Grammaticalization of Modal Nominal Predicates in Tatar

Chihiro Taguchi

University of Notre Dame / Kellogg Institute
ctaguchi@nd.edu
https://ctaguchi.github.io

August 4, 2023
21st International Conference on Turkish Linguistics
at Johannes Gutenberg-Universität Mainz

UNIVERSITY OF NOTRE DAME | ENGINEERING

KELLOGG INSTITUTE FOR INTERNATIONAL STUDIES
exploring DEMOCRACY and HUMAN DEVELOPMENT

Abstract

This study...

- Describes grammaticalized modal nominal predicates in Tatar;
- Demonstrates that the nominal predicates are control predicates;
- Provides a syntactic analysis of the constructions (synchronic account);
- Proposes a possible explanation to the emergence of the grammaticalized modal nominal predicates (diachronic account).

The slides are available at: `https://ctaguchi.github.io`

Introduction

- (Kazan) **Tatar** < Kipchak–Bulgar < Kipchak (Northwestern) Turkic < Turkic
- Four (or five) nouns can be used as modal predicates
  - *isäp* "idea, thought" > "plans to", **attitude**
  - *nijät* "idea, thought" > "plans to" (more bookish), **attitude**
  - *röxsät* "permission" > "be allowed to", **deontic**
  - *waqït* "time" > "must (now)" **deontic**
  - ?*mäslixät* "advice" (some speakers do not allow this construction)
- Used with an infinitive verb (*-(I)rGA*)
- Agent is marked by an oblique case
  - **Genitive** for *isäp* and *nijät*
  - **Dative** for *röxsät*, *waqït*, and *mäslixät*

(1) *isäp* "idea, thought"
Marat-nïŋ joqla-rɣa isäb-e.
Marat-GEN sleep-INF idea-POSS.3
'Marat plans to sleep.'

Genitive – Infinitive – Noun: *isäp*, *nijät*

The sentences below are complete finite sentences.

(2) *isäp* “idea, thought”
Marat-nïŋ *joqla-rɣa* *isäb-e.*
Marat-GEN sleep-INF idea-POSS.3
‘Marat plans to sleep.’

(3) *nijät* “idea, thought”
Marat-nïŋ *joqla-rɣa* *nijät-e.*
Marat-GEN sleep-INF idea-POSS.3
‘Marat plans to sleep.’

Dative – Infinitive – Noun: *röxsät*, *waqït*, *mäslixät*

Some constructions take a dative (pseudo-)subject.

(4) *röxsät* "permission"
Marat-qa *joqla-rɣa* *röxsät.*
Marat-DAT sleep-INF permission
'Marat is allowed to sleep.'

(5) *waqït* "time"
Marat-qa *joqla-rɣa* *waqït.*
Marat-DAT sleep-INF time
'It is time for Marat to sleep; Marat must sleep now.'

(6) ?*mäslixät* "advice"
Marat-qa *joqla-rɣa* *mäslixät.*
Marat-DAT sleep-INF advice
'It is recommended for Marat to go to sleep.'

Similar cases: Bashkir

(7) *niät* “idea, thought”

unïŋ	*niät-e*	*kit-ergä.*
3SG:GEN	idea-POSS.3	leave-INF

‘He plans to leave.’

(8) *iθäp* “idea, thought”

unïŋ	*iθäb-e*	*kit-ergä.*
3SG:GEN	idea-POSS.3	leave-INF

‘He plans to leave.’

(9) *röxsät* “permission”

uɣa	*kit-ergä*	*röxsät.*
3SG:DAT	leave-INF	permission

‘He is allowed to leave.’

(10) *waqït* “time”

uɣa	*kit-ergä*	*waqït.*
3SG:DAT	leave-INF	time

‘It’s time for him to leave.’

Similar cases: Non-Turkic languages

Similar constructions can be found in a number of non-Turkic languages:

(11) Welsh

rhaid	*i*	*fi*	*godi'n*		*gynnar.*
rhaid	i	fi	godi	yn	gynnar
necessity	to	me	wake_up.VN	in	early

'I have to wake up early.'

(12) Scottish Gaelic

b'	*àbhaist*	*dhi*	*snàmh.*
COP.PST	custom	to:her	swim.VN

'She used to swim.'

