

(2.64) **jañɣïr-ɣa jaw-arɣa*
      rain-DAT fall-INF

(2.65) **šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa*
      that letter-DAT Marat side-POSS.3-ABL write-PASS-INF

2.1.4 Nominal predicates

The following constructions exhibit similar forms, but their predicate is a noun, whose semantics is somewhat bleached to a functional one. Though some of these are often found in text and speech, they have not been described in existing grammars of Tatar (Burbiel, 2018; Zäkiev & Xisamova, 2015).

1. {-IrGA} *röxsät*: deontic modality, permission 'be allowed to ...'. The lexical meaning of *röxsät* is 'permission'. The AGENT is expressed with dative case-marking. THEME, both unaccusative and promoted by passivization, cannot appear in the AGENT's position, unless it is somehow reanalyzed as AGENT.

(2.66) *Marat-qa šul xat-nï jaz-ar*Ga *röxsät*
      Marat-DAT that letter-ACC write-INF permission
    'Marat is allowed to write the letter.'

(2.67) *#jaŋɣïr-ɣa jaw-arɣa röxsät*
      rain-DAT fall-INF permission
    '(*lit.* The rain is allowed to fall.'

(2.68) *#šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa röxsät*
      that letter-DAT Marat side-POSS.3-ABL write-PASS-INF permission
    '(*lit.* The letter is allowed to be written by Marat.'

2. {-IrGA} *waqït*: deontic modality 'it is time to ..., now one must ...'. The AGENT, or the bearer of the obligation, is marked by dative. Non-AGENT arguments cannot occupy the same position.

(2.69) *marat-qa šul xat-nï jaz-arɣa waqït*
　　　 Marat-DAT that letter-ACC write-INF time
　　　 'Now Marat must write the letter.'

(2.70) *??šul xat-qa (Marat taraf-ïn-nan) jaz-ïl-ïrɣa waqït*
　　　 that letter-DAT Marat.NOM side-POSS.3-ABL write-PASS-INF time
　　　 'Now that letter must be written by Marat.'

(2.71) *??jaŋɣïr-ɣa jaw-arɣa waqït*
　　　 rain-DAT fall-INF time
　　　 'Now it must rain.'

3. {-IrGA} *isäp*: attitude 'intend to ...'. The AGENT (the attitude holder) is marked by genitive, and *isäp* may agree with the possessor by means of possessive suffix. Non-AGENT arguments cannot occupy the same position. Its striking feature that is different from other infinitival constructions is that it can take a question clitic {=mI} right after the infinitive, while it is usually pronounced at the sentence-final position.

(2.72) *Marat-nïŋ mäktäp-kä bar-ïrɣa isäb-e*
　　　 Marat-GEN school-DAT go-INF thought-POSS.3
　　　 'Marat is going to go to school.'

(2.73) **jaŋɣïr-nïŋ jaw-arɣa isäb-e*
　　　 rain-GEN fall-INF thought-POSS.3

(2.74) **šul xat-nïŋ (Marat taraf-ïn-nan) jaz-ïl-ïrɣa isäb-e*
　　　 that letter-GEN Marat side-POSS.3-ABL write-PASS-INF thought-POSS.3

(2.75) *kit-ärgä=me isäp?*
　　　 leave-INF-Q thought
　　　 'Are you going to leave?'

4. {-IrGA} *nijät*: attitude 'intend to ...'. Its function and usage are similar to *isäp*, but *nijät* tends to be used in a more formal context.

(2.76) *Marat-nïŋ mäktäp-kä bar-ïrɣa nijät-e*
　　　 Marat-GEN school-DAT go-INF intention-POSS.3
　　　 'Marat is going to go to school.'

(2.77) **jaŋɣïr-nïŋ jaw-arɣa nijät-e*
　　　 rain-GEN fall-INF intention-POSS.3

(2.78) **šul xat-nïŋ (Marat taraf-ïn-nan) jaz-ïl-ïrɣa nijät*
　　　 that letter-GEN Marat side-POSS.3-ABL write-PASS-INF intention-POSS.3

(2.79) *kit-ärgä=me   nijät?*
      leave-inf-q   intention
      'Are you going to leave?'

5. {-IrGA} *mäslixät*: deontic modality, 'should, had better ...'. The AGENT is marked by dative.
   Non-AGENT arguments cannot occupy the same position. *mäslixät* is not a commonly used
   noun, and some speakers do not know this construction.

(2.80) *Marat-qa   šul   xat-nï   jaz-arɣa   mäslixät*
      Marat-DAT   that   letter-ACC   write-INF   recommendation
      'Marat should write the letter.'

(2.81) ??*šul   xat-qa   (Marat   taraf-ïn-nan)   jaz-ïl-ïrɣa   mäslixät*
      that   letter-DAT   Marat   side-POSS.3-ABL   write-PASS-INF   recommendation
      'The letter should be written by Marat.'

Predicate	Category	AGENT's case	THEME	Passivization
kereš-	V	NOM	✓	*
totïn-	V	NOM	✓	*
žïjïn-	V	NOM	✓	*
jör-	V	NOM	*	*
bul-	V	NOM	*	*
jarat-	V	NOM	*	*
telä-	V	NOM	*	*
turï kil-	V	DAT	*	*
qal-	V	DAT	*	*
jara-	V	DAT	*	*
bul-	V	DAT	*	*
it-	V	NOM	*	*
kiräk	A	DAT	*	*
tiješ	A	NOM	✓	✓
mömkin	A	DAT	*	*
mömkin	A	NOM	✓	✓
Ø	?	DAT	*	*
ide	Aux	DAT	*	*
röxsät	N	DAT	*	*
waqït	N	DAT	*	*
isäp	N	GEN	*	*
nijät	N	GEN	*	*
mäslixät	N	DAT	*	*

Table 2.1: Summary of infinitival constructions in Tatar. The check marks and asterisks in the right
two columns represent accepted or not accepted, respectively, by native speakers, and they are not
necessarily equal to syntactic grammaticality.

The following subsections will concentrate on infinitival constructions with a non-verbal predi-
cate, and show that these constructions are instances of raising and control constructions.

2.2   Raising

This section looks into the raising construction with an adjectival predicate in Tatar. As we have already seen in the descriptive overview in Section 2.1, adjectives *tiješ* and *mömkin* seem to be triggers for subject-to-subject raising (SSR), because they allow non-AGENT subjects including a THEME/PATIENT promoted to subject by means of passivization.

Semantics characteristic to raising predicates is that the raised argument is not a part of the semantic argument relation specified by the predicates. This is confirmed by other (often English-centric) syntactic tests. Consider an example of raising with *seem* (2.82). Raising verbs like *seem* allow an expletive *it* to be the subject, and *John* is the syntactic subject of the embedded CP. Furthermore, while control verbs like *try* and *want* can be derived with an agentive suffix *-er* (*trier* 'one who tries' and *wanter* 'one who wants', respectively), raising verbs like *seem* cannot naturally derive with the same suffix (??*seemer*). It shows that *seem* does not semantically host a agent argument.

(2.82) John seems to be leaving.

(2.83) It seems that John is leaving.

From these facts, it seems that raising predicates do not subcategorize for a subject, for example *seem*(x, y) for some nominal arguments x and y like a transitive verb. Rather, it takes a whole proposition and returns a subjectified (evidential) evaluation; in the notation of formal semantics, it should look like $\lambda P.\text{seem}(P)$, where P is some proposition of type t. Therefore, though *John* in (2.82) superficially looks like a subject of the verb *seem*, these tests confirm that it is in fact not semantic argument of *seem*. In Generative theories of syntax, the subject surfacing up in the main clause is *raised* from the original specifier position of the embedded TP to the higher Spec,TP. Minimalist syntax explains that this movement is caused by the "strong" uninterpreted feature $[u\text{D*}]$ (Adger, 2003). The movement of *John* from the lower Spec,TP is driven by the same mechanism as the EPP.
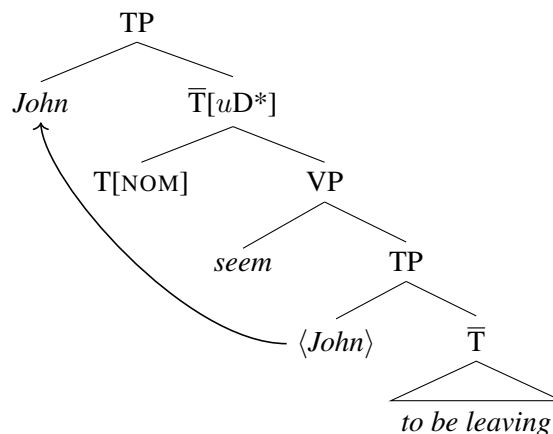
Figure 2.1: The structure of (2.82).

Similarly, *tiješ* subcategorizes for neither a subject nor any other nominal arguments. The syntactic subject of *tiješ* or *mömkin* is raised from the internal TP. Upon deriving the syntactic structure of

this construction, we assume that an external argument of an adjectival predicate is not inside AP but is introduced by some functional head as expressed in Figure 2.2, following Ramchand (2018). Here the functional phrase is represented as PredP, because the adjectival phrase serves as the predicate of a sentence (Al-Dobaian, 2006). In the present tense, its head Pred is normally phonetically unpronounced, but elsewhere it is realized as a copular predicate *bul-* (Figure 2.2). Thus, the structure of (2.84) is represented as Figure 2.3.
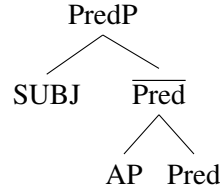
Figure 2.2: Introduction of PredP.

(2.84) *Marat$_i$* [t$_i$ *kit-ärgä*] {*tiješ*   / *mömkin*}
        Marat   leave-INF obligatory / possible
        'Marat must leave. / It may be the case that Marat leaves.'
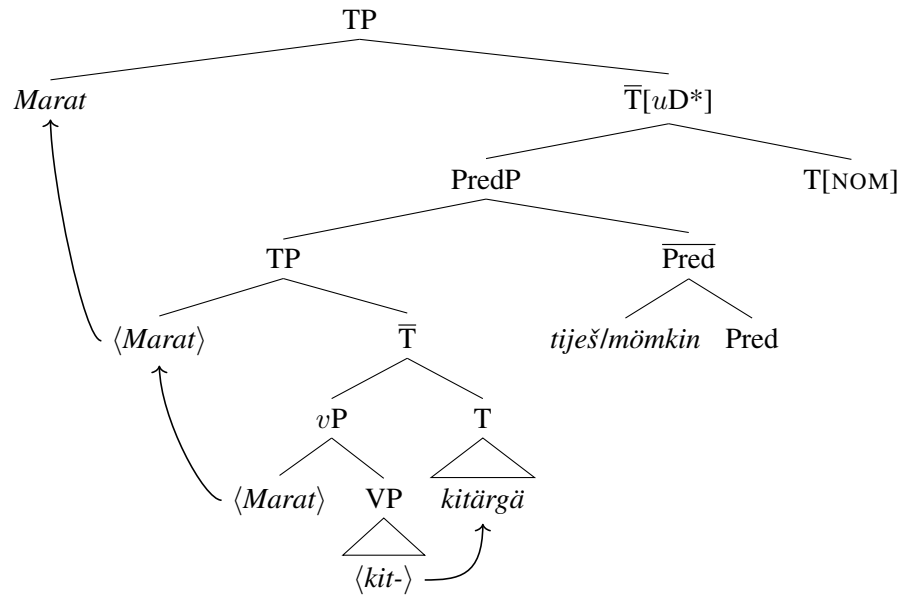
Figure 2.3: The structure of (2.84).

The next subsection will delve into the control construction with *kiräk*. We argue that the *kiräk-*type infinitival construction, unlike *tiješ*, is control, where a controller is an internal oblique argument of the control predicate, which is then moved to the higher Spec,TP position.

32

2.3   Control

This section theoretically accounts for the control construction with an embedded infinitival clause in Tatar. We have seen in the previous section that seemingly control non-verbal predicates that do not allow passivization include adjectives (*kiräk* and *mömkin*) and nouns (*röxsät*, *waqït*, and *mäslixät*). This observation raises two intriguing issues. First, control with a nominal predicate is against the common description in the literature of control that typically deals only with verbal and adjectival predicates (e.g., *try*, *promise*, *(be) likely*, etc.). Second, it is often a typological and syntactic fact that nominal and adjectival predicates have different functions and structures. Keeping these issues in mind, our aim in this section is to provide a consistent theoretical explanation for these two types of control predicates in Tatar. The first subsection deals with the structure and the mechanism of the control with the adjectival predicates (*kiräk* and *mömkin*). The second subsection, based on the arguments and consequences in the first subsection, proposes a syntactic structure of control nominal predicates in Tatar.

2.3.1   Adjectival Control

This section explains the control with adjectival predicates: *kiräk* and *mömkin*. The previous section has descriptively shown that these adjectival predicates fail the passivization test outlined in the beginning of this chapter. This fact is partial evidence in favor of a control analysis rather than a raising. In most textbook cases, an phonetically null syntactic element PRO is controlled by a (nominative) subject or an (accusative) object, which are called subject control and object control, respectively. However, the point that differs from them in these adjectival predicates of Tatar is that the controller is marked by neither nominative nor accusative, but is dative. Then, how should this control be represented in the Minimalist approach?

Recall the UTAH by Baker in Figure 2.4. Assuming that this hypothesis holds true in Tatar, the GOAL (the bearer of a state described by the adjective) is introduced under AP (Figure 2.5).

Just like canonical verbal predicate sentence, tense and other inflectional features are introduced by the head of TP, whose specifier is filled by the raised subject by virtue of the Extended Projection Principle (EPP). Thus, a prototypical example in (2.85) has the structure represented in Figure 2.6.

(2.85) *su*   *Marat-qa*   *kiräk.*
       water.NOM   Marat-DAT   necessary
     'Water is necessary to Marat'

We will derive a tree structure for sentence (2.47) from this template. The necessary thing (*su* 'water' in the example above) now is a proposition expressed by an infinitival clause (Figure 2.7).

After elaborating the subtree from the subordinate TP, we have the tree structure represented in Figure 2.9. At this point, we have not assumed the EPP at the level of the matrix clause, so the specifier of the matrix TP remains empty. Then, which constituent (or phrase) is the right candidate for occupying the Spec,TP position?
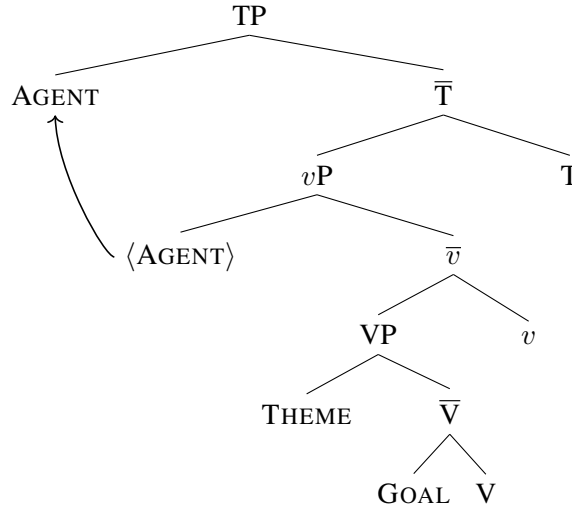
Figure 2.4: UTAH
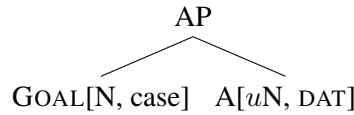
Figure 2.5: A minimal structure of an AP.

We have two possibilities that could be subject to the movement: the NP in the embedded clause (*šul xat-nï*) or the NP (*Marat-qa*) under AP.[3] The first candidate cannot be raised to the specifier of the higher TP, because it needs to go beyond the lower vP, which violates the Phase Impenetrability Condition (PIC) (Chomsky, 2001) defined in (2.86).

(2.86) PHASE IMPENETRABILITY CONDITION

Once a phase has been completed and sent to the interfaces, the internal domain of a phase [i.e. the complement of the phase head] is not accessible to operations at/above the next higher phase. Only the edge of the phase [the head plus any number of specifiers] remains accessible at the next higher phase. (Boeckx & Grohmann, 2004)

Following Chomsky (2001), phases are propositional, θ-complete phrases; namely, vP and CP.[4] The object of the embedded clause *šul xat-nï* cannot be raised to Spec,TP because it must cross the phase boundary between $\overline{v}$ and vP. The remaining candidate is the oblique argument of the sentence, i.e., *Marat-qa*; it is not a canonical candidate for raising by the EPP, because the NP argument subject to the EPP typically bears a nominative case-marking by T (Adger, 2003). Furthermore, *Marat-qa* seems to be already case-licensed (at least morphologically). This challenges the necessity of the

---

[3]The tactic with an expletive subject is generally not available in Tatar.

[4]Some argue for DP also being a phase.

TP

su

T̄

PredP   T

⟨*su*⟩   Pred̄

AP   Pred

Marat-qa   kiräk
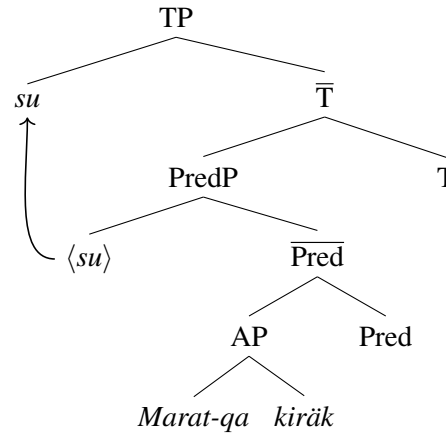
Figure 2.6: A tree representation of sentence (2.85).

TP

T̄

PredP   T

TP   Pred̄

PRO *šul xat-nï jaz-arɣa*   AP   Pred

Marat-qa   kiräk
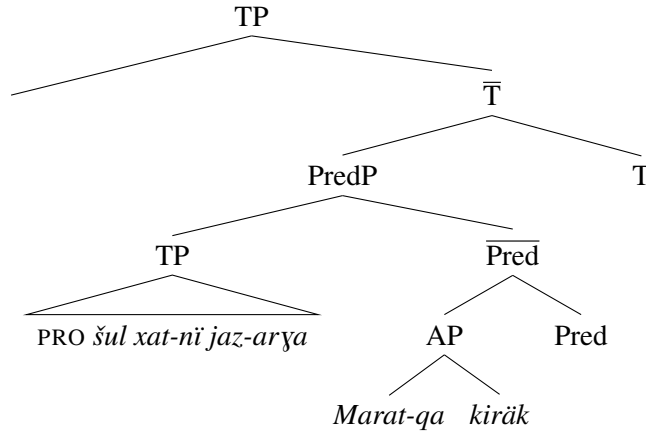
Figure 2.7: The infinitival clause (TP) is introduced under PredP.

movement, because Case assignment is a key factor that motivates the raising to Spec,TP (Bošković, 2011). Yet, if we assume that the EPP in Tatar can select (at least morphologically) case-marked NP when there is no other possible candidate, we can still predict the desired form. In other words, T has a strong uninterpretable case feature [$u$case:(NOM)*] which prefers to Agree with a nominative case-marking but can also attract an NP with a different case-marking when there is no other possible constituent.

