C-3 Mermaid Construction in Tatar

Chihiro Taguchi
Nara Institute of Science and Technology
taguchi.chihiro.td0@is.naist.jp

Abstract

This study reports the presence of the Mermaid Construction (MMC) in Tatar (< Kipchak < Turkic). MMC, first coined by Tsunoda (2011) [4], is a mono-clausal compound-predicate construction that superficially appears to contain two clauses. The compound predicate consists of a predicate, a noun, and a copula, of which the noun has undergone grammaticalization to some extent and has lost or is partially losing its meaning or nominal function. The first comprehensive study on the cross-linguistic distribution of the MMC was conducted in Tsunoda (2020a) [5], in which 27 languages are reported to have the MMC. Although Tatar is not mentioned in the work [5], it is shown in this study that Tatar has the MMC. In addition, the Tatar MMCs can be categorized into two types in terms of their syntactic behavior. The grammaticalization occurring in the Tatar MMCs suggests new paths of grammaticalization regarding the MMC's emergence.

1 Introduction

1.1 Mermaid Construction

According to Tsunoda (2020b), the Mermaid Construction (MMC) is a compound-predicate construction that typically consists of a predicate, a noun, and a copula [6]. Though the construction superficially looks as though it comprises two clauses, its syntactic function shows a mono-clausal behavior. Sentence (1) is an example of the MMC in Modern Standard Japanese (MSJ), which is reportedly the language having the most nouns that can trigger the MMC [6] (LT stands for a literal translation, and FT for a free translation). At a first glance, (1) seems to contain two clauses, the first of which is an adnominal clause *Hanako=ga ik-u* that modifies the matrix nominal predicate *yotee=da*. However, this intuitive analysis turns out to fail when the nominal predicate with a copula *yotee=da* is used independently, because (2) is a non-sensical sentence, albeit not syntactically ungrammatical.

(1) [*Hanako=ga ik-u*] *yotee=da*.
Hanako-NOM go-NPST plan-COP.NPST
LT: "[Hanako goes] a plan is."
FT: "Hanako plans to go."

(2) ? *yotee=da*.
plan-COP.NPST
"(It is) a plan."

Therefore, Tsunoda analyzed the construction as a mono-clausal structure with a compound predicate *ik-u yotee=da*, rather than a bi-clausal structure with an adnominal clause expressed in (1).

From empirical observations across languages worldwide, Tsunoda (2020b) lists up five common characteristics prototypical of the MMC [6]. Along with this, he also mentions a geographical and typological tendency that the MMC is more observed in languages of East Asia and SOV languages. These properties are useful as a checklist for determining an MMC, but, as Tsunoda (2020b) notes, they are a tentative list and several peripheral variations of MMCs (also called "quasi-MMC") that do not necessarily conform with all of these five are also reported [4]. Languages having marginal or nonconforming features to the MMC include, for example, Thai [3], an SVO language using a nominalizer morpheme instead of a Noun slot, and Tagalog, a VSO/VOS language without a copula.

(3) a. The structure is [Clause][1] Noun Copula, superficially at least.
b. The Noun is an independent word (not a clitic) that is a noun.
c. The subject of the Clause and the Noun are non-coreferential.
d. The Clause can be used as a sentence by itself.
e. The Clause is not the subject of the "Noun + Copula".

[1]Note that this capitalized Clause does not mean that it comprises a single independent clause, and that it is merely a term for its superficial appearance. Tsunoda (2020) paraphrases it as [predicate] Noun Copula, which is combined into a compound predicate.

1.2 Tatar language

Tatar is a Kipchak language of the Turkic language family. It is an agglutinative language that employs suffixes for morphological case-marking and verbal inflection. Its canonical word order is SOV, and a modifier precedes its head. In colloquial or emotional speech, however, arguments of a sentence may be scrambled. Sentence (4) shows an example of a subject inverted after a main predicate, and sentence (5) shows an inversion of a possessive phrase.

(4) *Nikter söt produkt-lar-ï-n küp+itep satïp+al-a anïŋ äti-se.*
somehow milk product-PL-POS.3-ACC many buy-PRS.3 3SG.POS father-POS.3
"Somehow his (her) father buys many dairy products."

(5) *Belem däräžä-se šaqtïj tübän i-de šul bu klass-nïŋ.*
knowledge level-POS.3 very low COP-PST.3 EMP this class-GEN
"The level of study in this class was very low."