(13) Russian

pora	*nam*	*by-l-o*	*uxodi-t'*
time	us.DAT	COP-PST-N	leave.IPFV-INF

'It is time for us to leave.'

IPFV: imperfective, N: neuter, VN: verbal noun

Peculiarities and research questions

Peculiarities

- Difference from the literal translations (non-compositional)
- No finite verb
- Non-nominative agents; Quirky subjects? (Sigurðsson, 1992)
- **Not reported** in descriptive grammars (Poppe, 1961; Ersen-Rasch, 2009; Burbiel, 2018)

Questions

- What is the syntactic structure of these constructions? How can these be finite sentences?
- These constructions are only found in Tatar and Bashkir among the Turkic languages; how did they emerge?

What are those nouns?

Four possibilities:

- Mere canonical nouns: *It is **time** for him to sleep.*
- Grammaticalized auxiliaries: *He **has to** sleep.*
- Raising predicates: *He **seems to** be sleeping.*
- Control predicates: *He **tries to** sleep.*

This presentation shows that the **control predicate** hypothesis is the most plausible.

The nouns are not canonical nouns

If they were canonical nouns, they would need a finite predicate:

(14) *Marat-nïŋ* *joqla-rɣa* *isäb-e* *bar.*
Marat-GEN sleep-INF idea-POSS.3 exist
'Marat has a thought to sleep.'

(15) *Marat-qa* *joqla-rɣa* *röxsät* *bir-de.*
Marat-DAT sleep-INF permission give-PST.3
'S/he gave Marat permission to sleep.'

In contrast, the constructions at issue lack any finite predicate.
Also, these constructions are **lexically highly selective**, unlike English: *It is (good/high) time to sleep.*
Therefore, these nouns are not just normal nouns.

The nouns are not auxiliaries

Modal auxiliaries do not change the semantic role relationship of nominal arguments; passivization is possible without affecting the semantic roles.

(16) *Tom might catch Jerry in the future.*

(17) *Jerry might be caught by Tom in the future.* (passivized)

However, the nouns at issue **do** seem to assign a semantic role to the genitive/dative:

- *isäp* and *nijät*: the **agent** of planning; planner
- *röxsät*: the **recipient** of permission
- *waqït*: the **initiator / non-volitional agent** of an urgent event

This observation is verified by the interpretation of the passivized sentences (next slide).

Passivization test: *isäp, nijät*

isäp/nijät: The **agent** (planner) changes after passivization, as is evident in the translations:

(18) *Dus-lar-nïŋ* *Marat-nï* *jarat-ïrɣa* *isäb-e.*
friend-PL-GEN Marat-ACC like-INF idea-POSS.3
'The friends plan to like Marat.'

(19) *Marat-nïŋ* *dus-lar-ï* *tarafïnnan* *jarat-ïl-ïrɣa*
Marat-GEN friend-PL-POSS.3 by like-PASS-INF
isäb-e.
idea-POSS.3
'Marat plans to be liked by his friends.' (less natural)[1]

Furthermore, it is unacceptable when the genitive is inanimate:

(20) **šul* *kitap-nïŋ* *čïɣar-ïl-ïrɣa* *isäb-e*
this book-GEN publish-PASS-INF idea-POSS.3
'The book is planned to be published.' (intended)

[1] The informant preferred to say *Marat jarat-ïl-ïrɣa teli* "Marat wants to be liked" or *Marat-nïŋ isäb-e üz-e-n jarat-tïr-ïrɣa* "Marat's idea is to make (people) like himself."

Passivization test: *röxsät, waqït*

Passivization with *röxsät* and *waqït* is often dispreferred by many speakers.

(21) **Marat-qa jarat-ïl-ïrɣa röxsät.*
Marat-DAT like-PASS-INF permission-POSS.3
'Marat is allowed to be loved.' (Intended)

(22) *?Marat-qa jarat-ïl-ïrɣa waqït.*
Marat-DAT like-PASS-INF time
'It is time for Marat to be loved.[2]'

In fact, the use of passivized verbs with these constructions (*isäp*, *nijät*, *röxsät*, *waqït*) are rarely found in the corpus (Saykhunov et al., 2023).[3] Therefore, passivization clearly affects the semantic role relationships. Thus, the nouns are not auxiliaries.