Alternatively, following Sigurðsson (2003), we can sever an inherent case (semantically associated case, such as dative, ablative, and genitive) and a structural case (structurally assigned case, i.e., nominative and accusative). In this view, *Marat-qa* under the AP already has the dative case-marking as an inherent case, but has yet to be marked structurally. Therefore, it can Agree with the uninterpreted valued feature [$u$case:NOM*] and raised to Spec,TP while retaining the inherent case-marking[5].

---

[5]Sigurðsson (2003) assumes that the selection as to which case-marking is blocked, inherent or structural, is language-specific. The case selection of Tatar is similar to the Icelandic-type, where an inherent case blocks a structural case.
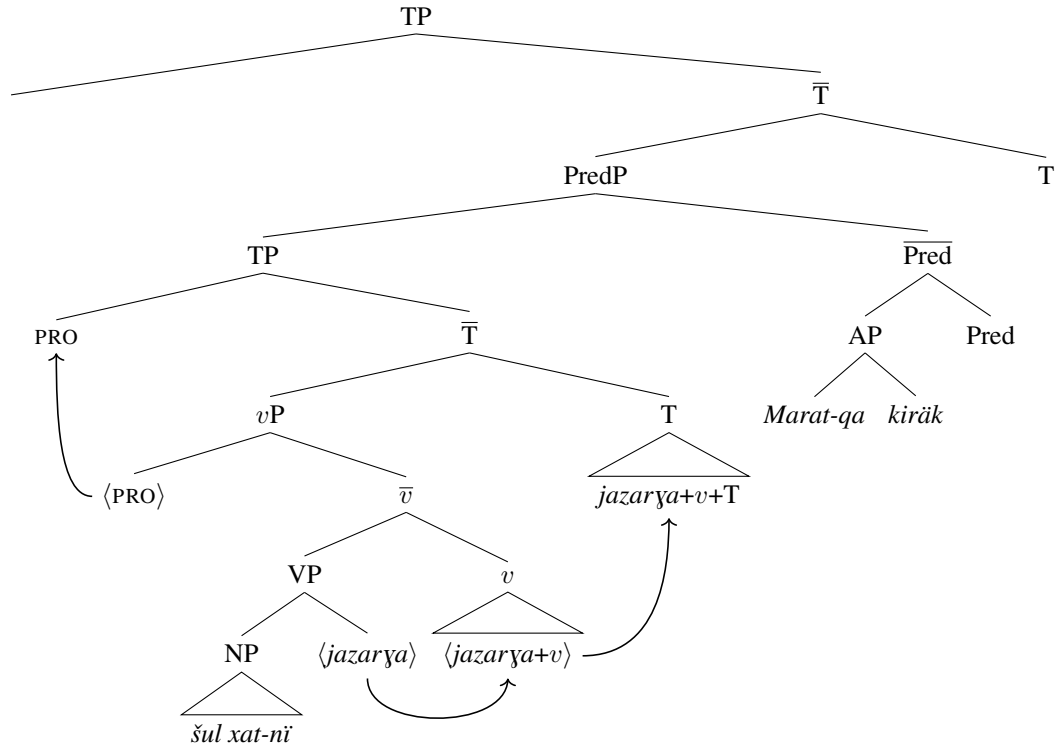
Figure 2.8: The tree structure of (2.47) before the movement to SpecTP.

It follows that *Marat-qa* in this construction is a quirky subject, where the abstract Case[6] is nominative while the morphological case-marking is not (Sigurðsson, 1992).

Either way, this obligatory movement to the Spec,TP position of the dative argument correctly predicts the unmarkedness in terms of focus. Though it is possible to place *Marat-qa* after the infinitival clause like (2.87), it is a marginal expression with a marked focus on *Marat-qa*. It is therefore reasonable to suppose that the dative argument is raised to Spec,TP as its default structure, and that post-syntactic operations at the level of information structure may further move it to the position immediately preceding the adjectival predicate.

The passivized counterpart (2.49) is ungrammatical because the subject of the passivized TP *šul xat* 'the letter' can neither be raised to the specifier of the higher TP, as illustrated in Figurenor be anaphorically controlled by any other NP.

(2.87) *šul   xat-nï   jaz-arɣa Marat-qa   kiräk*
       that letter-ACC write-INF Marat-DAT necessary
      'It is Marat who needs to write the letter (focus on Marat)'

However, our assumptions so far predict barely acceptable sentences like (2.49).

---

[6]It has been pointed out that abstract "Case" is a confusing misnomer; Sigurðsson (2003) instead calls it "argument licensing".
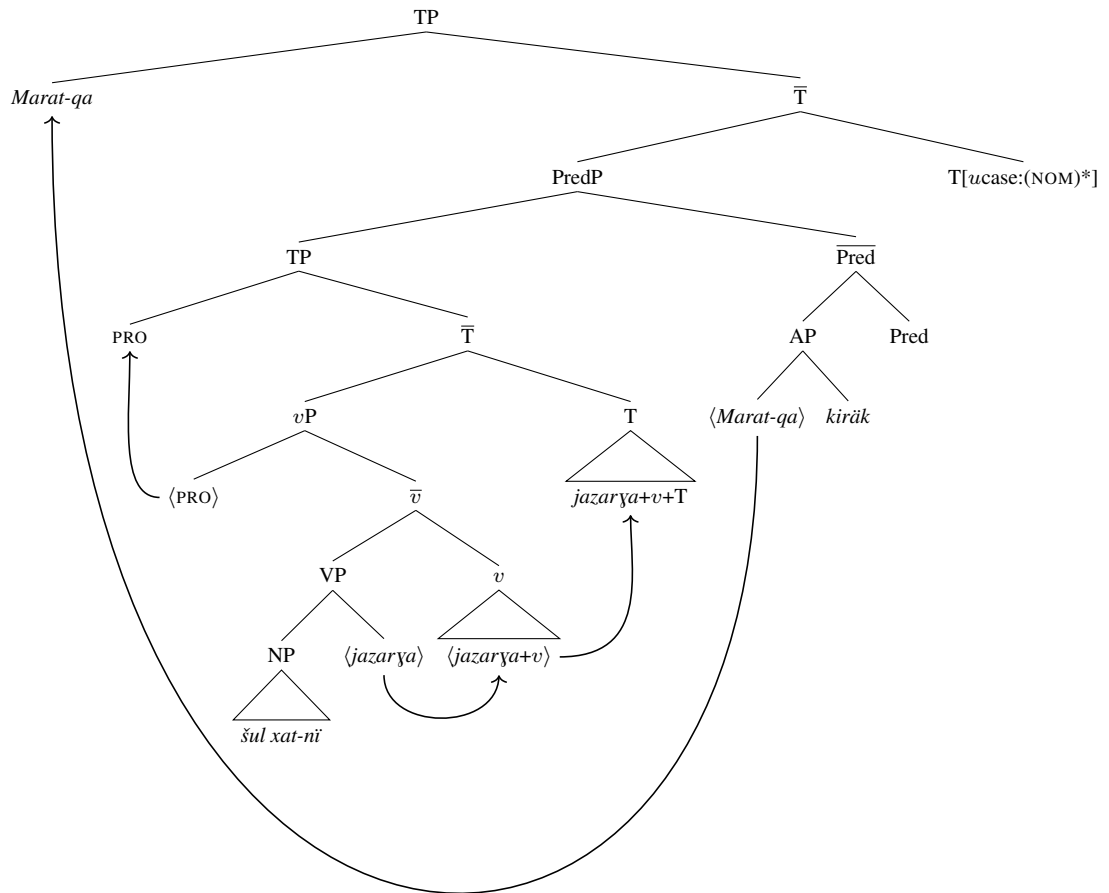
Figure 2.9: The tree structure of (2.47) before the movement to SpecTP.

This prediction arises two problems. First, although the derived syntactic tree shows no syntactic ill-formedness, its acceptability by native speakers is strikingly low. Second, it does not mean that the construction is ungrammatical purely in a syntactic manner, but is grammatical when a suitable context is forcibly specified as in (2.88).[7] Both observations provide us with a hint that the low acceptability is caused in the domain of semantics.

(2.88) Context: Marat is a stubborn boy who does not listen to adults at all. It is necessary for his dad to beat him to make him listen to others.
    Marat-qa   *(äti-se*   *taraf-ïn-nan)*   *qïjna-l-ïrɣa*   *kiräk*
    Marat-DAT   father-POSS.3 side-POSS.3-ABL beat-*pass-inf* necessary
    'Marat needs to be beaten up by his father.'

This observation tells us that the varying acceptability is subject to semantic and pragmatic manipulation rather than to syntactic operation. For this reason, it looks more appropriate to look for the solution in the semantic aspects of the control theory.

---

[7]The data for the example sentence together with the context was provided by a native Tatar speaker.
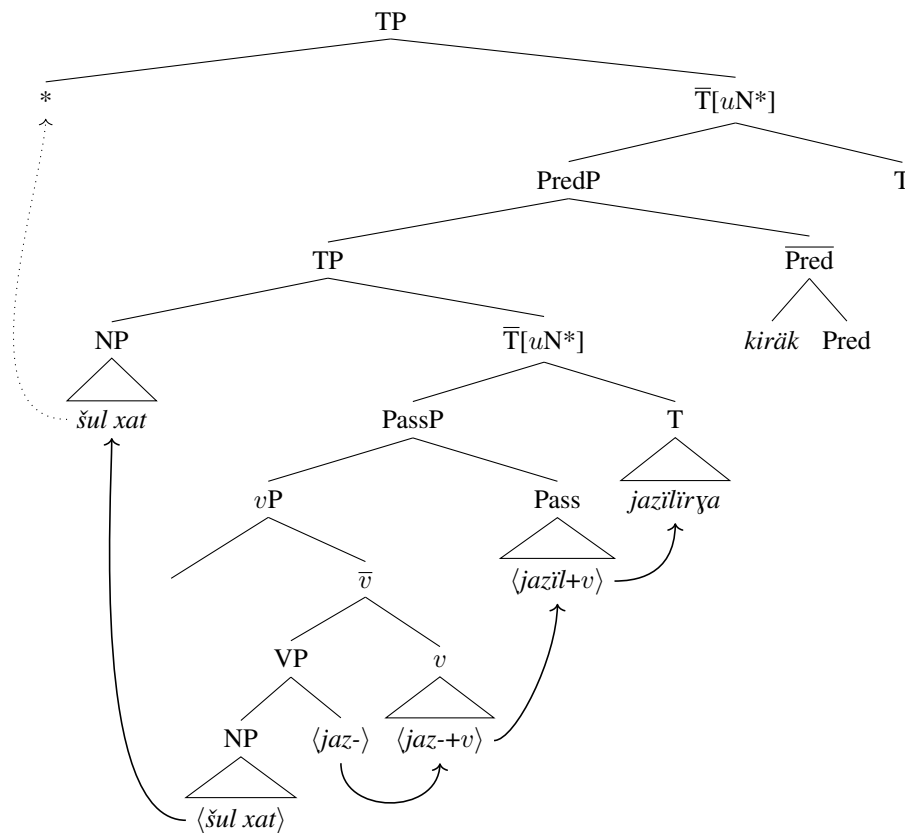
Figure 2.10: THEME in the lower TP, promoted to subject by passivization, bears no nominative case from the infinitive, but has nowhere to go because further upward movement is blocked by the Phase Impenetrability Condition. Therefore the representation predicts the ill-formedness of the passivized *kiräk*-construction.

Farkas (1988) introduces a new concept of the responsibility relation, which is a two-place relation $RESP(i, s)$, where s is a situation and i an individual. The relation is defined as "if $RESP(i, s)$ holds, i is seen as both initiating and controlling s, i.e., the realization of s crucially depends on i (*ibid.*). The prominence of this relation becomes clear when we compare the sentences in (2.89), where (2.89a) is an acceptable expression while (2.89b) sounds unnatural without assuming a specific contextual situation. This semantic difference emerges from the addressee's (i.e., the only participant) responsibility over the situation described by the predicate. The addressee in (2.89a) is responsible for being polite or not and is able to change their attitude by their own will, whereas the addressee in (2.89b), generally speaking, has almost no control over the ability to be tall. Farkas calls the responsible participant the *initiator* of the given situation s.

(2.89)   (a) Be polite!

         (b) # Be tall!

TP

* $\overline{\text{T}}[u\text{N*}]$

PredP T

TP $\overline{\text{Pred}}$

PRO $\overline{\text{T}}[u\text{N*}]$ AP Pred

PassP T NP *kiräk*

*v*P Pass *jazïlïrɣa* *šul xat*

$\overline{v}$ $\langle jazïl{+}v \rangle$

VP *v*

$\langle \text{PRO} \rangle$ $\langle jaz\text{-} \rangle$ $\langle jaz\text{-}{+}v \rangle$
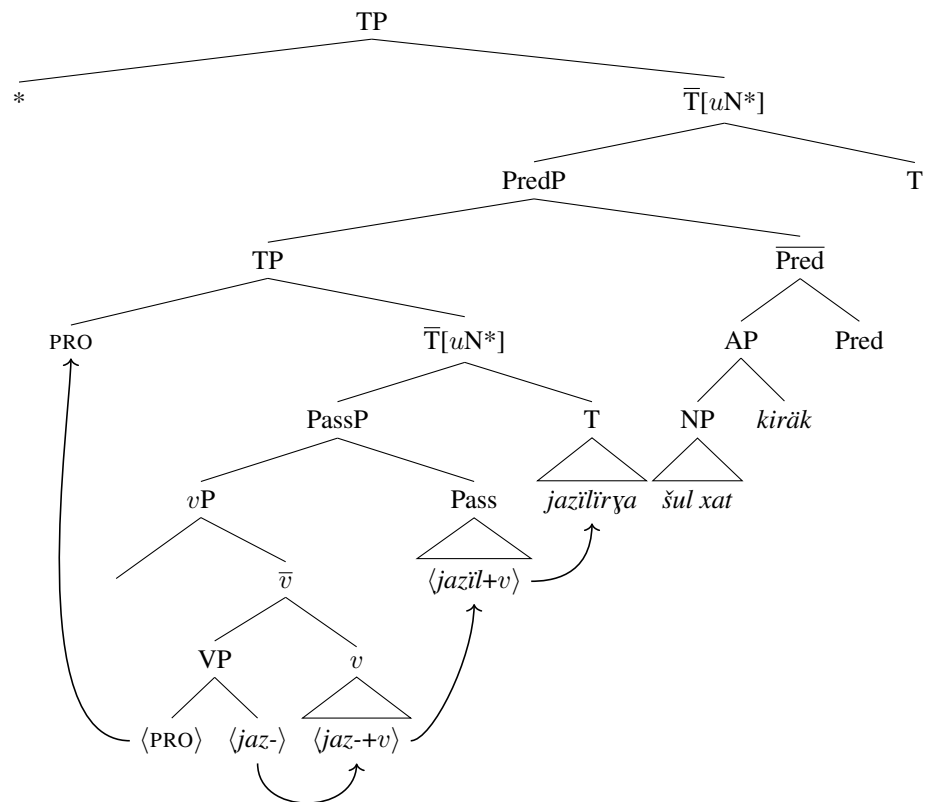
Figure 2.11: THEME in the lower TP, promoted to subject by passivization, bears no nominative case from the infinitive, but has nowhere to go because further upward movement is blocked by the Phase Impenetrability Condition. Therefore the representation predicts the ill-formedness of the passivized *kiräk*-construction.
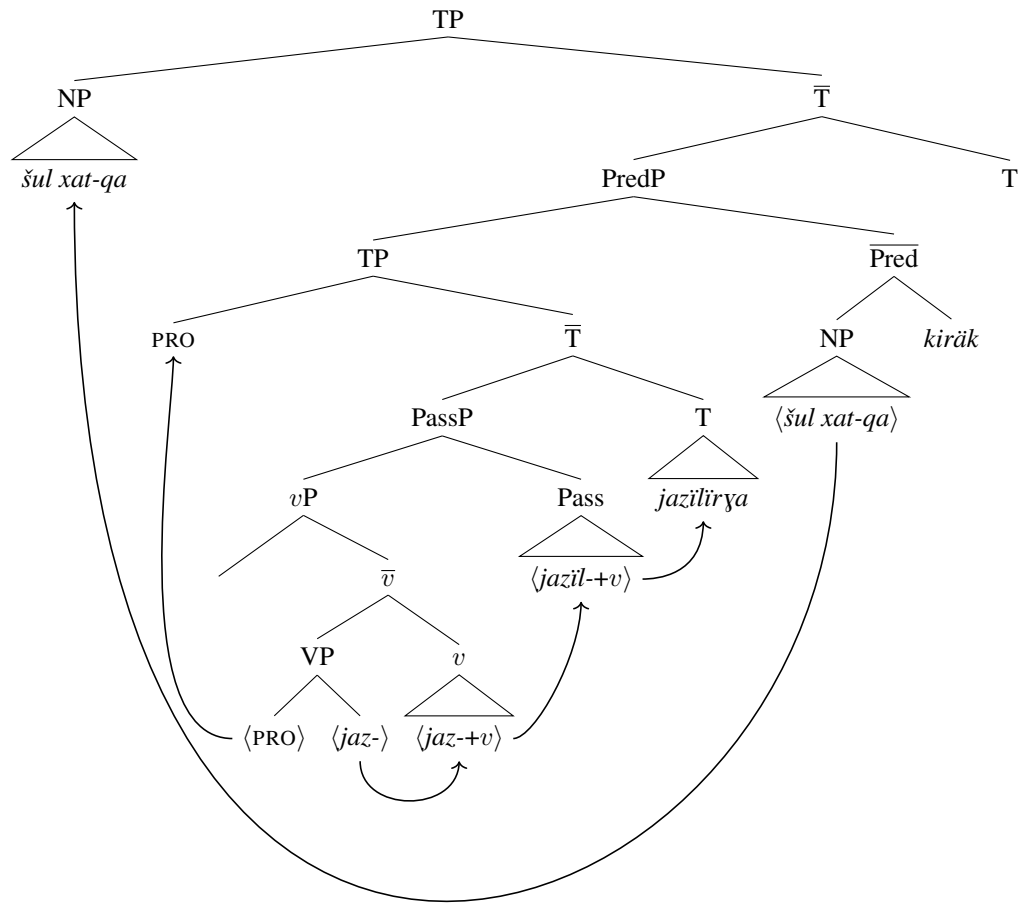
Figure 2.12: Wrong prediction produced by the same derivation as in Figure 2.9.