2 MMC in Tatar

This section discusses the conformity of the Tatar MMCs with the prototypical properties of MMCs listed in the previous section. It will be demonstrated that the Tatar MMCs possess all the features in (3). First of all, a typical example of the Tatar MMC is shown in (6). The Clause slot of the general formula in (3)a is filled by an infinitival phrase, the Noun slot by *isäp* "counting, thought" or *nijät* 'intention", the Copula slot by a zero copula[2]. The superficial prototypical structure of the Tatar MMC is given in (7). It is necessary to note that the Tatar MMC contains an infinitive verb, because the syntactic-semantic function of Tatar infinitives is polysemous and is understudied. In subsection 2.1, the semantics of Tatar infinitives and their syntactic behaviors are illustrated. Then, a detailed analysis of this construction will be given in subsection 2.2.

(6) *Öj-gä qajt-ïrɣa isäp / nijät* ∅ .
house-DAT return-INF counting / intention COP
LT: "[To go home] the intention is."
FT: "(One) plans to go home."

(7) Superficial structure of the Tatar MMC
[Infinitival Clause] Noun (Copula)

2.1 Tatar Infinitives

The infinitive in Tatar is a non-finite form of the verb. It does not inflect for person, number, or tense. Morphologically, it is expressed by means of the suffix *-(I)rGA* or *-ArgA*, whose negative forms are *-mAsKA*[3].

2.1.1 Independent infinitive

Even though the Tatar infinitive is non-finite, it can exceptionally appear by itself in a matrix clause to convey a deontic modal meaning, as exemplified in (8) and (9). In addition, when combined with a past tense copula, it expresses a speaker's wish. These constructions with a bare infinitive are restricted to present tense, and the subject of the verb is not syntactically overt[4].

[2]Tatar has a zero copula in present tense, which is unmarked. When the copula is in other tenses than present, an auxiliary *i-* or an alternative copular verb *bul-* is used.

[3]Letters in capital denote that they have one or more alternations due to vowel harmony. In Tatar, *I* is either *ï* or *e*, *A* either *a* or *ä*, *G* either *ɣ* or *g*, and *K* either *q* or *k*; the former candidate is employed when the vowel of the preceding syllable is back, and the latter when it is front. The parentheses express that the phonemes in them appear in some cases; in Tatar, *I* is inserted when the verb stem ends with consonant. *-ArgA* is used when the verb stem is monosyllabic with a short vowel and does not end with a rhotic (i.e., *l* or *r*).

[4]For the bare-infinitive structure with a deontic meaning, the subject can be expressed as a dative argument. Whether this is a dative subject or an oblique argument remains unclear.

(8) *Tïrïš-ïp uqï-rɣa.*
endeavor-CVB study-INF
"Study hard." (imperative, deontic modality)

(9) *nišlä-rgä?*
do:what-INF
"What to do?" (deontic modality)

(10) *Öj-gä qajt-ïrɣa* i-de.
house-DAT return-INF COP-PST
"(I would like to go home."

2.1.2 Infinitive as a modifier

An infinitival clause serves as an adverbial clause that modifies a verbal predicate to express intent or purpose (e.g., (11))[5]. For a limited number of verbal predicates, it may carry a modal (permission; e.g., *jara-* "may") or aspectual (inchoative; e.g., *totïn-*, *kereš-* "to start") meaning. It can also modify an adjectival predicate; when it modifies a non-modal adjective, it provides a degree of comparison for carrying out the action (e.g., (12))[6]. Modification of modal adjectives is more common in use, and it adds a modality (possibility, obligation, necessity, permission) to the event described by the infinitival phrase (e.g., (13)[7].

(11) *Dust-ïbïz-nï aeroport-qa qarší+alïrɣa bar-dï-q.*
friend-POS.1PL-ACC airport-DAT pick.up-INF go-PST-1PL
"We went to the airport to pick up our friend."

(12) *Ul Ginness kitab-ïn-a ker-ergä lajïq.*
3SG.NOM Guinness book-POS.3-DAT enter-*inf* worth
"He/She/It is worth being listed in Guinness World Records."

(13) *Irtägä tügel, ä bügen ük min kitärgä tiješ-men.*
tomorrow not but today EMP 1SG.NOM leave-INF obligatory-1SG
"Not tomorrow but today I have to leave."

In general, Tatar infinitives do not modify nouns directly without a verbal predicate; however, they seem to modify a noun when the noun is abstract and is the subject of an existential predicate *bar* "there is/are" or its negative counterpart *yuq* "there is/are not", as shown in (14), (15)[8] and (16). The structure of the Tatar MMC also fits in this category.

(14) *Äjt-ergä xaq-ïm bar.*
say-INF right-POS.1SG exist
"I have a right to say."