[2]The informant noted that, while this sentence might be grammatical, they had never heard of this sort of sentence and would instead use other synonymous expressions.

[3]Truly passive cases were *ot-ïl* (defeat-PASS) "to lose (*lit.* be defeated)" and *ƶiŋ-el* (win-PASS) "to lose (*lit.* be won)", both negativized. However, in such cases, it is almost synonymous to *win*, which is agentive.

The nouns are not raising predicates

Data shown so far also suggest that the nouns at issue are not raising predicates, either. Since raising predicates do not assign any semantic role to nominal arguments, passivization with the English raising verb *seem* in (24) does not affect the relationship of the agent and the patient.

(23) *Tom seems to catch Jerry.*

(24) *Jerry seems to be caught by Tom.*

However, passivization with the constructions at issue either changes the semantic role relationship or is unacceptable as discussed in the preceding slides.
Therefore, the nouns are not raising predicates either.

The nouns are **control predicates**

So far the three hypotheses have been rejected:

- ~~Mere noun hypothesis~~
- ~~Grammaticalized auxiliary hypothesis~~
- ~~Raising predicate hypothesis~~
- Control predicate hypothesis

Control predicates: *try*, *plan*, *want*, ...

- An unpronounced pronoun PRO is the subject of the infinitive
- PRO is anaphorically bound by the matrix subject
- In (25), Tom is the agent of *try*, and PRO is the agent of *catch*

(25) *Tom*$_i$ *tries* [PRO$_i$ *to catch Jerry*]$_{CP}$.

(26) *Jerry*$_i$ *tries* [PRO$_i$ *to be caught by Tom*]$_{CP}$.

The nouns are **control predicates**

If we assume that they are control predicates, they would look like:

(27) *Marat$_i$-nïŋ* [PRO$_i$ *joqla-rɣa*]$_{CP}$ *isäb-e.*
Marat-GEN sleep-INF idea-POSS.3
'Marat plans to sleep.'

(28) *Marat$_i$-nïŋ* [PRO$_i$ *joqla-rɣa*]$_{CP}$ *nijät-e.*
Marat-GEN sleep-INF idea-POSS.3
'Marat plans to sleep.'

(29) *Marat$_i$-qa* [PRO$_i$ *joqla-rɣa*]$_{IP}$ *röxsät.*
Marat-DAT sleep-INF permission
'Marat is allowed to sleep.'

(30) *Marat$_i$-qa* [PRO$_i$ *joqla-rɣa*]$_{IP}$ *waqït.*
Marat-DAT sleep-INF time
'It is time for Marat to sleep.'

Syntactic structure: a Minimalist account

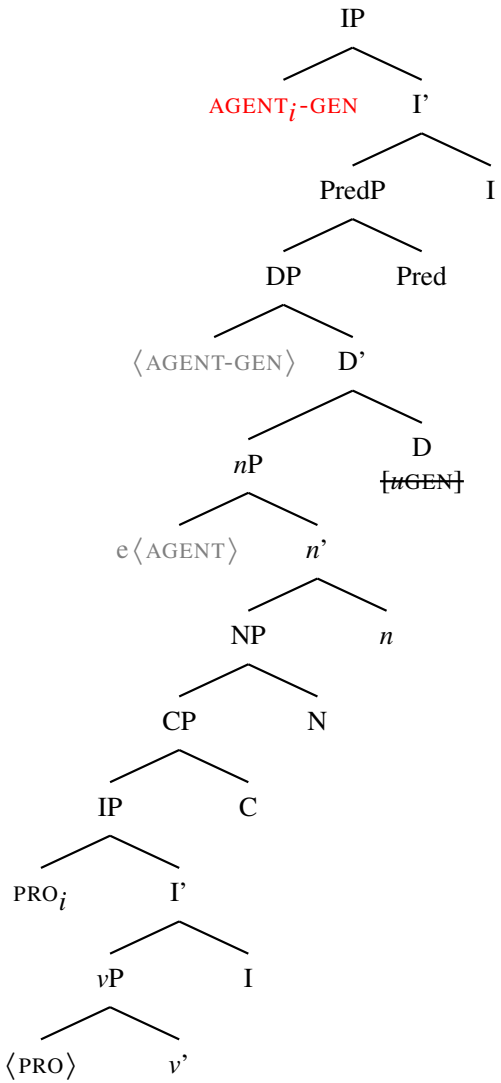
To delve into the structure of the constructions at issue, this section illustrates their syntactic structures based on Minimalism.[4]