Though it is often the case that the initiator coincides with the AGENT of the main predicate, Farkas carefully distinguishes them. For example, verbs like *chase* and *run away* assumes two intentional agents: the one who runs after and the one who escapes. It tells us that the syntactic argument structure does not stand solely on the thematic relation but on the perspective as well. In addition, a syntactic subject is not always the initiator of a situation, as it is obvious in passivization, where an object (typically a non-AGENT argument) is promoted to a subject. Thus, the RESP-relation is a semantic notion different from both the argument structure and the thematic relation.

This theoretical device can also explain the switch of the controller as exemplified in (2.90). Note here that the verb *ask* in these two sentences has slightly different meanings; in (2.90a), *ask* is synonymous to *convince*, *order*, and *persuade*, while *ask* in (2.90b) is used in a similar way to *ask for permission* and *appeal*. Intuitively, the responsible participant of the situation introduced by the *to*-infinitive complement with *ask* in the former sense is the second argument, namely Mary. In contrast, in the second sense, it is the first argument, John, who is responsible for the situation described in the *to*-infinitive complement. Thus, the choice of the controller is semantically conditioned by the

RESP-relation. This view implies that what motivates control has to do more with semantics than syntax. Indeed, it seems reasonable given the impression that the predicates that trigger control are often common across languages in terms of their meaning although their language-specific syntactic structure may differ to a great extent.

(2.90) (a) John asked Mary to leave. (PRO = Mary)

(b) John asked Mary to be allowed to leave. (PRO = John)

Now, when we look back at the Tatar sentence and its provided context in (2.88), it seems that the context affects the locus of the responsibility for the situation described by the infinitival clause. That is, while the initiator of the *beating-up* situation is typically its AGENT (here Marat's father), the context forcibly twists the interpretation such that the responsibility of the situation comes from the disobedient child. More technically, the control predicate *kiräk* lexically specifies that the dative argument is a responsible individual for the situation expressed in the infinitival complement, and the context emphasizes that the responsible individual is the patient (the one beaten up). To locate this semantic notion at the interface with syntax, I propose that the control predicate lexically assigns an uninterpreted unvalued semantic feature [uResp:] which specifies the responsible argument in the external matrix clause, and the situation s in the embedded clause, sometimes with the help of the context specification, assigns an uninterpreted valued feature [uResp:+] to the responsible individual i so that the relation RESP(i, s) holds. In order for the sentence to be semantically well-formed, these semantic features must be checked.

2.3.2 Nominal control

This subsection provides an explanation of control with nominal predicates *röxsät*, *waqït*, and *mäslixät*, largely based on the discussion in the previous subsection. Nominal predicates and adjectival predicates, often accompanied by a copula, seem to be structurally different. In semantics, a semantic denotation ⟦**boy**⟧ is a set of individuals who are boys, while ⟦**tall**⟧ denotes a property. In syntax, at least in languages where nouns and adjectives are different categories, nouns and adjectives exhibit a number of different characteristics. For example, nouns can be possessed by a possessor with some syntactic marking, while adjectives cannot. The discussion in this section follows this general view that nouns and adjectives of control in Tatar do have a different internal structure. However, it proposes that the mechanism of control with a nominal predicate is the same as the one with an adjectival predicate.

(2.91)   a John is tall.

b John is a boy.

In Section 2.1, we saw that Tatar can have peculiar infinitival constructions with five nouns: *röxsät*, *waqït*, *isäp*, *nijät*, and *mäslixät*. We can further categorize these nouns into two groups: the group that takes the AGENT in dative (*röxsät*, *waqït*, and *mäslixät*) and the one that takes the AGENT in genitive (*isäp* and *nijät*). The rest of this subsection focuses on each of these two groups and provides consistent explanations of their syntactic mechanism.
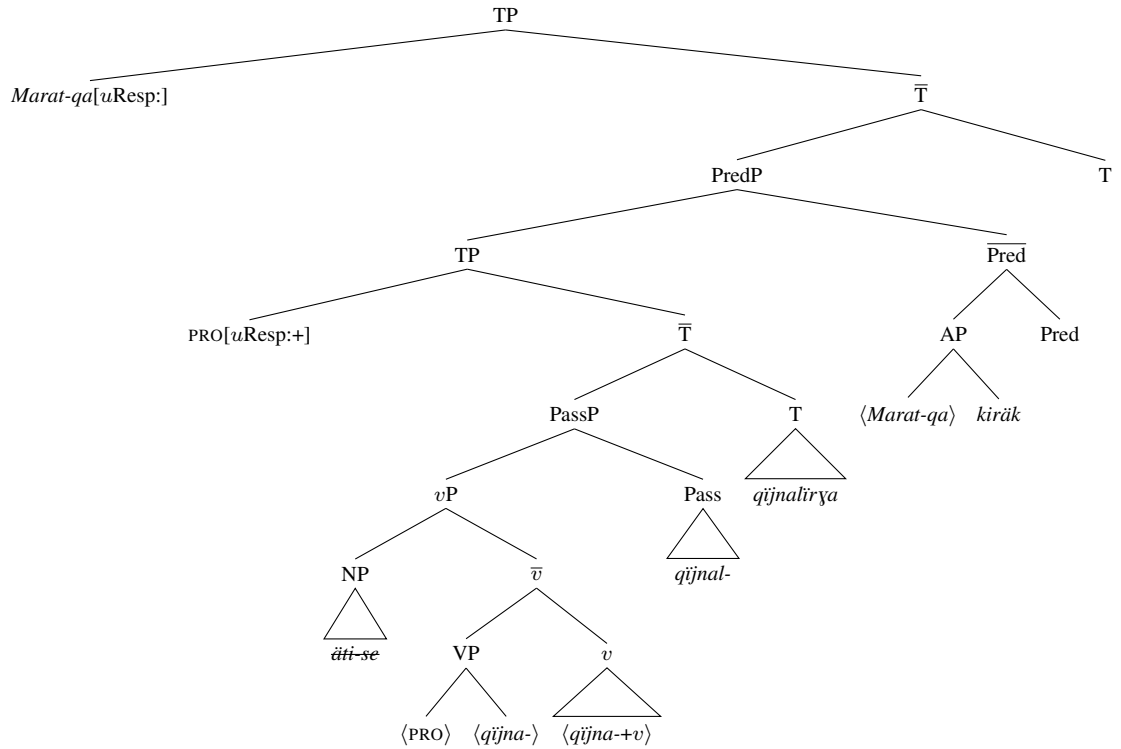
Figure 2.13: A syntactic representation of (2.88) with the proposed semantic feature [Resp] for the responsibility of the situation.

Röxsät-type nominal control with a dative agent

We have seen before that the nominal predicate *röxsät* can compose a control construction as in (2.92). Now, we start with an observation of a similar sentence with a synonymous verbal predicate *röxsät bir-* 'to give a permission' as in (2.93), which has a mediocre structure where the verb is *bir-* which subcategorizes for a subject (AGENT), a direct object (THEME), and an indirect object (RECIPIENT). Assuming that V and the specifier of VP (the direct object) are attracted to v, the syntactic representation of (2.93) is illustrated in Figure 2.14. The position of the arguments in the tree is following the UTAH.

(2.92) *Marat-qa   kit-ärgä   röxsät*
        Marat-DAT   leave-INF   permission
        'Marat is allowed to leave.'

(2.93) *Bulat   Marat-qa   kit-ärgä   röxsät   bir-de*
        Bulat.NOM   Marat-DAT leave-INF permission give-PST.3
        'Bulat allowed Marat to leave.'

Now, when we compare this canonical construction with the nominal control construction in (2.92), we notice the following differences. First, the AGENT (the provider of the permission) is
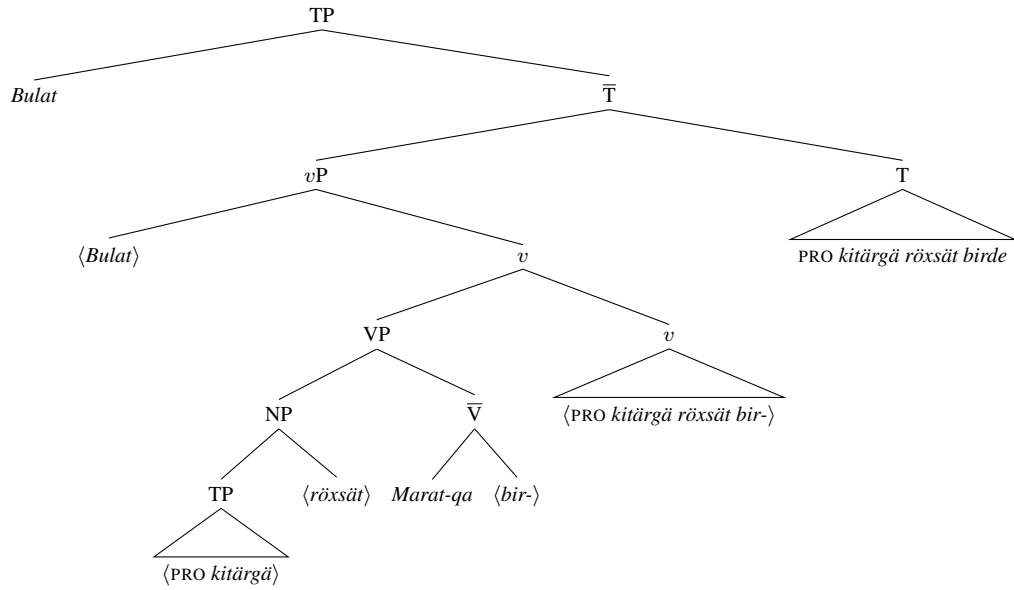
Figure 2.14: A syntactic tree for (2.93).

not specified in the nominal control. Second, the THEME noun *röxsät* in the canonical construction is the predicate itself in the control construction. This (dis-)similarity reminds us of the little n, a functional category used to account for the thematic relation around a noun instead of a verb. When we consider sentence (2.94), it is clear that *the enemy* is the agent of *destruction* and *the city* the patient thereof. This fact tells us that certain nouns like *destruction* can have semantic arguments and have some means to encode their AGENT under some projection. To distinguish from vP which is an extension of the projection of VP, we call this nP. A template of the structure of nP is illustrated in Figure 2.15.

(2.94) The enemy's destruction of the city

Based on this observation and the syntactic derivation attempted in the previous subsection, we can think of a structure for the nominal control in parallel with the canonical construction. That is, in (2.92), the AGENT is empty, and the content of the permission occupies the position for THEME. Now, we need to extend the structure of nP to cover the RECIPIENT (GOAL) of the permission. Fortunately, taking the UTAH in the verbal structure for granted, we can just assume that the RECIPIENT is under an intermediate projection of the noun. Figure 2.16 is a visualized structure of the extended version with branching fit to the Tatar syntax.

Since there is no AGENT in this construction, the sole nominal argument *Marat-qa* is raised to the specifier of TP to satisfy the EPP, just like the mechanism in the *kiräk*-construction described in the previous subsection.
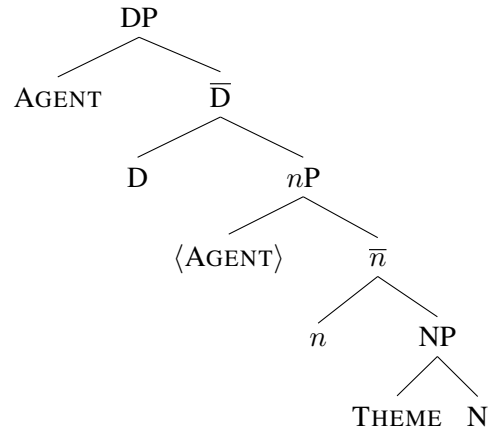
Figure 2.15: A structure of nP under DP (based on (Adger, 2003)).
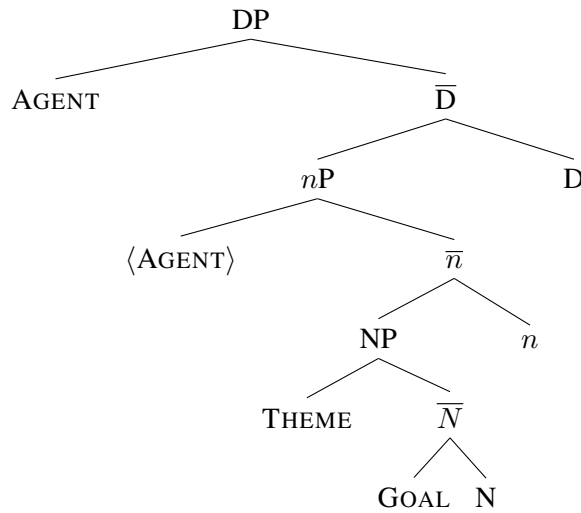
Figure 2.16: Caption

This analysis has merits in the following points. First, it represents well the fact that the noun *röxsät* in this construction is different from that in (2.93). The structure of the nominal control construction appears to involve a functional category (here nP) higher than a lexical category (NP), which reflects thematic relation like vP. This enables the controller to be maintained as an oblique dative argument functioning as the RECIPIENT which is moved to the sentence-initial position by virtue of the EPP. Most importantly, the core idea of this structure is similar to the one we assumed in the *kiräk*-type adjectival control construction in the sense that the dative argument is generated inside the projection of the lexical category (NP or AP).

TP

Marat-qa   T̄

PredP   T

*n*P   Pred

Ø   n̄

NP   *n*

TP   N̄

PRO   T̄   ⟨*Marat-qa*⟩   *röxsät*

kitärgä
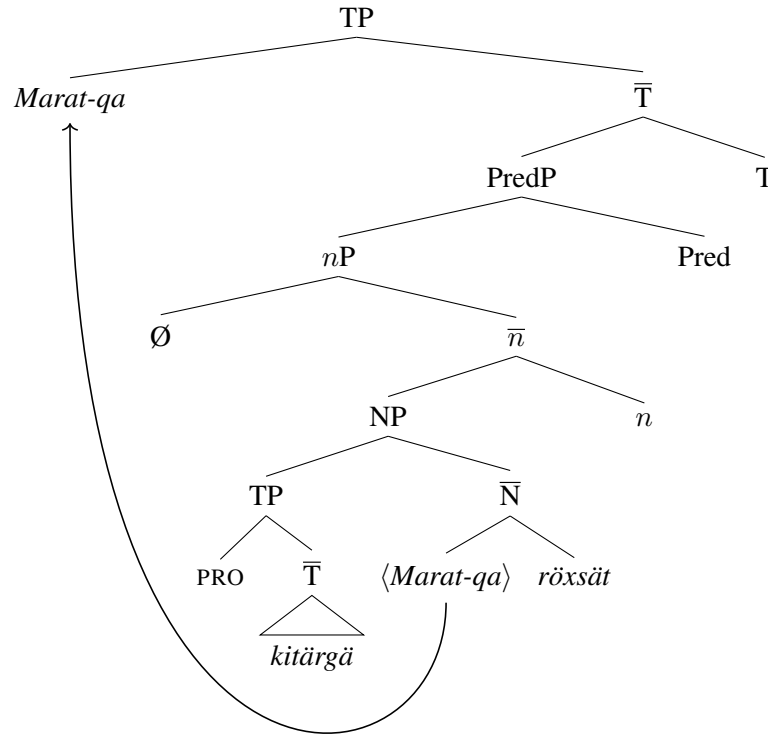
Figure 2.17: A proposed syntactic structure of (2.92).

Isäp-type nominal control with a genitive agent

One of the unique characteristics of the *isäp*-type construction, including nouns *isäp* and *nijät*, is that the AGENT is marked by genitive. More interestingly, the nominal predicate *isäp* can concord with the genitive (possessor) in person and number by a possessive suffix as concisely shown in (2.95). Although it looks like just a nominal phrase ('Marat's thought to leave'), it is in fact a full-fledged sentence. Upon analyzing this construction, we start with the assumption that it also involves *n*P. This is more similar to the textbook example in (2.94) in the way that the AGENT is expressed in a possessive form. In fact, marking an AGENT with a possessive form is also found in other related constructions such as a verbal noun (2.96). Taking these into account, we will assume a templatic structure in Figure 2.18. The external argument is introduced at a higher branch than NP, namely as the specifier of *n*P. For example, (2.96) can be represented as in Figure 2.19.

(2.95) *Marat-nïŋ   kit-ärgä   isäb-e*
      Marat-GEN   leave-INF   thought-POSS.3
      'Marat is going to leave. (*lit.* Marat's thought to leave)'

(2.96) *Marat-nïŋ   äjt-üw-e*
      Marat-GEN   say-NMLZ-POSS.3
      'what Marat says'
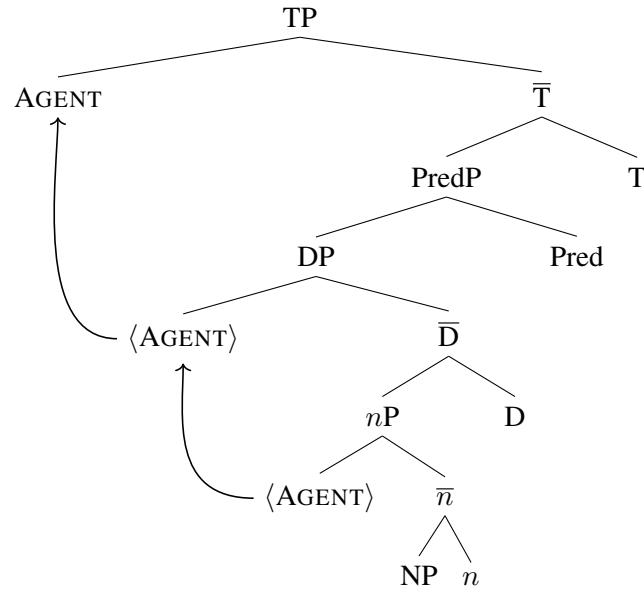
Figure 2.18: A templatic structure of the *isäp*-type control construction.

Another peculiar point of this construction is that the polar question enclitic {=mI} can be attached to the infinitive (2.98), though it is generally a strictly sentence-final element . Some speakers may allow the enclitic {=mI} at the sentence-final position as in (2.99), but others definitively judge that it is ungrammatical. According to the Corpus of Written Tatar (Saykhunov et al., 2021), the number of hits with the form "`<inf>`, исәпме" (*isäp=me* directly follows an infinitive) is 16, while the one with "`<inf><q>`, исәп" ({=mI} is attached to the infinitive, followed by a bare *isäp*) is 37.[8] Though this is not a decisive factor to stand on either side, at least it shows that the both forms can be acceptable. However, attaching the polar question enclitic to the infinitive like (2.98) is not grammatical with the *röxsät*-type construction (2.100).