(15) *Äjt-ergä nigez-em bar.*
say-INF base-POS.1SG exist
"I have a reason to say."

(16) *Tizdän Qazan-ɣa taɣïn kil-ergä isäb-em bar.*
soon Kazan-DAT again come-INF counting-POS.1SG exist
"I am going to come to Kazan again soon."

2.2 Contrastive analysis

In this subsection, we look in depth at the behavior of the Tatar MMC in (6) from a contrastive point of view based on the five properties given by Tsunoda (2020b), as well as its mono-clausal structure possibly caused by grammaticalization [6].

[5]This function is widely observed. Common collocation verbs are *jarat-* "to love", *tïrïš-* "to try", *telä-* "to want", etc.

[6]As for non-modal adjectives, the infinitive cooccurs with a rather smaller set of adjectives. Observed collocations are *xaqlï* "having a right, deserving", *äzer* "ready", *sälätle* "talented, able", etc.

[7]This construction is frequently observed. Modal adjectival predicates that often cooccur with infinitives are *kiräk* "necessary, must" and *mömkin* "possible, can".

[8]*nigez* means "basis, fundamental" as a concrete noun, but it changes to an abstract meaning "reason, rationale" in the infinitival construction.

2.2.1 Five cross-linguistic properties of MMC

The Tatar MMCs satisfy all the five prototypical conditions as illustrated below.

a. Superficial structure: [Clause] Noun Copula.
In (6), *Öj-gä qajt-ïrɣa* (“to go home”) is an infinitival clause. Since the copula in the present tense is usually zero, (6) superficially has the structure of [Clause] Noun Copula (cf. (3)a).

b. The Noun is an independent word.
This property is confirmed by the fact that the Noun slot is filled by either *isäp* (“thought, intention”) or *nijät* (“intention”), both of which are independent nouns.

c. The subject of the Clause and the Noun are non-coreferential.
We cannot observe the subject of the Clause since there is no overt subject in Tatar infinitival clauses. However, the intender can be expressed by adding a personal possessive suffix to the Noun as in (17). If we assume a covert PRO as the subject of the Clause in (17), it is non-coreferential with the Noun in the sense that the intender is not the intention itself.

(17) [*Öj-gä qajt-ïrɣa*] *isäb-em* / *nijät-em* ∅.
house-DAT return-INF counting-POS.1SG / intention-POS.1SG COP
LT: “[To go home] my intention is.”
FT: “(One) plans to go home.”

d. The Clause can be used as a sentence by itself.
Tatar MMCs also fit the fourth diagnostic, as the infinitive clause can be independently used with a modal meaning as shown in (18). Furthermore, the Clause slot can also be filled by an infinitival structure with *ide* like (10), as expressed in (19).

(18) [*Öj-gä qajt-ïrɣa*].
house-DAT return-INF
LT: “To go home.”
FT: “(It is in the cards that one) goes home.”

(19) [*Öj-gä qajt-ïrɣa* i-de] *isäp* / *nijät*.
house-DAT return-INF COP-PST intention
LT: “[(One) would like to go home] the plan is.”
FT: “(One) plans to go home.”

e. The Clause is not the subject of the “Noun + Copula”.
The fifth property (3)e also applies to Tatar MMCs. Infinitival clauses in Tatar do not have a function as a nominal phrase, and therefore they are never the subject of the “Noun + Copula”.

2.2.2 Grammaticalization and mono-clausality

The mono-clausality of the Tatar MMCs is supported by the fact that deleting the Clause as in sentence (20) merely results in a non-sensical sentence.

This evidence supports that the structure [INF + *isäp* / *nijät* (Copula)] is a mono-clausal MMC. As Tsunoda (2020) notes, MMCs show (at least) a weak degree of grammaticalization [6]. This is true of Tatar, too, as the Noun slot is highly selective, and other semantic synonyms, for instance *plan* as in (21), cannot be used in the slot. From a contrastive viewpoint, the Tatar MMC [INF + *isäp* / *nijät* (Copula)] has the meaning of intention or plan, and it is the most typical meaning commonly shared among MMCs listed by Tsunoda (2020) [6].

(20) ? *Isäp* / *Nijät i-de*.
intention COP-PST
“It was the intention.”

(21) *[*Öj-gä qajt-ïrɣa*] *plan*.
house-DAT return-INF plan
Intended: "(One) plans to go home."

In addition, the Tatar MMC typically allows present tense, and the acceptability MMCs in past tense varies among speakers. The restrictiveness of tense in some MMCs is also found in other languages. Sentence (22) is an example in Japanese, which has roughly the same meaning as (8), and its past tense counterpart in (23) does not make sense.