- The (genitive) agent is generated under a phrase of a functional category *n* (i.e., Spec,*n*P).
  - This is a standard way in Minimalism to analyze the agent of a noun phrase (Adger, 2003):
  - English: *The army's*$_{\text{AGENT}}$ *destruction of the city*$_{\text{PATIENT}}$.
  - Tatar:

    (31) *Batu-nïŋ*$_{\text{AGENT}}$ *Bulɣar-nï*$_{\text{PATIENT}}$ *ƶimer-üw-e*
    Batu-GEN Bolghar-ACC destroy-VN-POSS.3
    'Batu's destruction of Bolghar' (Wikisource)
- The (genitive) agent is moved to Spec,DP, where it agrees with the GENitive case feature of D.
- The genitive agent is further moved to the Spec,IP position by the Extended Projection Principle (EPP).

[4]This is not necessarily to claim that Minimalism is the only framework that can account for the constructions. See Taguchi (2022) for an account with Lexical Functional Grammar.

Syntactic structure of the genitive type



Syntactic structure of the genitive type

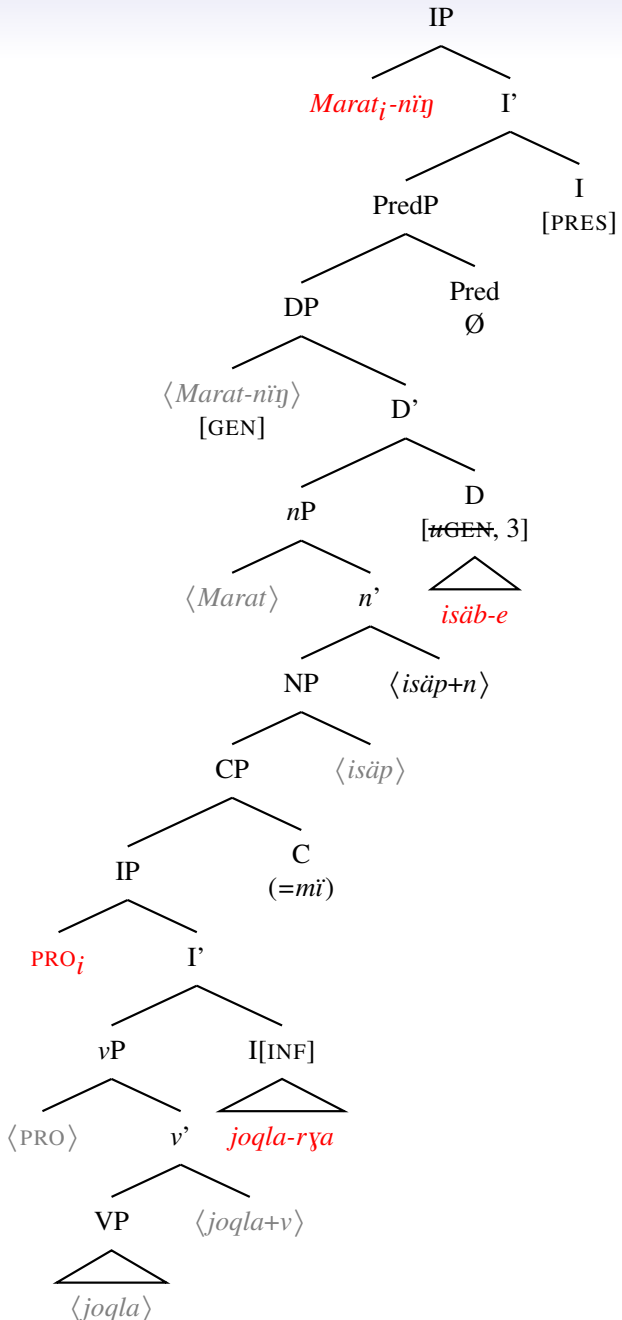
Also, the infinitival clause in the genitive type is assumed to be CP.