(2.97) *Marat   kit-äčäk=me?*
      Marat.NOM   leave-FUT.PTCP=Q
      'Will Marat leave?'

(2.98) *Marat-nïŋ   kit-ärgä=me   isäb-e?*
      Marat-GEN   leave-INF=Q   thought-POSS.3
      'Is Marat going to leave?'

(2.99) ?*Marat-nïŋ   kit-ärgä   isäb-e=me?*
      Marat-GEN   leave-INF   thought-POSS.3=Q
      'Is Marat going to leave?'

(2.100) **Marat-qa   kit-ärgä=me   röxsät?*
      Marat-DAT   leave-INF=Q   permission

---

[8]Retrieved on 13 August, 2022.
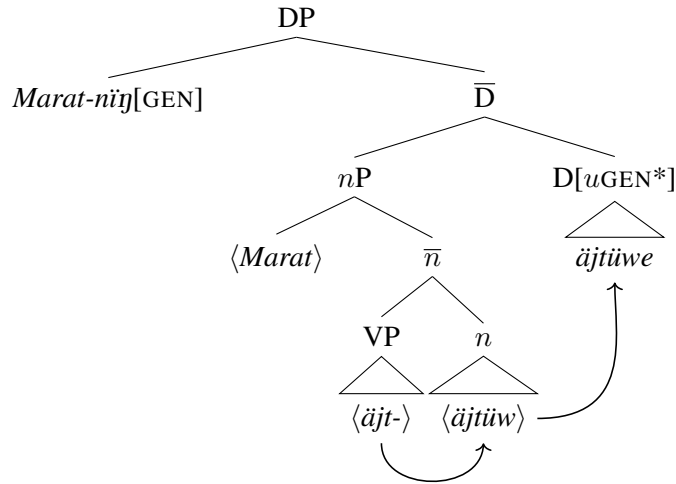
Figure 2.19: An exemplary structure of (2.96).

This constrast shows that the structure of the *isäp*-type control construction is somewhat different from other control constructions. A possible explanation to the nature of the infinitive in the *isäp*-type construction is that it involves CP, not TP as in other control constructions in Tatar. CP, a functional category headed by a complementizer, is known to be responsible for the clause type in some languages. For example, in English, subject-auxiliary inversion in polar question sentences headed by C having the interrogative feature (Adger, 2003). In the same manner, we can assume the enclitic {=mI} is a responsible constituent for selecting the clause type at C. In this case, for example, the structure (2.97) should look like Figure 2.20.
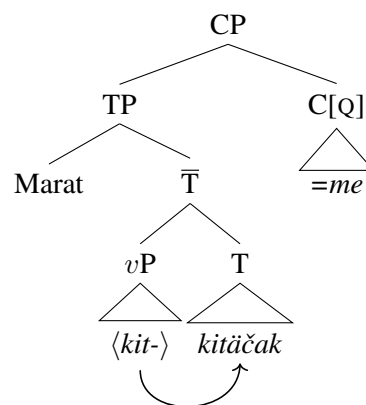
Figure 2.20: A tree representation of (2.97) with CP.

A motivation to assume that {=mI} is C, in addition to the fact that it operates on the clause type, is that it is strongly sentence-final (i.e., word-final as well). This is a different morphotactic from some other Turkic languages. In Uyghur, for example, the cognate interrogative marker *-m(u)* is also a sentence-final element, but precedes a personal suffix. Therefore, it seems that the Uyghur

polar interrogative morpheme is part of the morphology of a predicate. In contrast, the enclitic {=mI} in Tatar is independent of the verbal or adjectival morphology, albeit prosodically dependent on its preceding element.

(2.101) Uyghur
  yaxshi-mu-siz?
  good-Q-2SG.HON
  'Are you well?'

(2.102) Uyghur
  bar-am-siz?
  go-Q-PRS.2SG.HON
  'Do you go?'

Summarizing the points, Figure 2.21 illustrates the structure of (2.98), which contains CP as a daughter of NP.
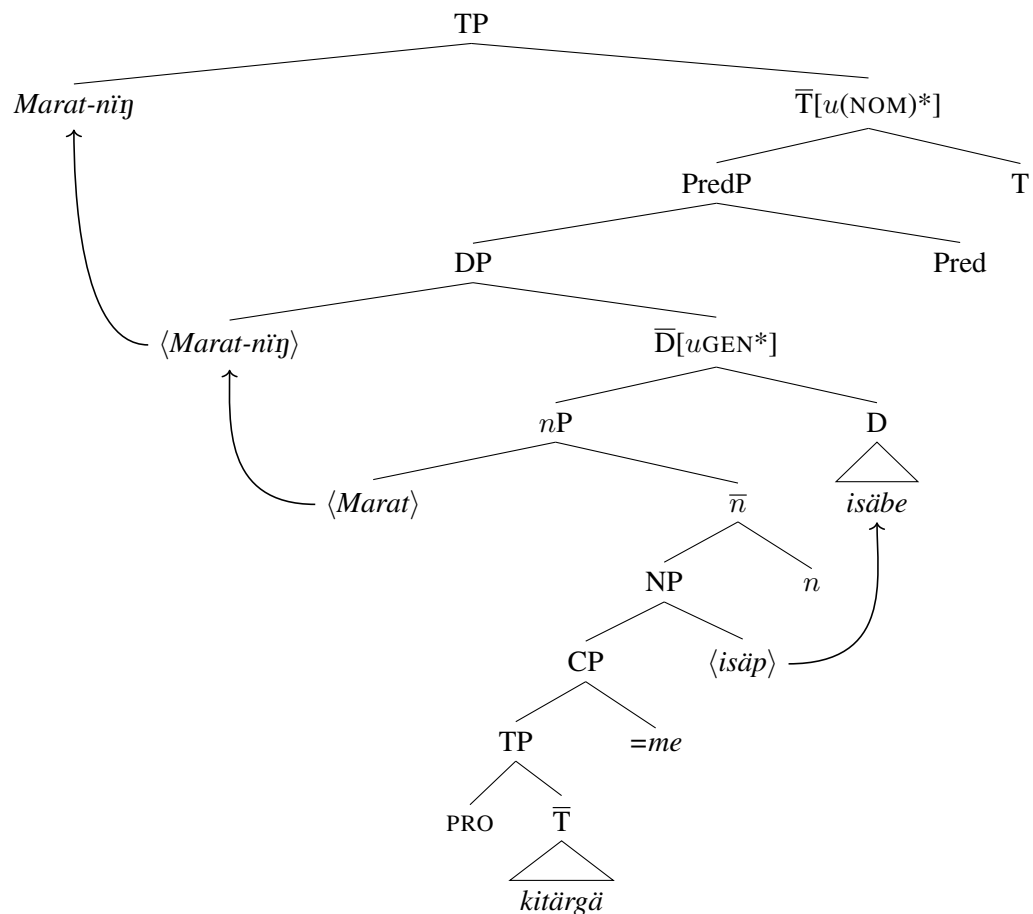
Figure 2.21: The structure of (2.98).

2.4   Interlude: Dative–Infinitive Construction

In the previous section, we observed control constructions with non-verbal predicates in Tatar. Before moving on to the raising construction, we will take a short look at a closely related construction: the dative–infinitive construction (DatInf henceforth). DatInf refers to a sentence that contains a bare infinitive and its agent coded in the dative case. Burbiel (2018), which provides a comprehensive description of DatInf, categorizes its usage as follows:

1. A categorical command or prohibition (in negative), e.g. (27);

2. A Decision, e.g. (28);

3. An emphatic personal advice, suggestion, instruction etc., e.g. (29);

4. An official instructions, orders, rules etc., e.g. (30);

5. Inquiries, questions, e.g. (31).

(27) *qubrat xan   berenče bul-ïp   awaz sal-dï*:   *quzɣal-ïrɣa*!
      Qubrat Khan first   be-CVB mouth put-*pst.3* move-INF
      'Qubrat Khan was the first to cry: "Move!"'(Burbiel, 2018)

(28) *mondïj qurqïnïč uj-lar-nï   quw-arɣa*!
      such   scary   thought-PL-ACC chase-INF
      'Such scary thoughts must be chased away!' (Burbiel, 2018)

(29) *daimi häjbät tuqlan-ïrɣa, bolïn   hawa-sï   sula-rɣa,   čišmä suw-ï   eč-ärgä, ...*
      always well   eat-INF   meadow air-POSS.3 breathe-INF spring water-POSS.3 drink-INF
      '(You should) always eat well, breathe meadow air, drink spring water, ...' (Burbiel, 2018)

(30) *uqïtuw-tärbïjä   process-ï-n   daimi kamilläš-ter-ergä*
      teaching-education process-POSS.3-ACC always improve-CAUS-INF
      'Education process should be constantly improved' (Burbiel, 2018)

(31) *źïjïntïq   turïnda söjlä-š-ergä=me*?
      collection about   speak-REC-INF=Q
      '(Are you here) to talk about the collection (of poems)?' (Burbiel, 2018)

The characteristic common among all the types is the deontic flavor that is further specified by context. What is striking about the DatInf constructions is that it includes neither typical finite verbal predicate that is usually inflected by tense and person nor nominal-adjectival (copulaless) predicate. In addition, the agent of the infinitive verb is marked by the dative case, not nominative that agents typically bear. It is also an interesting issue in terms of semantics that, at the level of the surface, the

meaning does not seem to be compositional, because there is no pronounced element contributing to the modal meaning.

In fact, similar DatInf constructions are found in several Slavic languages, for example Russian as a language relatively well-studied on this construction (Bailyn, 2004, 2011; Timberlake, 2004; Wiemer, 2017). The semantics of the Russian DatInf constructions also chiefly functions as modal, and the modality itself is polysemous (e.g., deontology, imperative, possibility, etc.) and depends on its context. Timberlake (2004) notes that the DatInf construction has historically developed to the use of the infinitive with certain nonverbal predicates, such as *nado* "necessary", *možno* "possible", *vozmožno* "possible", and *nel'zja* "impossible, impermissible ", which are to-day more commonly used than the DatInf construction.

Are the DatInf constructions in Tatar and Russian biclausal or monoclausal? Before proceeding to the discussion in Tatar, let us first take a look at the treatment of DatInf in Russian, on which there is more abundant literature. The analysis of its clausal structure is still controversial in Russian syntax (Bailyn, 2011). Some linguists stand on the monoclausal analysis for the Russian DatInf construction (Bailyn, 2004). Bailyn argues that the dative case is assigned by a higher category C, and the infinitive does not morphologically agree with nominative. Partial evidence supporting that the dative case is assigned by C is that an infinitival clause with an overt complementizer *čtoby* "in order to" triggers a dative agreement with a reflexive element. In sentence (2.103) which is assumed to contain no C in the infinitival clause, dative case-marking on the reflexive pronominal element *odin* "alone" is not allowed. On the other hand, the grammaticality of agreement with dative in (2.104), which has an overt C *čtoby*, implies that it is C that causes this difference by assigning the dative case.

CP
- C[+Modal]
- TP
  - NP.DAT T̄
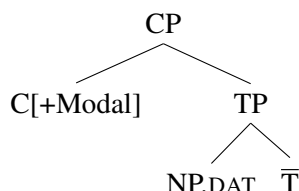
Figure 2.22: A monoclausal structure hypothesis of the DatInf construction given by Bailyn (2004).

(2.103) Russian (Bailyn, 2011)

Ivan	*xočet*	[*tanceva-t'*	*odin*	/	**odnomu*]
Ivan.NOM	want.PRS.3SG	dance-INF	alone.NOM	/	alone.DAT

'Ivan wants to dance alone.'

(2.104) Russian (Bailyn, 2011)

Ivan	*prišel,*	[*čtoby*	[*tanceva-t'*	**odin*	/	*odnomu*]]
Ivan.NOM	arrive.PFV:PST.SG.M	in.order.to	dance-INF	alone.NOM	/	alone.DAT

'Ivan arrived in order to dance alone.'

An alternative explanation posits that this construction is in fact biclausal (Fleisher, 2006; Sigurðsson, 2002). Based on the fact that Russian is a null-copula language in the present tense, this

position assumes that there is an unpronounced copular predicate in the main clause in addition to the infinitival predicate in the embedded clause. For instance, Fleisher (2006) gives the following structure in Figure 2.23.[9] In this structure, the unpronounced *byt'* has the dative argument as an internal argument, not an external argument, which controls the PRO in the infinitival clause. The dative case-marking of the pronoun in the infinitival clause (also known as "second dative") (2.104, 2.105) is as a result of the agreement with the PRO which the dative case is assigned to as a structural case and which is the only constituent that can bear a structural dative case from an infinitive (Fleisher, 2006).
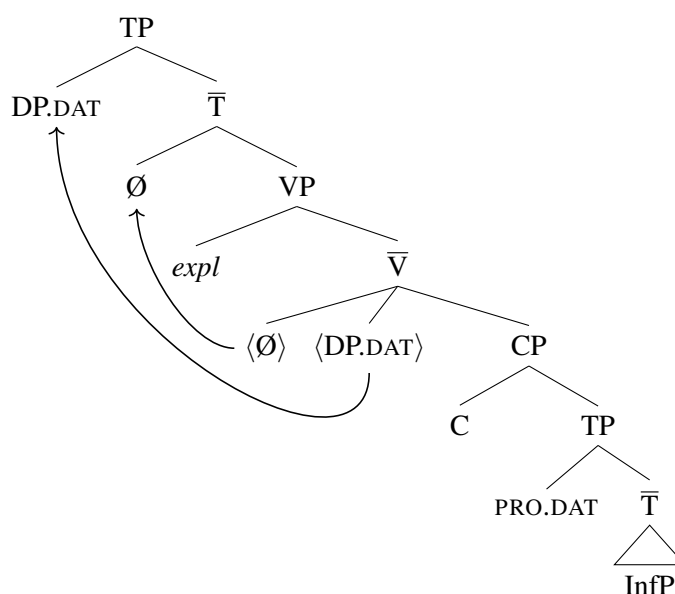
Figure 2.23: The structure of DatInf proposed by Fleisher (2006).

(2.105) Russian (Timberlake, 2004)

vozmožnost' [PRO   *poexa-t'*   *sam-omu*]
possibility   PRO.DAT   PFV:go-INF   oneself-DAT.SG
'the possibility of going alone'

A motivation for this analysis is the restriction on the position of a negation particle *ne*. In a sentence with a finite verb, the negation particle precedes the main verb (2.106). However, in the DatInf construction in non-present tenses, the negation particle preceding the overt copular verb *byt'* is not grammatical (2.107); it has to be placed after *byt'* (2.108). This contrast evidently shows that the syntactic structure of the DatInf construction in Russian is different from canonical sentences with a finite verb. Bailyn (2011) and Fleisher (2006) mention this as partial evidence for the biclausal structure of the Russian DatInf. Some monoclausalists like Moore & Perlmutter (1999), however, dodge the biclausal analysis of this data, and argue for the monoclausal analysis claiming that *bylo* and *budet* (past and future forms of *byt'*, respectively), seemingly inflected verbs, are in fact temporal particles, and therefore they cannot host negation.

---

[9]The notation for grammatical categories is slightly changed to be consistent with the notation used in this thesis.

(2.106) Russian
 gruzovik-i ne proexa-l-i
 truck-NOM.PL NEG go.through.PFV-PST-PL
 'The trucks did not get through.'

(2.107) Russian (Moore & Perlmutter, 1999)
 *gruzovik-am ne by-l-o proexa-t'
 truck-DAT.PL NEG be-PST-SG.N go.through.PFV-INF

(2.108) Russian (Moore & Perlmutter, 1999)
 gruzovik-am by-l-o ne proexa-t'
 truck-DAT.PL be-PST-SG.N NEG go.through.PFV-INF
 'It was not in the cards for the trucks to get through.'

As we have seen, the clausal status of the DatInf construction in Russian has seen no conclusion agreed upon. However, at least in Tatar, based on these arguments and the provided syntactic evidence, we argue that the DatInf in Tatar has a biclausal structure similar to the one Fleisher (2006). To begin with, Tatar infinitives do not assign a structural dative case. In Russian, when a controller is non-nominative, the controlled PRO is assumed to have a dative case because the reflexive pronoun exhibits an agreement with dative (2.109). However, in Tatar, the reflexive pronoun is nominative (2.110); therefore, the infinitives do not structurally assign dative to their subject.

(2.109) Russian (Timberlake, 2004)
 ona {*prikaza-l-a mne ~ umolja-l-a menja*} [PRO *poexa-t'*
 3SG.F order-PST-SG.F 1SG.DAT beseech-PST-SG.F 1SG.ACC PRO.**DAT** PFV:go-INF
 sam-omu]
 oneself-**DAT**.SG
 'She {ordered me ~ beseeched me} to go myself'

(2.110) *Marat* *Bulat-nï* [PRO {*berüz-e* / **berüz-en-ä*}
 Marat.NOM Bulat-ACC PRO oneself-POSS.3.**NOM** / *oneself-POSS.3-DAT
 joqla-rɣa] *künder-de*
 sleep-INF persuade-PST.3
 'Marat persuaded Bulat to sleep by himself (alone).'

Second, the grammatical condition of the Tatar DatInf construction resembles that of the control with adjectival predicates discussed in the previous section. That is, it is not generally accepted for non-AGENT NPs to be the sentence-initial dative argument (cf. Table 2.1), whereas it is possible in Russian (2.111). This less subject-like characteristic supports Fleisher's (2006) hypothesis that the dative argument in the DatInf construction is an internal argument of the phonetically null predicate. Therefore, the structure of the Tatar DatInf (2.112) can be written out as Figure 2.24. Namely, (i) the DatInf construction in Tatar has a biclausal structure; (ii) the main predicate is phonetically null in the present tense; (iii) the dative argument originates as an internal argument of the main null predicate and is moved to the specifier of TP by virtue of the EPP; (iv) the dative argument controls the PRO in the infinitival clause; (v) the PRO is structurally nominative.

(2.111) Russian (Timberlake, 2004)
by-t'　*groze!*
be-INF　thunder.DAT.SG
'There will be thunder!'