(22) *Seezee benkyoosuru koto=da*.
hard study.NPST thing=COP
"Study hard.", or "You should study hard."

(23) * *Seezee benkyoosuru koto=datta*.
hard study.NPST thing=COP.PST
Intended: "You should have studied hard."

2.3 Other nouns that allow the MMC

There are two more nouns that are found in the Noun slot of the MMC prototype: *waqït* "time" and *röxsät* "permission", which, as opposed to the previous MMCs with no overt subject, can have (apparently) a subject in dative.

(24) *Jarar*, *miña joqla-rɣa waqït*.
well 1SG.DAT sleep-INF time
"Well, it's time for me to sleep."

(25) *Aña jeget bülmä-se-nä ker-ergä röxsät*.
3SG.DAT boy room-POS.3-DAT enter-INF permission
"He/she is allowed to enter the boy's room."

However, these MMCs select a different position for a verb-modifying enclitic to be attached to. For example, while the yes-no interrogative enclitic *=mI* succeeds the noun in (26) and (27), it is fixed to the infinitive verb for the cases of *isäp* and *nijät*, as described in (28) and (29). Similarly, the enclitic *=TIr* that adds an irrealis (presumptive) modality meaning appears in the same position as exemplified in sentence (30).

(26) *Joqla-rɣa waqït=mi*?
sleep-INF time=Q
"Is it time for sleep?"

(27) *Ker-ergä röxsät=me*?
enter-INF permission=Q
"Is it allowed to enter?"

(28) *Öj-gä qajt-ïrɣa=mï isäp?*
house-DAT return-INF=Q counting
"Are you planning to go home?"

(29) * *Öj-gä qajt-ïrɣa isäp=mï*?
house-DAT return-INF counting-Q
Intended: "Are you planning to go home?"

(30) *Šunda qunaqxanä-dä jäšäp tor-ïrɣa=dïr isäb-e*.
in_there hotel-LOC live-CVB stay-INF=IRR counting-POS.3
"He/She is going to stay in the hotel there (maybe)."

This discrepancy in the distribution provides evidence that these two types of MMCs have a different syntactic structure. The details are discussed in the next section.

3 Flows of grammaticalization

Tsunoda (2020), though tentatively, mentions four possibilities that account for the emergence of the MMC: (i) adnominal clauses; (ii) noun-predicate sentences whose subject is a complement clause; (iii) cleft sentences; (iv) language contact [6]. In this section, a possible path of grammaticalization that caused the MMC in Tatar is proposed. The grammaticalization process first starts with the extension of the infinitive's usage ([INF] + Noun

+ Existential), and then is followed by dropping of an existential predicate ([INF] + Noun (+ Copula)). The adjacent combination of the infinitival predicate and the noun forms a compound predicate that has acquired an abstract meaning related to tense–aspect–modality (TAM). In addition, I mention another possibility that the grammaticalization may be caused by subject inversion. These explanations of the MMC's emergence in terms of ellipsis and subject inversion are new possibilities that are not argued in Tsunoda (2020b) [6].

3.1 Extension of infinitival modification

As observed in subsection 2.1, the Tatar infinitive shows a syntactically and semantically limited distribution when it is used as an adnominal modifier. Modification of verbal and adjectival predicates is fairly flexible, but modification of a noun is usually possible when the noun is abstract and the noun is the subject of an existential predicate. Presumably, this is an extended usage from the former to the latter; however, at this point, this is merely a speculation that lacks a further diachronic analysis.

3.2 Ellipsis of an existential predicate

As the next step of grammaticalization after the extension of infinitival modification, there is assumed to be an ellipsis of the existential predicate for a limited number of nouns. This argument is inspired by a comment upon elicitation from native speakers that the MMC in (6) is a "shorter way" of saying (31). A piece of supporting evidence for this claim is that realization of the MMC in Tatar is restricted to present tense. Sentence (18) does not allow its past tense counterpart (32), and instead it has to be expressed in a full form as in (33). I thus propose an ellipsis as a grammaticalization process that a limited number of nouns (*isäp*, *nijät*, *waqït*, *röxsät*) underwent.

(31) *Öj-gä qajt-ïrɣa isäp / nijät bar.*
house-DAT return-INF counting / intention EXIST
"I am planning to go home."

(32) * *Öj-gä qajt-ïrɣa isäp / nijät i-de.*
house-DAT return-INF counting / intention COP-PST.3
Intended: "I was planning to go home."

(33) *Öj-gä qajt-ïrɣa isäp / nijät bar i-de.*
house-DAT return-INF counting / intention exist COP-PST.3
"I was planning to go home."