- Typically, the question enclitic *=mI* is sentence-final as shown in (32), presumably functioning as the head (C) of CP.
- In the genitive type construction, *=mI* is usually attached to the infinitive as in (33), though some speakers marginally accept sentence-final *=mI* in this construction as in (34).

(32) *Marat* *joqla-dï=mï?*
Marat.NOM sleep-PST.3=Q
'Did Marat sleep?'

(33) *Marat-nïŋ* [*joqla-rɣa=mï*]$_{CP}$ *isäb-e?*
Marat-GEN sleep-INF=Q idea-POSS.3
'Does Marat plan to sleep?'

(34) *??Marat-nïŋ* *joqla-rɣa* *isäb-e=me?*
Marat-GEN sleep-INF idea-POSS.3=Q
'Does Marat plan to sleep?'



- The (genitive) agent is generated at Spec,*n*P
- The (genitive) agent is moved to Spec,DP, where it agrees with the GENitive case feature of D.
- The genitive agent is further moved to Spec,IP by the Extended Projection Principle (EPP).

Syntactic structure of the dative type

There is a difference in the semantics of the dative argument in the *röxsät* and *waqït*-constructions:

- *Röxsät*: the dative argument is the **recipient** of permission
- *Waqït*: the dative argument is the **initiator / non-volitional agent** of the temporally urgent event

Cuervo (2020): **recipients** and **non-volitional agents** are introduced by applicative phrases (ApplP) in different positions:

- **Recipients**: ApplP higher than *v*P (HApplP)
- **Non-volitional agents**: ApplP lower than *v*P (LApplP)
  - cf. Russian impersonal construction with non-volitional agent dative:

(35) ***Boris-u*** *pe-l-o-s^{j}*
Boris-DAT sing-PST-N-REFL
'Boris felt like singing.'

Applying this to our argument:

- *Röxsät*: the dative argument is the Specifier of LApplP under *n*P.
- *Waqït*: the dative argument is the Specifier of HApplP above *n*P.

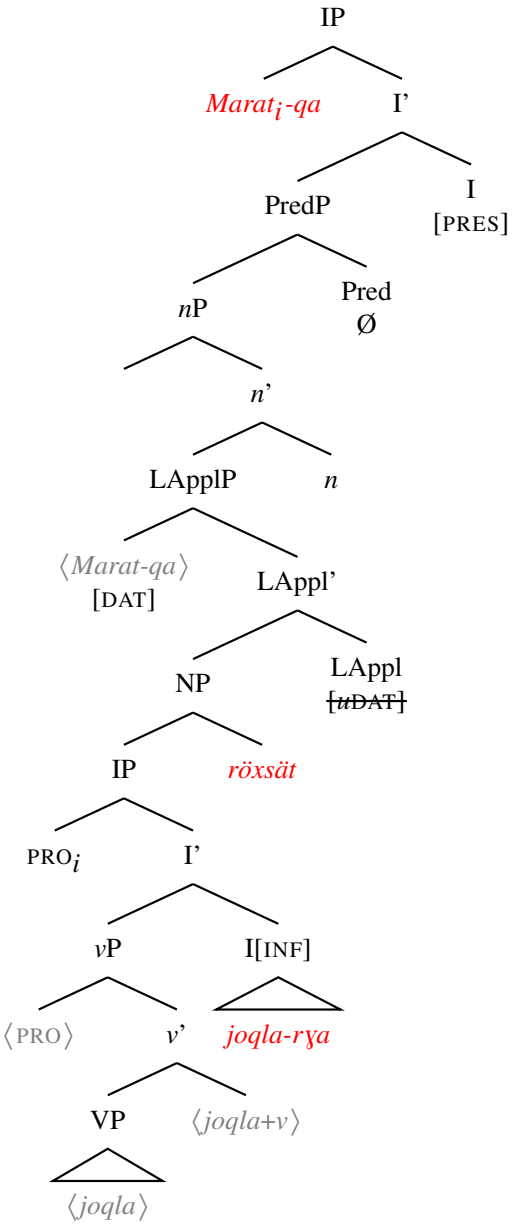
Explaining the difference of acceptability in *röxsät* and *waqït*