(2.112) *Marat-qa*　*kit-ärgä*　*Ø*
　　Marat-DAT　leave-INF　PRED
'Marat must leave!'
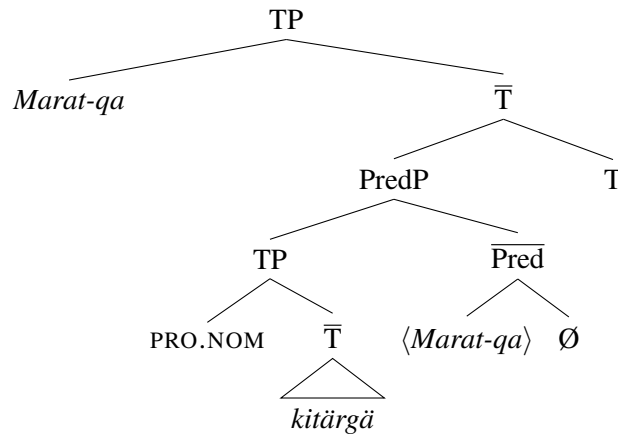
Figure 2.24: The tree structure of (2.112).

The biclausal analysis of the Tatar DatInf is also favorable to analyze a double negation in (2.113). In this case, each negation phrase is inserted after $\overline{\text{T}}$ following the Hierarchy of Projections T > Neg > PRED, where PRED includes functional categories v and Pred.

(2.113) Zaripov (2013)
[PRO *eč-mäskä*]　　*Ø*　　*tügel,*　([PRO　*iser-mäskä*]　　　*Ø*)
PRO　drink-NEG.INF　PRED　NEG　PRO　get.drunk-NEG.INF　PRED
'It is not that one shouldn't drink, it is that one should not get drunk'

Last but not least, the assumption that the DatInf constructions in Tatar involve a phonetically null predicate can solve the problem of non-compositionality of meaning. If we take a view that the meaning of a sentence is constructed by applying the meaning of each constituent, the DatInf construction is problematic because there is no overt lexical source of the modal meaning. Although Bailyn's (2004) monoclausal analysis accounts for the compositionality with the hypothesis that an unpronounced C has a feature of some modality, it is not very clear why C itself is responsible for the modality because C is typically filled by a complementizer.[10] By assuming a null predicate, we can interpret that the null predicate is lexically responsible for the modal meaning just like overt predicates.

---

[10]The view that C can be responsible for the modal feature might be related to the verbal morphosyntax of subjunctives in several Indo-European languages. For example, French subjunctives seem to have a strong relationship with the nature of an embedded CP under *que* with a subjectified meaning.
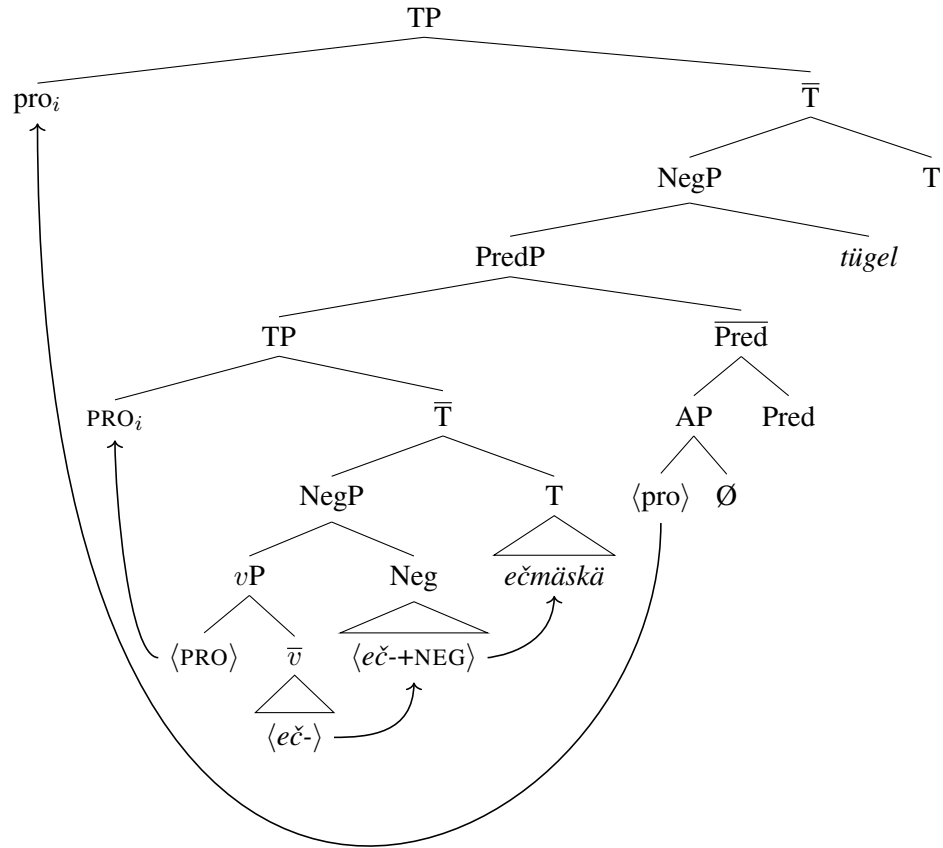
Figure 2.25: The structure of (2.113).

2.5 Summary of this chapter

The peculiar fact that we saw as the first observation of these constructions was that, although *kiräk* and *tiješ* share the more or less similar deontic meaning, they are superficially different in case-marking, where *kiräk* is accompanied by a dative (pseudo-)subject and *tiješ* by a nominative subject. To this issue, this chapter has given an answer that these constructions are in fact of different kinds. Specifically, the infinitival construction with *kiräk* is involves control, while *tiješ* triggers subject-to-subject raising. In addition, this chapter provided a descriptive report of control with nominal predicates: *röxsät*, *waqït*, *isäp*, *nijät*, *mäslixät*. The functional use of these Tatar nouns in the control construction itself is a newly reported phenomenon that has not been discussed in the descriptive grammars of Tatar. Categorizing these nouns in two types, the chapter further gave theoretical analyses as to why these constructions are instances of control, how their control syntactically works, and how it can be derived consistently within the framework of Minimalist syntax.

It is also worth mentioning the chapter revealed that the non-verbal control constructions syntactically resemble the DatInf construction in Tatar. Section 2.4 showed that, with the comparison to

the similar DatInf construction found in Russian, the Tatar DatInf construction is best analyzed as biclausal, where the dative oblique argument of a null predicate in the main clause controls the PRO of the subordinate clause.

Chapter 3

Mermaid Construction: Control and Raising with a Nominal Predicate

In the previous chapter, we dived into control and raising constructions with non-verbal predicates in Tatar. Above all, it was an interesting finding that several control constructions employ noun predicates: *röxsät*, *waqït*, *isäp*, *nijät*, *mäslixät*. While it is common to find control constructions with a verbal or adjectival predicate across languages, those with a nominal predicate are understudied linguistic phenomena, though of course it is no wonder to find such a construction given that nouns can constitute a one-place predicate just like adjectives. Upon an attempt to theoretically account for the idiosyncratic syntactic phenomena, it is often helpful to look for similar cases in other languages. That is, are there any similar constructions in other languages?

A keyword in this chapter is the Mermaid Construction. The main argument of this chapter is that the control constructions with a noun predicate we saw in the previous chapter can be integrated with the study of the Mermaid Construction — or the latter be integrated with the former. Section 3.1 will concisely explain what the Mermaid Constructions are and provide an overview of the descriptive studies on them that have been accumulated so far and Tsunoda's (2020a) monoclausalist hypothesis thereof. In doing so, we aim to confirm that the syntactic phenomenology of the Tatar nominal control predicates proves to be part of the crosslinguistically reported Mermaid Constructions. In Section 3.2, by comparing raising and control with the Mermaid Constructions generally, I will further argue that the Mermaid Constructions consist of not only control constructions but also raising constructions. This is a fundamentally opposing argument to the monoclausal analysis, because raising and control have a biclausal structure assuming that they have an embedded TP. Then, section 3.3, in favor of the biclausal analysis, provides direct counterarguments against the monoclausality of the Mermaid Constructions. Sections 3.4 and 3.5 further argue for the biclausalist hypothesis of the Mermaid Constructions as raising and control. The former section demonstrates that semantics of the Mermaid Constructions are the same as that of raising and control by deriving their scopal (un)ambiguity with an existential quantifier. Then, the latter juxtaposes the nature of Mermaid Constructions with raising and control in the study of grammaticalization.

3.1   Overview of Mermaid Construction

Mermaid Construction, coined by the linguist Tasaku Tsunoda (2011, 2012, 2020a), is a construction that typically has the form shown in (3.1). Though there are crosslinguistic differences, the Clause's verb is usually in a non-finite form, and the accompanying noun predicate consists of a noun and a copula. Note that the structure in (3.1) does not specify the linear order of each element, and the order can vary in accord with language-specific word order. At least, Tsunoda (2020a) assumes that the Clause's verb and the noun are linearly adjacent and they compose a "compound predicate", which we will argue against in detail in Section 3.2.

Even though the structure looks like a nominal sentence lacking a finite verb, it is semantically similar to a verbal sentence with some auxiliary meaning. An example of Mermaid Constructions in Japanese is given in (3.2).[1] If we take it as a sentence containing a noun predicate which a verb phrase adnominally modifies, it would literally mean "Hanako is the feeling to grill the fish", which is not what (3.2) truly means.

(3.1) [(Non-finite) Clause] + Noun (+ Copula)

(3.2) Japanese (< Japonic); Japan; SOV
    Hanako=ga   *sakana=wo* *yak-u*   *ki=da*
    Hanako=NOM fish=ACC   grill-NPST.ADN feeling(N)=COP
    'Hanako intends to grill the fish. (*lit.* Hanako is the feeling to grill the fish.)'

In fact, the Mermaid Construction shown in (3.2) is not the same as a sentence with a typical nominal predicate like in (3.3). Compare the sentences in (3.2) and (3.3). In (3.2), the nominal predicate is not coreferential with the subject, as succinctly represented in semantic notation in (3.4). On the other hand, in the typical nominal predicate sentence (3.3), the denotation of an individual "Hanako" is included in the denotation of the nominal predicate (a set of individuals who grill the fish), as formalized in (3.5).

(3.3) Japanese
    Hanako=ga   [*sakana=wo* *yak-u*]   *hito=da*
    Hanako=NOM person=ACC grill-NPST.ADN person=COP
    'Hanako is the person who grills the fish.'

(3.4) ⟦**Hanako**⟧ ⊄ ⟦**feeling-to-grill-the-fish**⟧ (3.4)

(3.5) ⟦**Hanako**⟧ ⊆ ⟦**person-who-grills-the-fish**⟧ (3.5)

---

[1]Tsunoda (2011) notes that the construction is named "Mermaid Construction" because the [Clause] looks like a verbal sentence while the part outside the [Clause] looks like a nominal sentence. However, this is a confusing naming because of the following two reasons. First, the [Clause] does not look like a verbal sentence because it is not a finite verbal clause but is non-finite in most of the cases. Second, metaphorical usage of "mermaid", with the upper body of human and the tail of fish, creates a false concept that the construction consists of two linearly organized parts, a verbal part and a nominal part. However, a crosslinguistic scrutiny reveals that the elements can be coordinated in a more complex manner.

3.1.1 Crosslinguistic Data

Mermaid Constructions are found in languages across language families. Tsunoda (2020a), a crosslinguistic descriptive study of the Mermaid Constructions, mentions a tendency that the distribution of the Mermaid Constructions is concentrated in languages of Asia, in particular of East Asia, as listed in (3.7)-(3.14). The examples below show the language's name with its genetic group, chiefly spoken area, canonical word order, and the data source.

(3.6) Mitsukaido dialect (< Japanese < Japonic); Ibaraki (Japan); SOV (Sasaki, 2020)
are{=Ø / =wa / =ŋa / =no} *uresiŋar-u* *segi=zja=nee*
3SG=NOM / =TOP / =POSS / =GEN be.pleased=NPST seat=COP.IRR=NEG.NPST
'S/he has no right to be pleased.'

(3.7) Irabu (< Miyako < Ryukyuan < Japonic); Irabu (Japan); SOV (Shimoji, 2020)
kai=ga=du *sac=n* *idi-r* *kutu*
3SG=NOM=FOC first=DAT go_out-ADN.NPST thing
"S/he should go first"'

(3.8) Korean (< Koreanic); Korean Peninsula; SOV (Kim, 2020)
chinkwu=ka ilpon=ey ka-l yeyceng=i-ta
friend=NOM Japan=DAT go-ADN.PROS plan=COP-DECL
"(My) friend is going to Japan."'

(3.9) Amdo Tibetan (< Tibeto-Burman < Sino-Tibetan); Amdo (China); SOV (Ebihara, 2020)
arʄa=kə nor ptsoŋ-ʄu ntɕʰarʑə rɛ
father=ERG yak sell.IPFV-NMLZ.GEN plan COP.B
"(My) father plans to sell yaks."'

(3.10) Burmese (< Tibeto-Burman < Sino-Tibetan); Myanmar; SOV (Kato, 2020)
ʈù dì híN=gò sá=dɛ̂ póuN(=bɛ́)
3SG this curry=KO eat=ADN shape(=EMP)
"It seems that s/he ate this curry"'

(3.11) Kurukh (< North Dravidian < Dravidian); East India; SOV (Kobayashi & Tsunoda, 2020)
aːs-hiː tambaː-s-in ilc-kaː caɖɖeː rahc-a:
3SG.M-GEN own.father-M-ACC fear.PS-PST.VADJ necessity COP.PS-PST.3SG.NM
"It was because he was scared of his father."'

(3.12) Sidaama (< Cushitic < Afroasiatic); Sidaama Region (Ethiopia); SOV (Kawachi, 2020)
íse faraš6 guluf-f-annó gara-a=ti
3SG.F.NOM horse.ACCOBL ride-3SG.F-IPFV.3 manner-LV=NPC.PRED.MOD
"She seems to ride a horse (habitually)"'

(3.13) Tagalog (< Malayo-Polynesian < Austronesian); Philippines; VSO (Katagiri, 2020)
plano ni=Noy na b‹um›isita sa=Davao bukas
plan GEN=Noy LK visit‹AF:INF› OBL=Davao tomorrow
"Noy plans to go to Davao."

58

(3.14) Bengali (< Indo-Iranian < Indo-European); Bangladesh , West Bengal (India); SOV
   (Huziwara, 2018)
   tar   ṭokio=te   jawar   kɔtʰa
   3SG.GEN Tokyo=LOC go.NMLZ.GEN word
   "It is planned that he will go to Tokyo.'"

Nevertheless, it turns out that the following three languages (3.15)-(3.17), unreported in Tsunoda (2020a), also have similar constructions. None of these languages is a language of East Asia, but they share in common the feature of the Mermaid Constructions where the non-finite clause is subordinate to the nominal predicate. Furthermore, if we presuppose that these examples, particularly of Russian and Welsh, are Mermaid Constructions, they are counterexamples against Tsunoda's (2020a) assumption that the verb and the noun are adjacent in the word order to compose a compound predicate.

(3.15) Tatar (< Kipchak < Turkic); Tatarstan (Russia); SOV
   siŋa   joqla-rɣa röxsät
   2SG.DAT sleep-INF permission
   "You are allowed to sleep.'"

(3.16) Russian (< Slavic < Indo-European); Russia; SVO
   pora nam   pogovorit'
   time   1PL.DAT PFV:talk:INF
   "It is time for us to talk.'"

(3.17) Welsh (< Celtic < Indo-European); Wales (United Kingdom); VSO
   rhaid   i   fi   godi'n   gynnar
   rhaid   i   fi   godi   yn gynnar
   necessity to 1SG wake_up in   early
   "I need to wake up early'"

Therefore, though it is true that languages found to have these constructions so far seem to center around Asia, these kinds of constructions are not necessarily an areal phenomenon, but can be found in languages that are unlikely to have had language contact.

3.1.2 Marginal examples of the Mermaid Construction

The Mermaid Constructions we have seen have a morphologically full noun in the predicate. However, Tsunoda (2020a) also includes related examples that do not have a morphologically full noun but involves a clitic or an affix grammaticalized from a noun or a nominalized form of a verb. Though we do not delve into these types of constructions in this thesis, it generally remains of interest whether it is part of the Mermaid Constructions.

Case 1: The noun is grammaticalized into a clitic or an affix

The examples below show the grammaticalization process of a full Mermaid Construction noun in Early Middle Japanese *sama* (3.18), through a reportative evidential clitic in Modern Japanese =*soo*

(3.19), to an inferential evidential suffix in Modern Japanese *-soo* (3.20). Although the latter two examples are not canonical types of the Mermaid Constructions in the sense that they do not have an independent Mermaid Construction noun, they can be treated as continuous phenomena to the Mermaid Constructions (see Section 3.5 for more details).

(3.18) Early Middle Japanese (Tsunoda 2020a; originally from *The Tale of the Bamboo Cutter* (9–10th century AD)
    tune=yori=mo   mono-omop-i-tar-u   sama=nar-i
    usual=ABL=ETOP thing-think-LINK-STAT-NPST.ADN appearance=COP-NPST.DECL
    '(Princess Kaguya) seems/appears to be in deeper thought than usual.'

(3.19) Japanese (Tsunoda, 2020a)
    Hanako=ga   Ezinbara=ni   ik-u=soo=da
    Hanako=NOM Edinburgh=DAT go-NPST.ADN=REPORT=COP.NPST.DECL
    '(I heard that) Hanako will go to Edinburgh.'

(3.20) Japanese (Tsunoda, 2020a)
    Hanako=ga   Ezinbara=ni   ik-i-soo=da
    Hanako=NOM Edinburgh=DAT go-LINK-INF-INFER=COP
    'It seems/appears that Hanako will go to Edinburgh.'

Case 2: The predicate is a nominalized verb

Some languages mentioned in Tsunoda (2020a) also include nominalized verbal predicate as part of the Mermaid Constructions (3.21)–(3.23). These do not have any morphologically independent noun, but the verb is nominalized and adds to the sentence some grammatical meaning.

(3.21) Japanese
    *Hanako=ga   kono   karee=wo   taber-u=**no**=da*
    Hanako=NOM this.ADN curry=ACC eat-NPST.ADN=NMLZ=COP.NPST.DECL
    '(It is the case that) Hanako eats this curry.'

(3.22) Burmese (Kato, 2020)
    *t̪ù   dì   hín̪=gò   sá=**dà**̀*
    3SG this curry=KO eat=NMLZ.RLS
    '(It is the case that) s/he eats this curry.'

(3.23) Hindi (< Indo-Iranian < Indo-European); India; SOV (Yasunari, 2020)
    *mãĩ   yuunivarsiṭii jaa-ne=**vaalaa**   hũ̃ũ*
    1SG university   go-INF.OBL=NMLZ.M.SG COP.PRS.1SG
    'I am about to go to the university.'