3.3 Formation of a compound predicate: noun vs. quasi-particle noun

As we have seen in subsection 2.3, there is a difference in the syntactic distribution of *isäp*, *nijät* and *waqït*, *röxsät*. The syntactic behavior of the former type is similar to that of particles modifying the verb phrase. For example, *=inde* "already" in sentence (34) is a particle functioning as a modifier to the verb phrase *qajttïŋ*. In this case, the interrogative enclitic *-mI*, which is normally attached to the sentence-final word, is not attached after the particle *inde*, but after the verb phrase *qajttïŋ*.

(34) *qajt-tï-ŋ=mï inde?*
return-PST-2.SG=Q already
"Are you back already?"

(35) * *qajt-tï-ŋ inde=me?*
return-PST-2.SG already=Q
Intended: "Are you back already?"

This syntactic restriction resembles that of *isäp* and *nijät*. Therefore, I posit that, in the MMC, these two nouns have lost their canonical nominal status, and instead have a particle-like — yet subject to nominal morphology — function that plays a role of adding TAM information (namely, intention, future, irrealis, etc.). As for *waqït* and *röxsät*, they form a compound predicate in the MMC maintaining the syntactic nominal properties.

3.4 Another possibility: subject inversion

It is worth noting that we can assume another scenario as to how the MMCs of *isäp* and *nijät* were formed; that is, the noun-ending structure of the MMCs with *isäp* and *nijät* might come from the subject inversion. As mentioned in subsection 1.2, a subject is sometimes inverted after a main verb in a sentence. Even with the transposed subject, the particle remains in its own position, as contrasted in (36) and (37). From this fact, it is possible to hypothesize another way of grammaticalization described in (38). However, this is theoretically less plausible than that proposed in subsection 3.2, because Tatar infinitives do not have a nominal function and, therefore, the hypothetical structure in (38) is ungrammatical.

(36) *min* *Tatarča döres* *söjlim=me?*
1SG.NOM in_Tatar correct speak:PRS.1SG=Q
"Am I speaking correctly in Tatar?"

(37) *Tatarča döres* *söjlim=me* *min?*
in_Tatar correct speak:PRS.1SG=Q 1SG.NOM
"Am I speaking correctly in Tatar?"

(38) Hypothetical structure: * [*isäp* [INF COP]] → Subject inversion: [[INF COP] *isäp*]

4 Conclusion

This study reported the presence of the MMCs in Tatar, and analyzed their cross-linguistic uniqueness through the comparison with the five properties of the MMC proposed in Tsunoda (2020b) [6]. It was demonstrated that there are at least four nouns that can be used in the Tatar MMCs: *isäp* "counting", *nijät* "intention", *waqït* "time", and *röxsät* "permission". These nouns can be categorized in two groups in terms of their syntactic behavior in the MMCs. The first two nouns in the compound predicate of the MMC show an idiosyncratic behavior which syntactically rather resembles modal particles. In contrast, the latter two form a compound predicate with an infinitive verb, while maintaining a property of a nominal predicate. Considering these differences, this study provided two new possible explanations for the grammaticalization of the MMC.

Abbreviations 1, 2, 3 – first, second, third person; COP – copula; CVB – converb; DAT – dative; EMP – emphatic particle; GEN – genitive; INF – infinitive; IRR – irrealis; NEG – negative; NOM – nominative; NPST – non-past; PL – plural; POS – possessive suffix; PRS – present; PST – past; Q – yes-no question clitic; SG – singular.

References

[1] Ebata, Fuyuki. (2020). Sakha (Yakut). In T. Tsunoda (ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance*. De Gruyter, pp. 363–85.

[2] Katagiri, Masumi. (2020). Tagalog. In T. Tsunoda (ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance*. De Gruyter, pp. 781–816.

[3] Takahashi, Kiyoko. (2020). Thai. In T. Tsunoda (ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance*. De Gruyter, pp. 737–80.

[4] Tsunoda, Tasaku. (2011). Mermaid Construction: A contribution from Japanese linguistics to general linguistics [in Japanese]. *NINJAL research papers* (1). National Institute for Japanese Language and Linguistics, pp. 53–75.

[5] Tsunoda, Tasaku. (2020a). *Mermaid Construction: A compound-predicate construction with biclausal appearance*. De Gruyter.

[6] Tsunoda, Tasaku. (2020b). Mermaid Construction: Introduction and summary. In T. Tsunoda (ed.), *Mermaid Construction: A compound-predicate construction with biclausal appearance*. De Gruyter, pp. 1–62.