Case 3: The predicate is an adjectivalized noun

This marginal type of the Mermaid Constructions does not have a full noun but a noun adjectivalized by a proprietive suffix. The proprietive suffix N-PROP means "having N", and this type of construction is reported in Mongolian (3.24) and Sakha (3.25), for example. Though they are treated as

quasi-Mermaid Constructions in Tsunoda (2020a), I do not consider them so, because they have an adjectival predicate but not a nominal predicate, which is characteristic to the canonical Mermaid Constructions. Also, they do not seem to exhibit a similar kind of the peculiar behavior both in syntax and semantics. They are just an adjectival predicate derived from a noun, and their semantics also seems to be normally compositional, as can be confirmed in the literal translation.

(3.24) Mongolian (< Mongolic); Mongolia; SOV (Umetani, 2020)
    bi   *margaaš*   *xödöö*   *jav-a-x*   *tölövlögöö-tej* *baj-na*
    1SG.NOM tomorrow countryside go-EP-VN.NPST plan-PROP   be-TV.NPST
    'I am planning to go to the countryside tomorrow. (*lit.* I am with the plan to go to the countryside tomorrow.)'

(3.25) Sakha (< Turkic); Yakutia (Russia); SOV (Ebata, 2020)
    min   *tokio-ʁa*   *bar-ar*   *bɯlaan-naax-pɯn*
    1SG.NOM Tokyo-DAT go-PTCP.PRS plan-PROP-COP.1SG
    'I plan to go to Tokyo. (*lit.* I am with the plan to go to Tokyo.)'

3.1.3   Tatar nominal control predicates as Mermaid Constructions

We have covered the descriptive overview of the Mermaid Constructions. When we turn back to the control constructions with a nominal predicate in Tatar that we saw in Chapter 2, we notice the structural similarity between the two constructions. Namely, both constructions have the following form:

(3.26) (Subject or AGENT) + [Non-finite Clause] + Noun (+ Copula)

In addition, it is also similar that both constructions contribute to some grammatical meaning, in particular modality. We saw in Section 2.3.2 that the meaning of nominal control constructions in Tatar is modality (*röxsät*, *waqït*, and *mäslixät*) and intentionality (*isäp* and *nijät*). Thus, we claim that the nominal control constructions in Tatar are part of the Mermaid Constructions. From this section on, we continue our discussion assuming that this claim is plausible.

3.1.4   Tsunoda's Monoclausal Analysis of the Mermaid Constructions

Tsunoda (2020a) analyzes the Mermaid Constructions as having a monoclausal structure. In the construction, the predicate in the rightmost or leftmost of the (pseudo-)Clause, the Mermaid Construction noun, and optionally a copula constitute a compound predicate. (3.27) illustrates a general form of the Mermaid Constructions in this monoclausalist view, where the underlined components are a compound predicate.

(3.27) Subject ... <u>Pred N (Copula)</u>

Several authors in Tsunoda (2020a) presents evidence for the monoclausality of the construction (Kawachi, 2020; Kim, 2020; Sasaki, 2020; Tsunoda, 2020b). For example, for the Mermaid Constructions of Japanese, Tsunoda (2020a) demonstrates that both Pred + N and N + Copula form an

inseparable unit, and therefore Pred, N, and Copula together form a single compound predicate. For the former unit, it is confirmed that Pred and N cannot be intervened by any element, such as an adnominal adjective (compare (3.28) and (3.29)). As for the latter unit, it is impossible to insert an independent lexical phrase (e.g., AP or NP) between N and Copula.

(3.28) Japanese
    tsuyo-i   *ki*
    strong-NPST.ADN   feeling
    'strong feeling'

(3.29) Japanese
    Hanako=ga   *Ezinbara=ni*   *ik-u*   (**tsuyo-i)*   *ki=da*
    Hanako=NOM Edinburgh=DAT go-NPST.ADN strong-NPST.ADN feeling=COP.NPST.DECL
    '(intended:) Hanako (strongly) intends to go to Edinburgh.'

3.2 Mermaid Constructions as Control and Raising

We saw in Chapter 2 that the five nouns *röxsät*, *waqït*, *isäp*, *nijät*, *mäslixät* can be control predicates with an infinitival clause. Next, the previous section introduced the Mermaid Constructions (Tsunoda, 2020a)) and positioned those nominal control predicates as part of the Mermaid Constructions. Then, can other canonical Mermaid Constructions be treated as control? The following subsections argue that, while some of the Mermaid Constructions are control, others are raising.

3.2.1 Mermaid Constructions as control

This subsection will see if the treatment of the Mermaid Constructions as control that we saw in Tatar also holds true for other Mermaid Constructions. Since it is realistically difficult to analyze all of the Mermaid Constructions we saw in Section 3.1, we chiefly puts a focus on the Mermaid Constructions in Japanese, because Tsunoda (2020a) mentions that it is the language with the largest number of reported Mermaid Construction predicates.

    Recall the Mermaid Construction noun *ki* mentioned in the previous section (3.2). It is a morphologically independent noun that means "feeling", but when it is used in the Mermaid Construction, it means that the subject is intending to do something. We assume that *ki* is a nominal control predicate; namely, the embedded clause has a phonetically null syntactic argument PRO anaphorically bound by an argument in the main clause (3.30). In order to check if this is also a kind of control, we apply the classic diagnostics for distinguishing control and raising: the passivization test and the idiom chunk test.

(3.30) Japanese
    Hanako$_i$=ga   [PRO$_{i/*j}$ *Taroo=o*   *tatak-u*]   *ki=da*
    Hanako=NOM PRO   Taro=ACC hit-ADN.NPST feeling=COP.NPST.DECL
    'Hanako intends to hit Taro. (*lit.* Hanako is the feeling to hit Taro.)'

Passivization test for *ki*

The diagnostics of the passivization test are that a predicate is triggering raising if passivizing the embedded verb does not change the semantic role relation of the (main) subject and the (embedded) object, otherwise it is control. Sentence (3.31) is a passivized counterpart of (3.30). The subject *Hanako* is demoted to an oblique argument in the embedded clause, and *Taro* is promoted to the subject of the main clause. As is evident by the comparison of their free translation, the semantic role relation changes after the passivization. While (3.30) has *Hanako* as both the hitter (the AGENT of *hit*) and the intender (the AGENT of *intend*), (3.31) has *Hanako* as just the hitter, and it is now *Taro* who is the intender (see Table 3.1 for a visual comparison). Therefore, the Japanese Mermaid Construction noun *ki* has a control-like syntactic behavior.

(3.31) Japanese
    Taroo=ga   *Hanako=ni*   *tatak-are-ru*   *ki=da*
    Taro=NOM   Hanako=DAT   hit-PASS-ADN.NPST   feeling=COP.NPST.DECL
    'Taro intends to be hit by Hanako. (*lit.* Taro is the feeling to be hit by Hanako.)'

Voice	*hit*(AGENT, PATIENT)		INTEND(AGENT)
	AGENT	PATIENT	AGENT
Active (3.30)	*Hanako*	*Taro*	*Hanako*
Passive (3.31	*Hanako*	*Taro*	*Taro*

Table 3.1: A comparison of the thematic relation in (3.30) and (3.31).

Idiom chunk test for *ki*

An idiom is an expression with a figurative meaning that is not compositional. As we saw in the beginning of Chapter 2, the idiom chunk test judges that a predicate is raising if changing the copular element to the predicate does not lose the non-literal idiomatic meaning. We apply this test to the Japanese Mermaid Construction noun *ki*. Sentence (3.32) is a Japanese idiom that literally means "the legs become sticks" but idiomatically mean "one gets exhausted". As expected, the Mermaid Construction with *ki* fails this test (3.33).

(3.32) Japanese
    asi=ga   *boo=ni*   *nar-u*
    leg=NOM   stick=DAT   become-NPST.DECL
    '(One) gets exhausted by some leg-related physical activity (*lit.* the legs become sticks)'

(3.33) Japanese
    asi=ga   *boo=ni*   *nar-u*   *ki=da*
    leg=NOM   stick=DAT   become-NPST.ADN   feeling=COP.NPST.DECL
    'The legs intend to become sticks / *One intends to get exhausted'

    Other Japanese Mermaid Construction predicates that behave similarly to control include *tumori* "intention" and *ikoo* "intention".

3.2.2 Mermaid Constructions as raising

Mermaid Constructions we have analyzed in Japanese and Tatar have the traits of the control construction. Then, can all the Mermaid Constructions be reanalyzed as control? As the following data show, the answer is no. Some Mermaid Construction nouns such as *yotei* in Japanese pass both the passivization test (3.35) and the idiom chunk test (3.36).

(3.34) Japanese
 Hanako=ga Taroo=o tatak-u yotei=da
 Hanako=NOM Taro=ACC hit=NPST.ADN plan(N)=COP.NPST.DECL
 'Hanako is going to hit Taro.'

(3.35) Japanese
 Taroo=ga Hanako=ni tatak-are-ru yotei=da
 Taro=NOM Hanako=DAT hit-PASS-NPST.ADN plan=COP.NPST.ADN
 'Taro is going to be hit by Hanako.'

(3.36) Japanese
 asi=ga boo=ni nar-u yotei=da
 leg=NOM stick=DAT become-NPST.ADN plan=COP.NPST.DECL
 'One is going to be exhausted (by some leg-related physical activity)'

The data show that Mermaid Construction predicates like *yotei* behave similarly to raising predicates. In other words, *yotei* does not directly subcategorize for any semantic argument, and subordination with *yotei* does not affect the thematic relation of the arguments. It seems that most Mermaid Construction nouns in Japanese are of the raising type, for example *tokoro* "place" (progressive aspect), *mono* "thing" (habitual aspect), *kimari* "regulation" (deontic modality), moyoo "appearance, pattern" (epistemic modality), *katati* "shape" (epistemic modality), etc.

The claim that the Mermaid Constructions are in fact either control or raising is an updated view from Taguchi (2021), which argues that they are chiefly raising based on the Japanese data. A further investigation of the nature of the Mermaid Constructions in this thesis has revealed that some of them do have an agentive semantic argument like a control verb.

3.3 Counterarguments to the Monoclausal Hypothesis

If we take a position that the Mermaid Constructions are control and raising constructions with a noun, it almost automatically follows that the Mermaid Constructions are biclausal, given the assumption that an infinitival clause is headed by T.[2] This is an opposing argument to the monoclausal hypothesis (3.37) proposed in Tsunoda (2020a).

---

[2]Again, our argument at this point excludes smaller infinitival phrases (vP and VP) known as restructuring (Wurmbrand, 2001). However, we are not to abandon the possibility of a restructuring analysis for the Mermaid Constructions. This topic is left for future work.

(3.37) MONOCLAUSAL HYPOTHESIS OF THE MERMAID CONSTRUCTIONS:

Mermaid Constructions have a monoclausal structure where a verb, a Mermaid Construction noun, and optionally a copula constitute one compound predicate. That is, the Mermaid Constructions do not have two predicates but one.

Sticking with our analyses so far, we argue against this monoclausal hypothesis, and propose the biclausal analysis of the Mermaid Constructions (3.3).

(3.38) BICLAUSAL HYPOTHESIS OF THE MERMAID CONSTRUCTIONS:

Mermaid Constructions have a biclausal structure where the Mermaid Construction noun, together with a copula in some languages, constitute the main predicate, and the non-finite clause is the subordinate clause.

3.3.1 Alternative explanation with the Biclausal Hypothesis

In the Japanese Mermaid Constructions, Tsunoda (2020a) argues for the monoclausal hypothesis based mainly on the evidence in Japanese syntax. There are several points that he makes as the evidence for their monoclausality.

NOM/GEN alternation

First, in Japanese, it is possible to alternate the nominative suffix =*ga* and the genitive suffix =*no* in an adnominal clause (3.39). However, this is not the case in the Mermaid Constructions; the genitive subject is not allowed (3.40). From this fact, Tsunoda (2020a) argues that the bracketed phrase in (3.40) is not an adnominal clause.

(3.39) Japanese

pro [*Hanako*{=ga/=no} *sakana=o yak-u*] *nioi=da*
pro Hanako=NOM/GEN fish=ACC grill-NPST.ADN smell=COP.NPST.DECL
'(It is) the smell that (comes from where) Hanako grills the fish.'

(3.40) Japanese

[*Hanako*{=ga/*=no} *sakana=o yak-u*] {*ki/yotei*}=*da*
Hanako=NOM/GEN fish=ACC grill-NPST.ADN feeling/plan=COP.NPST.DECL
'Hanako {intends to/is going to} grill the fish.'

Second, Tsunoda (2020a) points out that the Mermaid Constructions are monoclausal because it is possible to put a focus on the subject with clefting in the Japanese Mermaid Constructions (3.42), while the clefting focus is not possible with an adnominal clause (3.41).

(3.41) Japanese

[*$*t_i$ sakana=o yak-u*] *nioi=na=no=wa*
fish=ACC grill-NPST.ADN smell=COP.NPST.ADN=NMLZ=TOP
Hanako$_i$=da
Hanako=COP.NPST.DECL

(3.42) Japanese
[*t_i* *sakana=o* *yak-u*] {*ki/yotei*}=*na*=*no*=*wa*
    fish=ACC grill-NPST.ADN feeling/plan=COP.NPST.ADN=NMLZ=TOP
Hanako_i=da
Hanako=COP.NPST.DECL
'It is Hanako who {intends to/is going to} grill the fish.'

Last but not least, Tsunoda (2020a) argues that the Mermaid Constructions are monoclausal because they can have only one subject, while an adnominal clause can have a subject for each predicate. In (3.43), another nominative argument *kore=ga* can stand as the subject of the main predicate. In (3.44), in contrast, if we add the same nominative argument, the Mermaid Construction reading is lost, and it is interpreted only in the literal meaning.

(3.43) Japanese
kore=ga [*Hanako*{=*ga*/=*no*} *sakana=o* *yak-u*] *nioi=da*
this=NOM Hanako=NOM/GEN fish=ACC grill-NPST.ADN smell=COP.NPST.DECL
'This is the smell that (comes from where) Hanako grills the fish.'

(3.44) Japanese
kore=ga [*Hanako*{=*ga*/*=*no*} *sakana=o* *yak-u*] {*ki/yotei*}=*da*
this=NOM Hanako=NOM/GEN fish=ACC grill-NPST.ADN feeling/plan=COP.NPST.DECL
'*Hanako {intends to/is going to} grill the fish. /
This is the feeling/plan for Hanako to grill the fish.'

From these points, Tsunoda (2020a) concludes that the Mermaid Constructions in Japanese do not involve an adnominal clause, but has a monoclausal structure with a compound predicate of the adjacent verb and noun. Nevertheless, these points can also be explained by the biclausal analysis. To begin with, these mentioned differences only demonstrate that the subject belongs to the main clause, and they are not direct evidence for the monoclausality. In the biclausal hypothesis, the overt subject is in the main clause, either as a controller or as a raised subject. Therefore, the restriction of the =*ga*/=*no* alternation is a natural consequence (cf. (3.45)). In addition, even if the sentence contains an adnominal clause, it is possible to make a cleft sentence if the dislocated argument is the main subject (3.46). The impossibility of having two subjects in (3.44) is just because the subject of the adnominal clause is either a covert PRO or a dislocated trace.

(3.45) Japanese
Hanako_i{=*ga*/*=*no*} [{PRO_i/*t_i*} *sakana=o* *yak-u*] {*ki/yotei*}=*da*
Hanako=NOM/GEN fish=ACC grill-NPST.ADN feeling/plan=COP.NPST.DECL
'Hanako {intends to/is going to} grill the fish.'

(3.46) Japanese
[*t_i* [*Hanako=ga* *sakana=o* *yak-u*] *nioi=na*]=*no*=*wa*
    Hanako=NOM fish=ACC grill-NPST.ADN smell=COP.NPST.ADN=NMLZ=TOP
kore=da
this=COP.NPST.DECL
'It is this that is the smell which (comes from where) Hanako grills the fish.'

The evidence that the subject of the adnominal clause is either a PRO or raised can be confirmed by the optional honorification agreement, which is known as a subjecthood test for Japanese (Matsumoto, n.d.).[3] Sentences (3.47)–(3.49) are all grammatical. That is, the optional honorification agreement can be morphologically expressed on either the embedded verb, the (main predicate) noun, or both. Therefore, if we assume that this construction has a compound predicate, it is not clear how two components of a predicate can agree with the subject independently. It is more plausible that the predicates work morphosyntactically independently rather than compose a compound predicate and that they share the same subject.

(3.47) Japanese
$Hanako_i$=*sama*=*ga* [{t_i/PRO$_i$} *Ezinbara*=*ni*   *irassyar-u*]
Hanako=HON=NOM   Edinburgh=DAT   come.HON-NPST.ADN
{*yotei/tumori*}=*da*
plan/intention=COP.NPST.DECL
'Hanako {is going to/intends to} come to Edinburgh (the speaker is showing a respect to Hanako)'

(3.48) Japanese
$Hanako_i$=*sama*=*ga* [{t_i/PRO$_i$} *Ezinbara*=*ni*   *kuru*]
Hanako=HON=NOM   Edinburgh=DAT   come.NPST.ADN
{*go-yotei/o-tumori*}=*da*
HON-plan/HON-intention=COP.NPST.DECL
'Hanako {is going to/intends to} come to Edinburgh (the speaker is showing a respect to Hanako)'

(3.49) Japanese
$Hanako_i$=*sama*=*ga* [{t_i/PRO$_i$} *Ezinbara*=*ni*   *irassyar-u*]
Hanako=HON=NOM   Edinburgh=DAT   come.HON-NPST.ADN
{*go-yotei/o-tumori*}=*da*
HON-plan/HON-intention=COP.NPST.DECL
'Hanako {is going to/intends to} come to Edinburgh (the speaker is showing a respect to Hanako)'

3.3.2 Other problems in the Monoclausal Hypothesis

The Monoclausal Hypothesis by Tsunoda (2020a) contains other problems than the compound predicate analysis. One issue is the strong assumption that the Clause is finite. As is admitted in Tsunoda (2020a), there are already a number of exceptions to this assumption. For example, when the embedded predicate is an adjectival noun predicate, which is a subclass of adjectival predicates, it appears in the adnominal form, which is distinct from the finite declarative form (cf. (3.50) and (3.51)).

(3.50) Japanese
Hanako=*ga*   *genki*=*da*
Hanako=NOM   fine=COP.NPST.DECL
'Hanako is fine.'

---

[3]I am grateful to Chen Xie (p.c.) for suggesting this diagnosis.

(3.51) Japanese

    Hanako=ga   *genki=na*   *yotei=da*

    Hanako=NOM fine=COP.NPST.**ADN** plan=COP.NPST.DECL

    'Hanako is going to be fine.'

In addition, we have already confirmed in Section 3.1 that most of the Mermaid Constructions crosslinguistically take a non-finite clause, though there are typological differences in what non-finite form they take, as illustrated in Table 3.2. The assumption that what seems to be the embedded clause is in fact a finite form might be a Japanese-centric confusion coming from the fact that the adnominal and declarative forms in Modern Standard Japanese are morphologically the same in the verbal paradigm.

Form	Languages
Adnominal	Japanese?, Mitsukaido Japanese?, Irabu Ryukyuan, Korean, Burmese, Sidaama?
Nominalized	Amdo Tibetan, Bengali
Infinitival	Tagalog, Tatar, Russian, Welsh
Adjectivalized	Kurukh

Table 3.2: A comparison of the forms of embedded predicates in the Mermaid Constructions. Languages with a question mark '?' mean that the form is by and large ambiguous with their finite (declarative) form.

3.4 Syntax–Semantics Interface of the Mermaid Constructions

Previous sections have syntactically analyzed the Mermaid Constructions as having a biclausal structure similar to the raising and control constructions. In turn, this section focuses on the semantics of the Mermaid Constructions. Specifically, employing a formal approach to semantics, it will demonstrate that the Mermaid Constructions can be treated as part of raising and control at least based on the data in Japanese.[4] Upon the proofs with semantic derivations, we utilize Glue semantics, a framework for the analysis of semantic composition at the syntax-semantics interface (Asudeh, 2022; Dalrymple et al., 2019). Glue has chiefly been developed in Lexical Functional Grammar (LFG), but it is designed to be valid for any syntactic theories that assume syntactic heads. Though we do not dive into the details of the introduction to Glue, readers with basic knowledge of logic and formal semantics should be able to follow the arguments henceforth, as LFG-specific theoretical devices such as f-structure will not be crucial to the points in this section. For detailed explanations and definitions, see the aforementioned literature.

3.4.1 Semantic scopes of raising and control

Raising and control are known to exhibit a difference in the semantic scope of an existential quantifier (Dalrymple et al., 2019). Compare the sentences (3.52) and (3.53) below.

---

[4]We focus on the data in Japanese, not Tatar, because no raising Mermaid Construction has been found in Tatar.

(3.52) Someone seemed to yawn. (*raising*)

(3.53) Someone tried to yawn. (*control*)

In (3.52), there are two possible readings. One reading is to assume that there is a specific person and he or she seemed to yawn. In this case, the existence of the person is a prerequisite, but it is possible that they did not yawn in fact. The other reading is that it seemed to be the case a person yawned. This reading does not presuppose a specific person, and there could actually be nobody in the situation. Let us call the former reading **wide** scope interpretation because the existential scope spreads to the whole, and the latter reading **narrow** scope interpretation because the existential scope is under the scope of the evidential modality. In the formal semantic notation, this scopal difference can be represented as in (3.54) and (3.55), respectively.

(3.54) Wide scope interpretation: $\mathsf{exist}(x, \mathsf{person}(x), \mathsf{seem}(\mathsf{yawn}(x)))$

(3.55) Narrow scope interpretation: $\mathsf{seem}(\mathsf{exist}(x, \mathsf{person}(x), \mathsf{yawn}(x)))$

In the case of control (3.53), on the other hand, only the wide scope interpretation of the existential quantifier is possible (3.56). That is, the sentence only allows the wide scope interpretation that presupposes the existence of some person, who attempted to yawn. The narrow scope interpretation (3.57), where the yawning attempt was executed impersonally, is impossible in English.[5]

(3.56) Wide scope interpretation: $\mathsf{exist}(x, \mathsf{person}(x), \mathsf{try}(x, \mathsf{yawn}(x)))$

(3.57) Narrow scope interpretation: $*\mathsf{try}(\mathsf{exist}(x, \mathsf{person}(x), \mathsf{yawn}(x)))$

Then, how can this scopal difference be derived in denotational semantics? Suppose that we have the lexicon $\mathcal{L} = \{\mathsf{yawn}, \mathsf{someone}, \mathsf{seem}, \mathsf{try}\}$ at hand. For the sake of facilitating the discussion, we borrow the meaning constructor of each semantic constituent in $\mathcal{L}$ used in Dalrymple (2019). A meaning constructor is a terminology in Glue that consists of a pair of two logical formula separated by a colon; the left hand side of a meaning constructor is a semantic denotation $\mathcal{M}$ in the form of lambda calculus, and the right hand side $\mathcal{G}$ instructs how variables should be consumed to derive another form based on a syntactic parse (Asudeh, 2022), which works in a similar manner to semantic types. The linear implication connective is represented by $x \multimap y$ to mean "to consume x to derive y", and by and large it works in the same manner as $\rightarrow$ in linear logic. The list of the lexicon $\mathcal{L}$ and their meaning constructors is summarized in Table 3.3.[6]

Now, for wide and narrow scope interpretations, we think of two simplified syntactic-semantic trees that ignore functional information such as tense and semantic roles for convenience, as illustrated in Figures 3.1 and 3.2. Note that the linear order of constituents do not reflect the actual word order, because we do not assume a movement here.

---

[5]However, it is certainly not the case in the restructuring infinitival construction in German (Wurmbrand, 2001).

[6]The functional variables e, t, and so on, are not the same as e (entity) t (truth-value) of semantic types.

lexicon $\mathcal{L}$	meaning constructors $\mathcal{M} : \mathcal{G}$
yawn	$\lambda x.\mathsf{yawn}(x) : e \multimap y$
someone	$\lambda S.\mathsf{exist}(x, \mathsf{person}(x), S(x)) : \forall H.(e \multimap H) \multimap H$
seem	$\lambda P.\mathsf{seem}(P) : y \multimap s$
try	$\lambda P \lambda x.\mathsf{try}(x, P) : (e \multimap y) \multimap (e \multimap t)$

Table 3.3: The semantic denotation of the constituents.
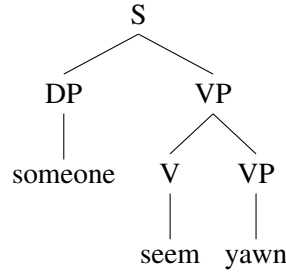
Figure 3.1: A simplified tree for the wide scope interpretation.

For the wide scope interpretation, we first apply *yawn* to *seem*. In doing so, we need to assume e as the conditional proof assumption (CPA), by which we will obtain $\mathsf{yawn}(x) : y$ (cf. Table 3.3) by the implication elimination. Then, we feed this to *seem* and get $\mathsf{seem}(\mathsf{yawn}(x)) : s$. Since e is an assumption, we get $\lambda x.\mathsf{seem}(\mathsf{yawn}(x)) : e \multimap s$ for ⟦**seem-yawn**⟧. Next, we feed this to *someone*. The meaning constructor of *someone* instructs that it consumes $e \multimap H$, where H is any argument, and returns the same H. Therefore, if we feed $e \multimap s$, we will get s. These steps are summarized in (3.58) below.

(3.58) a.   ⟦**seem-yawn**⟧ $= \lambda x.\mathsf{seem}(\mathsf{yawn}(x)) : e \multimap s$ (CPA)

b.   ⟦**someone-seem-yawn**⟧ $= \mathsf{exist}(x, \mathsf{person}(x), \mathsf{seem}(\mathsf{yawn}(x))) : s$

The narrow reading is derived from the structure in Figure 3.2. We first apply *yawn* to *someone*, from which we get a truth-conditional form $\mathsf{exist}(x, \mathsf{person}(x), \mathsf{yawn}(x))$ by the same rule above. Then, we apply this to *seem*, and obtain the narrow scope interpretation $\mathsf{seem}(\mathsf{exist}(x, \mathsf{person}(x), \mathsf{yawn}(x)))$. These steps are summarized in (3.59).

(3.59) a.   ⟦**someone-yawn**⟧ $= \mathsf{exist}(x, \mathsf{person}(x), \mathsf{yawn}(x)) : y$

b.   ⟦**seem-someone-yawn**⟧ $= \mathsf{seem}(\mathsf{exist}(x, \mathsf{person}(x), \mathsf{yawn}(x))) : s$

These semantic compositions predict that the narrow scope interpretation with a control predicate (e.g., *try*) is impossible, because the instructions in the meaning constructors allows application of *yawn* to *try* but not of *someone yawns* to *try*.

```
                              S
                          ╱       ╲
                       VP           IP
                        │         ╱    ╲
                      seem      DP      VP
                                 │       │
                              someone  yawn
```

Figure 3.2: A simplified tree for the narrow scope interpretation.

3.4.2   Predicting the scopal (un)ambiguities of the Mermaid Constructions

Interestingly, scopal (un)ambiguities of the same kind are also found in the Mermaid Constructions in Japanese. Recall that in previous sections we have argued that *yotei* is a raising predicate and *ki* a control predicate in Japanese. If this assumption is correct, it must follow that *yotei* allows both wide and narrow scope interpretation and *ki* only allows wide scope interpretation. Indeed, this prediction is correct. In (3.60), there are two possible scopal readings, i.e., wide and narrow. If we take a wide scope for *yotei*, its semantic denotation will look like: $\mathsf{exist}(x, \mathsf{person}(x), \mathsf{planned}(\mathsf{go}(x)))$, where there is at least one specific person (say, Hanako, but the speaker does not recognize her) who is going to go. This is a valid interpretation in Japanese. In contrast, if we take a narrow scope for *yotei*, its semantic denotation will be: $\mathsf{planned}(\mathsf{exist}(x, \mathsf{person}(x), \mathsf{go}(x)))$, where at least one person is going but the speaker does not care who. This is also a valid interpretation.

(3.60) Japanese
    dareka=ga   *ik-u*   *yotei=da*
    someone=NOM   go-NPST.ADN   plan=COP.NPST.DECL
    'Someone is going to go. (*raising*)'

For the control predicate *ki(=da)* (3.61), on the other hand, the narrow scope interpretation is not accepted in Japanese. The interpretation is restricted to a situation where there is a specific person who intends to go. This observation is in line with the semantic behaviors of raising and control verbs. Evidently, these behaviors can be semantically derived by assuming the same structure of meaning constructors as raising and control. That is, the meaning constructors for *yotei* (a raising noun) and *ki* (a control noun) should look like Table 3.4.

(3.61) Japanese
    dareka=ga   *ik-u*   *ki=da*
    someone=NOM   go-NPST.ADN   feeling=COP.NPST.DECL
    'Someone intends to go. (*control*)'

71

lexicon $\mathcal{L}$	meaning constructors $\mathcal{M} : \mathcal{G}$
yotei	$\lambda P.\mathsf{planned}(P) : y \multimap s$
ki	$\lambda P \lambda x.\mathsf{intend}(x, P) : (e \multimap y) \multimap (e \multimap t)$

Table 3.4: Meaning constructors for *yotei* and *ki* in Japanese.
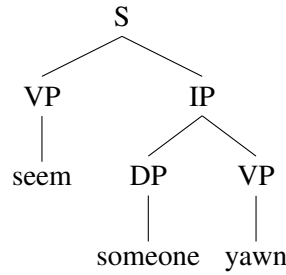
3.5 Diachronic Perspectives: Grammaticalization

The previous section has descriptively shown that the Mermaid Constructions have a syntactically similar structure to control and raising predicates. This section provides a diachronic analysis of the Mermaid Constructions with a focus on grammaticalization. We argue that, through the comparison of the degree of grammaticalization happening in Mermaid Constructions and control/raising predicates, the Mermaid Constructions have similar diachronic characteristics to the control and raising constructions. In particular, our arguments are based on the traits of grammaticalization formalized by Heine & Kuteva (2002) and Hopper (2003).

3.5.1 Four mechanisms of grammaticalization by Heine & Kuteva (2002)

Heine & Kuteva (2002) mention the following four points as the typical characteristics of grammaticalization phenomena.

1. Desemanticization

   A word loses its concrete meaning and is reinterpreted in specific contexts as more abstract, grammatical meanings.

2. Extension (context generalization)

   A word is used in new contexts with newly acquired grammatical meanings.

3. Decategorialization

   A word loses its original morphosyntactic properties characteristic of lexical or other less grammaticalized forms.

4. Erosion (phonetic reduction)

   A word loses its phonetic substance.

These four concepts are epitomized by the grammaticalization process of the future enclitic =*'ll* in Modern English. In Old English, *willan* was an independent verb meaning "to want". In the course of grammaticalization to =*'ll*, it has lost the concrete semantics of the action to wish something and gained an abstract function to mark the future tense (desemanticization). At the same time, *will(an)*, originally used to refer to an action or an attitude, is generalized to be also used in a new context to refer to an abstract future tense (extension). In terms of morphosyntactic categories, *will* used as a

future tense marker is no longer a verb, but is an auxiliary (decategorialization). Finally, in colloquial speech, *will* is phonetically abbreviated to =*'ll*, realizing as something like [ɫ] (erosion).

Based on these four points, this subsection compares the grammaticalization in the Mermaid Constructions and the control/raising constructions. We will cover the topics in the order of control, raising, and the Mermaid Construction. First, it is hard to say if verbal and adjectival control predicates have undergone any grammaticalization process. That is, it is not clear whether verbs like *try* and *want* have the infinitival complement as a result of grammaticalization from, say, a nominal argument, or they inherently subcategorize for it without grammaticalization.

Raising, on the other hand, seems to have undergone a weak degree of desemanticization and extension. For example, a verb *appear* used in the context of raising has lost its lexical meaning of "to emerge" and instead gained a grammatical meaning of some epistemic modality. Similarly, *tiješ* and *mömkin* as raising predicates in Tatar no longer have their concrete meaning of "obligated, indebted" and "possible", and instead have a function as deontic modality ("must") and epistemic modality ("it might be the case that ..."), respectively. Still, raising predicates do not exhibit decategorialization and erosion (excluding further grammaticalized forms such as auxiliaries). Raising predicates like *seem* in English and *tiješ* in Tatar retain their original lexical categories and morphosyntactically behave as a verb and an adjective, respectively.

In the Mermaid Constructions, it is certain that those nouns have been losing a concrete meaning as a full-fledged noun (desemanticization) and instead acquiring a new grammatical meaning in new contexts (extension). For example, *isäp* as a control predicate in Tatar no longer means "thought, counting" referring to some actual cognitive process, and marks the speaker's attitude (dynamic modality) to a situation described in the infinitival phrase. Yet, those nominal control predicates still retain morphosyntactic traits as nouns. It follows typical nominal declension morphology and syntactic rules, e.g., adnominal modification (no decategorialization). In addition, it can be seen from the data (3.2)–(3.17) that the nominal control predicates are not pronounced in a different (typically reducing) manner from the original noun form (no phonetic reduction). Thus, at least for nominal control predicates, it seems that they have undergone desemanticization and extension.

As a summary of the observation so far, Table 3.5 illustrates the comparison of control, raising, and the mermaid constructions in terms of the four mechanisms of grammaticalization proposed by Heine & Kuteva (2002).

	Control	Raising	Mermaid Constructions
Desemanticization	??	✓	✓
Extension	??	✓	✓
Decategorialization	–	–	–
Erosion	–	–	–

Table 3.5: A comparison of the grammaticalization in control, raising, and the Mermaid Constructions based on Heine & Kuteva's (2002) categorization. ✓ represents that a construction of the column has undergone the process of the row. ?? means that it is unknown whether the process has happened in the construction. Note that control and raising here do not include the Mermaid Constructions (i.e., nominal control predicates) for convenience.

3.5.2 The cline of grammaticalization by Hopper & Traugott (2003)

Hopper & Traugott (2003) showed that the grammaticalization processes have a unidirectional tendency shown in (3.62). That is, their theory assumes that (i) a content word is the least grammaticalized form; (ii) as grammaticalization proceeds, the semantics of the word shifts from a concrete content (e.g., a thing, a concept, an event, etc.) to a functional element (e.g., tense, aspect, modality, etc.); (iii) it then loses its lexical independence and realizes as a form attached to another host as a clitic; and, (iv) as the ultimate end of the grammaticalization, the attachment is no longer separable prosodically and morphosyntactically as an affix.

(3.62) Content word → grammatical word → clitic → affix (Hopper & Traugott, 2003)

This subsection compares control, raising, and the Mermaid Constructions by applying the "clines of grammaticality" shown above. To begin with, because control predicates retain their lexical concrete meaning, control predicates are at the least grammaticalized stage. On the other hand, raising predicates are more of grammatical words than content words. For instance, when a predicate like *begin* are used as a control verb, it semantically functions as a two-place predicate; if we take a proposition "John began to run" for example, its semantic denotation would be $begin_1(John, run(John))$. In contrast, when *begin* is used as a raising verb, its semantic function takes a proposition and returns a truth value. Namely, "John began to fall." would be $begin_2(fall(John))$, where *John* is not a direct semantic argument of *begin*. This usage of *begin* marks an inchoative aspect of the situation described by the infinitival phrase. In the same manner, Mermaid Construction nouns have lost their concrete semantic content and acquired a new function to mark an abstract grammatical feature. Table 3.6 is a summary of the discussion so far.

	Control	Raising	Mermaid Constructions
Content word	✓	✓	✓
Grammatical word	–	✓	✓
Clitic	–	–	–
Affix	–	–	–

Table 3.6: A comparison of the degree of the grammaticalization in control, raising, and the Mermaid Constructions based on the clines of grammaticality proposed by Hopper & Traugott (2003). Check marks in gray mean that the stage has passed.

3.5.3 Tatar nominal control constructions as grammaticalization

In Chapter 2, we analyzed the control constructions with a nominal predicate in Tatar in the Minimalist framework. During the discussion, we in fact assumed two points that are implicitly relevant to grammaticalization (3.63).

(3.63)   1. The *röxsät*-type construction, including *röxsät*, *waqït*, and *mäslixät*, have the dative argument as an internal argument, because it is a grammaticalized form that has lost its finite predicate that takes a dative GOAL.

2. The *isäp*-type construction, including *isäp* and *nijät*, have the genitive argument as an external argument. It is a consequence of grammaticalization by which it lost its finite predicate *bar/juq*.

The first assumption is that the *röxsät*-type construction is grammaticalized and has lost a dative-taking predicate. For example, the noun *röxsät* "permission" typically co-occurs with the verb *birü* "to give" that marks the recipient of the permission in dative. In the course of grammaticalization, this dative-taking predicate is lost, and the dative RECIPIENT/GOAL is reanalyzed as an internal argument of *röxsät*, because otherwise it does not have any host. The second assumption is in line with the discussion of little *n*P in the possessive-genitive construction in many languages. As the example in English "The enemy's destruction of the city" epitomizes, certain nouns seem to retain the thematic relation, which in turn is expressed in *'s*-possessive and *of*-genitive. Tatar does not use a *have*-verb to express possession, but marks the possessor in genitive and the possessed object in nominative, optionally agreeing with the possessor's person and number (3.64). We assume that the pre-grammaticalized form of the *isäp*-type construction has this form (3.65) with the existential predicate *bar* (*juq* in negative). Then, it lost the existential predicate, and the original DP is reanalyzed as a full TP while retaining its thematic relation that derives from *n*P (3.66).

(3.64) *minem   kitäb-em   bar*
     1SG.GEN book-POSS.1SG exist
     'I have a book (*lit*. There is my book).'

(3.65) [*minem   [kit-ärgä]*$_{VP}$ *isäb-em*]$_{DP}$ *bar*
     1SG.GEN leave-INF   thought-POSS.1SG exist
     'I have a thought (plan) to leave.'

(3.66) *minem   [kit-ärgä]*$_{VP}$ *isäb-em*
     1SG.GEN leave-INF   thought-POSS.1SG
     'I am going to leave.'

Evidence to assume an existential predicate as the pre-grammaticalized form is that the negated form of (3.66) employs *juq*, a negative form of *bar*, and not *tügel*, a negative particle placed after nominal and adjectival predicate (3.67).

(3.67) *minem   kit-ärgä   isäb-em   juq/*tügel*
     1SG.GEN leave-INF thought-POSS.1SG exist.NEG/*NEG
     'I am not going to leave.'

A similar grammaticalization process is also found in the Welsh Mermaid Construction. Though it takes the form of the Mermaid Construction in a present, non-interrogative, positive sentence (3.17), existential predicate (*oes*, *mae*, and their inflected forms) is necessary in other environments (3.68)–(3.69).

(3.68) Welsh
     oes   rhaid   i fi godi   nawr?
     exist.Q necessity to me wake_up now
     'Do I have to wake up now?'

(3.69) Welsh
     nac   oes, does   dim rhaid   i ti godi   nawr.
     NEG exist exist.NEG NEG necessity to you wake_up now
     'No, you don't have to wake up now.'

3.6   Summary of this chapter

This chapter has positioned the Tatar control constructions with a nominal predicate under the Mermaid Constructions that was outlined in Section 3.1. Then, Section 3.2 provided syntactic evidence for treating the Mermaid Constructions cross-linguistically as a subset of raising and control . By doing this, it is inevitable to contradict Tsunoda's (2020a) analysis that the Mermaid Constructions have a monoclausal structure because of our assumption that raising and control involve a subordinate CP/TP. Given this opposition, Sections 3.3–3.5 argued for the biclausal hypothesis from the aspects of syntax, semantics, and grammaticalization. All of these perspectives have supported that the Mermaid Constructions has a biclausal structure that is in parallel with the raising and control constructions.

Chapter 4

Conclusion

Before recapping the discussions in this thesis, let us remind ourselves of our initial research questions that the Tatar infinitival constructions raised.

1. What makes the difference in case-marking of the agent between the *tiješ*-type and *kiräk*-type infinitival constructions?

2. Why is there no finite predicate in the Dative–Infinitive construction? It seems to be non-compositional.

3. Why is the agent marked by dative in the Dative–Infinitive construction? What assigns the case?

4. The *röxsät*-type and *isäp*-type infinitival constructions also seem to lack a finite predicate. What is their structure like?

5. The AGENT is marked by dative in the *röxsät*-type construction, while it is in genitive in the *isäp*-type construction. What causes this difference?

6. Are these constructions monoclausal or biclausal?

After a general overview of the Tatar language in Chapter 1, in the course of our discussions through Chapters 2 and 3, we have found answers to these questions. Chapter 2 chiefly dealt with the issues related to Questions 1–5. Section 2.1 thoroughly outlined the Tatar infinitival constructions with a particular focus on non-verbal predicates. Based on this descriptive data, Section 2.2 argued that adjectival predicates *tiješ* and *mömkin* (with a nominative subject) are raising predicates. In contrast, Section 2.3 confirmed that other non-verbal predicates such as *kiräk* and *röxsät* are control predicates. The dative case-marking on the AGENT (or, to be precise, the initiator) of *kiräk* is assigned as an internal GOAL argument of the predicate. These points answer to our first question. Section 2.3 also pointed out that the five nominal predicates, *röxsät*, *waqït*, *mäslixät*, *isäp*, and *nijät* are in fact control predicates but of a noun. This is a cross-linguistically new finding, because most literature on control and raising analyzes verbal and adjectival predicates. We saw that the first three predicates take a dative AGENT (initiator), while the last two take a genitive AGENT (initiator). We provided an

77

explanation for this difference that the dative case-marking is a remnant of the marking on the internal GOAL argument and the genitive is a remnant of the marking on the POSSESSOR=AGENT argument in nP. In both cases, the original predicate responsible for these case-markings is lost, possibly as a result of grammaticalization (see also Section 3.5). This hypothesis answers to our fifth question. Then, Section 2.4 analyzes the Dative–Infinitive construction of Tatar as a subset of the control construction by assuming a phonetically null (adjectival) predicate that assigns a dative case in the same manner as the *kiräk*-type control construction. In doing so, the null predicate is analyzed to be responsible for the modal meaning, and the problem of non-compositionality is solved. These arguments answer to our second and third questions.

Chapter 3, particularly Section 3.1 showed that the Tatar control constructions with a nominal predicate is part of the Mermaid Constructions. Furthermore, Section 3.2 proposed that the Mermaid Constructions are a subset of the raising and control constructions. From this proposal, it follows that the Mermaid Constructions with a morphosyntactically full noun are biclausal, which is a contradicting argument against Tsunoda's (2020a) monoclausalist proposal. However, Section 3.3 pointed out the problems of the monoclausal analysis, and demonstrated the plausibility of the biclausal structure of the Mermaid Constructions. In support of this claim, Section 3.4 demonstrated from the aspect of the syntax-semantics interface that the scopal (un)ambiguity of the Mermaid Constructions with an existential quantifier can be correctly predicted in the same manner as the raising and control constructions. In addition, Section 3.5 confirmed that the Mermaid Constructions, control, and raising are closely related to grammaticalization, which has affected the case-marking of the Tatar nominal control constructions. These arguments answer our last question.

Thus, we have solved our initial questions regarding the raising and control constructions and the Mermaid Constructions in Tatar and other languages. However, it is true that we have consciously ignored several related aspects such as restructuring and marginal Mermaid Constructions for the sake of shaping our arguments straightforward. This thesis is not to declare that such points are irrelevant to the issues; rather, they are also points worth delving into. We shall end this thesis leaving these remaining tasks as future work.

Bibliography

Adger, D. (2003). *Core syntax: Minimalist approach*. Oxford University Press.

Al-Dobaian, A. S. (2006). A Minimalist analysis of three predicate types: Lexical, functional, and null. *Linguistic Research*, *11*, 22–52.

Asudeh, A. (2022). Glue semantics. *Annual Review of Linguistics*, *8*, 321–341.

Bailyn, J. F. (2004). The case of q. In O. Arnaudova, W. Browne, M. L. Rivero, & D. Stojanović (Eds.), *Formal approaches to slavic linguistics 12: The ottawa meeting* (pp. 1–36). Michigan Slavic Publications.

Bailyn, J. F. (2011). *The syntax of Russian*. Cambridge University Press.

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. The University of Chicago Press.

Boeckx, C., & Grohmann, K. K. (2004). Barriers and phases: Forward to the past?

Bošković, Ž. (2011). Last Resort with Move and Agree in derivations and representations. In C. Boeckx (Ed.), *The Oxford handbook of Linguistic Minimalism*. Oxford University Press.

Burbiel, G. (2018). *Tatar grammar: A grammar of the contemporary Tatar literary language*. Institute for Bible Translation.

Carnie, A. (2013). *Syntax: A generative introduction* (3rd ed.). Wiley-Blackwell.

Chomsky, N. (2001). Beyond explanatory adequacy. *MIT Occasional Papers in Linguistcis*, *20*, 1–28.

Chomsky, N. (2010). *Lectures on Government and Binding: The Pisa lectures*. De Gruyter Mouton.

Conklin, J. T. (2015). The interaction of palatal coarticulation and palatal harmony in Kazan Tatar. *Proceedings of Purdue Linguistics Association 10th Symposium*.

Dalrymple, M., Lowe, J. J., & Mycock, L. (2019). *The oxford reference guide to lexical functional grammar*. Oxford University Press.

Ebata, F. (2020). Sakha (yakut). In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 363–386). Mouton de Gruyter.

Ebihara, S. (2020). Amdo Tibetan. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 419–464). Mouton de Gruyter.

Farkas, D. F. (1988). On obligatory control. *Linguistics and Philosophy*, *11*, 27–58.

Fleisher, N. (2006). *Russian dative subjects, case, and control* (Master's thesis). University of California, Berkeley.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge University Press.

Henry, C. R. (2018). An optimality theoretic analysis of vowel harmony in Kazan Tatar. *92nd Annual Meeting of the Linguistic Society of America*.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd). Cambridge University Press.

Huziwara, K. (2018). A contrastive study of external adnominal clauses in Japanese and Bangla. *Bangabidya: International Journal of Bengal Studies*, *10*, 358–367.

Ibrahim, X. (2008). *Ялгыз тәкәрлек: роман, повестьлар, хикәяләр [Lonely lawping: Novel, stories, tales] (in Tatar)*. Татарстан Китап Нәшрияты.

Johanson, L. (2021). *Turkic*. Cambridge University Press.

Katagiri, M. (2020). Tagalog. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 781–816). Mouton de Gruyter.

Kato, A. (2020). Burmese. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 781–816). Mouton de Gruyter.

Kawachi, K. (2020). Sidaama. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 781–816). Mouton de Gruyter.

Kim, J. (2020). Korean. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 781–816). Mouton de Gruyter.

Kobayashi, M., & Tsunoda, T. (2020). Kurux. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 781–816). Mouton de Gruyter.

Kratzer, A. (1996). Severing the external argument from its verb (J. Rooryck & L. Zaring, Eds.). *33*.

Landau, I. (2013). *Control in generative grammar: A research companion*. Cambridge University Press.

Matsumoto, Y. (n.d.). Complex predicates in japanese: A syntactic and semantic study of the notion 'word'. *Studies in Japanese Linguistics series*.

Mihalicek, V. (2012). *Serbo-croatian word order: A logical approach* (Doctoral dissertation). Ohio State University.

Moore, J., & Perlmutter, D. M. (1999). Case, agreement, and temporal particles in Russian infinitival clauses. *Journal of Slavic Linguistics*, *7*(2), 219–246.

Poppe, N. (1961). *Tatar manual*. Indiana University Publication.

Ramchand, G. (2018). Alternating adjectives. *The Linguistic Review*, *35*(2), 283–306.

Republic of Tatarstan. (2013). 1-3PT от 12.01.2013 'Об использовании татарского языка как государственного языка Республики Татарстан' [On the use of the Tatar language as the state language of the Republic of Tatarstan].

Sakurama-Nakamura, M. (2018). Татарская диаспора и туган тел: ситуация татарского языка среди татар в странах Центральной Азии [Tatar diaspora and tugan tel: The situation of the Tatar language among Tatars in Central Asian countries] (in Russian). *Татар теле, әдәбияты, һәм тарихы: үткәне, бүгенгесе һәм киләчәге*, 23–38.

Sasaki, K. (2020). Mitsukaido dialect of Japanese. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 125–166). Mouton de Gruyter.

Saykhunov, M. R., Khusainov, R. R., Ibragimov, T. I., Luutonen, J., Salimzyanov, I. F., Shaydullina, G. Y., & Khusainova, A. M. (2021). Corpus of written Tatar.

Shimoji, M. (2020). Irabu Ryukyuan. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 781–816). Mouton de Gruyter.

Sigurðsson, H. A. (1992). The case of quirky subjects. *Working Papers in Scandinavian Syntax*, *49*.

Sigurðsson, H. A. (2002). To be an oblique subject: Russian vs. Icelandic. *Natural Language & Linguistic Theory*, *20*(4), 691–724.

Sigurðsson, H. A. (2003). Case: Abstract vs. morphological. In E. Brandner & H. Zinsmeister (Eds.), *New perspectives on case theory* (pp. 223–268). CSLI Publications.

Taguchi, C. (2021). Mermaid Construction in Tatar. *Proceedings of the 162nd Conference of the Linguistic Society of Japan*.

Timberlake, A. (2004). *A reference grammar of Russian*. Cambridge University Press.

Timerkhanov, A. A. (2019). *Tatar-English, English-Tatar dictionary*. G. Ibragimov Institute of Language, Literature; Art, Tatarstan Academy of Sciences.

Tsunoda, T. (2011). Ningyo koubun: Nihongogaku kara ippan gengogaku he no kouken. *NINJAL Research Papers*, *1*, 53–75.

Tsunoda, T. (2012). Ningyo koubun to meisi no bunpouka. *NINJAL Project Review*, *7*, 3–11.

Tsunoda, T. (2020a). *Mermaid Construction: A compound-predicate construction with biclausal appearance*. De Gruyter Mouton.

Tsunoda, T. (2020b). Modern standard japanese. *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 65–123). Mouton de Gruyter.

Umetani, H. (2020). Mongolian. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 387–417). Mouton de Gruyter.

Wertheim, S. (2002). Language "purity" and the de-Russification of Tatar. *UC Berkeley: Berkeley Program in Soviet and Post-Soviet Studies*.

Wiemer, B. (2017). Main clause infinitival predicates and their equivalents in Slavic: Why they are not instances of insubordination. In L. J. and U. Demske (Ed.), *Infinitives at the syntax-semantics interface: A diachronic perspective*. De Gruyter Mouton.

Wurmbrand, S. (2001). *Infinitives: Restructuring and clause structure*. De Gruyter Mouton.

Yasunari, I. (2020). Hindi. In T. Tsunoda (Ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance* (pp. 651–678). Mouton de Gruyter.

Zäkiev, M., & Xisamova, F. (2015). *Татар грамматикасы [Tatar grammar]*. Г. Ибраһимов исем. Тел, әдәбият һәм сәнгать институты.

Zaripov, Ä. (2013). Кадерле хәтирәләр. *Казан Утлары*, *1090*, 153–167.

Appendices

Appendix A

Conventions for the Transcription and Notations

The Latin transcription of Tatar used in this thesis is based on the Latin orthography of Tatar specified in the law of the Republic of Tatarstan 2013 with some modifications for the sake of consistency with the transcription used in Johanson (2021) that is compatible across Turkic languages. The detailed correspondence of the Cyrillic alphabet, the Latin alphabet, and the transcription used in this paper are illustrated in Table A.1 below. The rules of the Latin orthography are also summarised with examples in Timerkhanov (2019). In this thesis, transcription is italicised, phonological notation is enclosed by slashes /·/, and phonetic realization is enclosed by square brackets [·]. Suffixes and enclitics that have allophones by phonological alternations are enclosed by curly brackets {·}, e.g., {-GAn} includes *-qan*, *-kän*, *-ɣan*, and *-gän*. The notations for representing phonological alternations in the curly brackets {·} are listed in Table A.2 below.

Cyrillic	Latin	Transcription
А, а	A, a	*a*
Б, б	B, b	*b*
В, в	W, w	*w*
	V, v	†*v*
Г, г	G, g	*g*
	Ğ, ğ	*ɣ*
Д, д	D, d	*d*
Е, е	E, e	*e*
	Yı, yı	*jï*
	Ye, ye	*je*
*Ё, ё	Yo, yo	†*jo*
Ж, ж	J, j	*ž*
Җ, җ	C, c	*ź*
З, з	Z, z	*z*
И, и	İ, i	*i*
Й, й	Y, y	*j*
К, к	K, k	*k*
	Q, q	*q*
Л, л	L, l	*l*
М, м	M, m	*m*
Н, н	N, n	*n*
Ң, ң	Ñ, ñ	*ŋ*
О, о	O, o	*o*
Ө, ө	Ö, ö	*ö*
П, п	P, p	*p*
Р, р	R, r	*r*
С, с	S, s	*s*
Т, т	T, t	*t*
У, у	U, u	*u(w)*
	W, w	*w*
Ү, ү	Ü, ü	*ü(w)*
	W, w	*w*
Ф, ф	F, f	*f*
Х, х	X, x	*x*
Һ, һ	H, h	*h*
*Ц, ц	Ts, ts	†*c*
Ч, ч	Ç, ç	*č*
Ш, ш	Ş, ş	*š*
*Щ, щ	Şç, şç	†*šč*
ъ	—	—
Ы, ы	I, ı	*ï*
ь	—	—
Э, э	E, e	*e*
	'	'
Ю, ю	Yu, yu	*ju(w)*
	Yü, yü	*jü(w)*
Я, я	Ya, ya	*ja*
	Yä, yä	*jä*

Table A.1: Correspondence of the Cyrillic and Latin orthographies with the transcription employed in this thesis. An asterisk means that the character is used only in Russian-origin morphemes. A dagger means that the transcription is used only in Russian-origin morphemes.

84

Notation	Definition
{A}	Alternation of low vowels *a~ä*
{I}	Alternation of mid unrounded vowels *ï~e*
{U}	Alternation of high rounded vowels *u~ü*
{D}	Alternation of dental consonants *d~t*
{G}	Alternation of velar consonants *k~g~q~ɣ*

Table A.2: Notations for phonological alternations.

Appendix B

Glossing abbreviations

For detailed use of the glossing abbreviations, see the original literature.

ABL — ablative

ACCOBL — accusative-oblique

ADN — adnominal

AF — actor focus

AUX — auxiliary

B — form B in Amdo Tibetan

CAUS — causative

COP — copula

CVB — converb

DAT — dative

DECL — declarative

EMP — emphatic

EP — epenthesis

ERG — ergative

F — feminine

FUT — future

GEN — genitive

INDEF — indefinite

INF — infinitive

INFER — inferential evidential

IPFV — imperfective

IRR — irrealis

KO — clitic =kò/gò in Burmese

LINK — linking interfix

LK — linker

LOC — locative

LV — lengthened vowel

M — masculine

MOD — modified

NEG — negative

NM — non-masculine

NOM — nominative

NPC — noun-phrase enclitic

NPST — non-past

OBL — oblique

PART — particle

PASS — passive

PFV — perfective

PL — plural

POSS — possessive

PRED — predicative

PROG — progressive

PROP — proprietive

PROS — prospective

PRS — present

PS — past stem

PST — past

PTCP — participle

Q — polarity question clitic

REC — reciprocal

REPORT — reportative evidential

RLS — realis

SG — singular

TOP — topic

TV — terminating verbal

VADJ — verbal adjective

VN — verbal noun