This treatment also seems to explain why passivized *waqït* (✓passivization) and passivized *röxsät* (✗passivization) differ in acceptability:

- Although they are both datives, their degree of agentivity differs.
  - *Röxsät*: **recipient**; lower agentivity
  - *Waqït*: **initiator / non-volitional agent**; higher agentivity
- The constructions with higher agentivity are semantically able to allow passivization.[5]
- i.e., an issue of **semantics** rather than of syntax

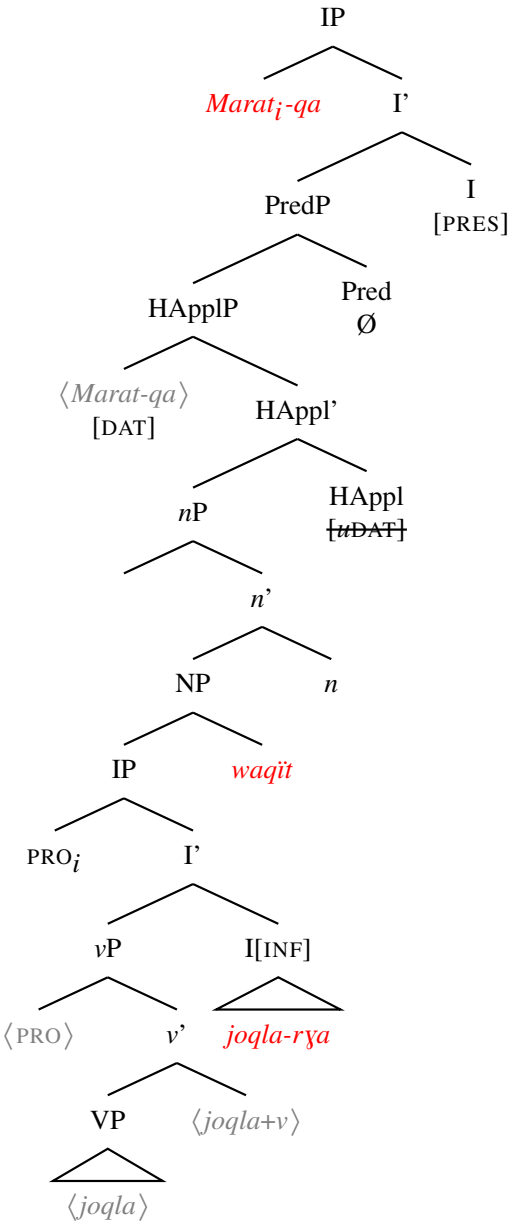
[5] An argument with "higher agentivity" in this context could be what Farkas (1988) calls **initiator**, on which the realization of the infinitival event crucially depends on.

Structure of the *röxsät* construction



- *Röxsät*: **recipient**; lower agentivity → low applicative
- The (dative) argument is introduced at Spec,LApplP

Structure of the *waqït* construction



- *Waqït*: **non-volitional agent**; higher agentivity → high applicative
- The (dative) argument is introduced at Spec,HApplP

Grammaticalization

These constructions are probably grammaticalized, where the main predicate is omitted:

- *Marat-nïŋ joqla-rɣa isäb-e ~~bar~~.*
- *Marat-qa joqla-rɣa röxsät ~~bir-de~~.*

The cline of grammaticalization (Hopper and Traugott, 2003):

- Content word > grammatical word > clitic > affix

The modal nominal predicates in Tatar:

- They have lost the original lexical meaning;
- They retain morphologically independent status;
- Therefore, they are at the stage of **grammatical word**.

Grammaticalization

Four main interrelated mechanisms of grammaticalization by Heine and Kuteva (2002):

- Desemanticization (or semantic bleaching): loss in meaning content;
- Extension (or context generalization): use in new contexts;
- Decategorialization: – loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms
- Erosion (or phonetic reduction): loss in phonetic substance.

The modal nominal predicates in Tatar:

- They are undergoing **desemanticization** (loss of the nominal meaning), **extension** (usage in modal semantics), and **decategorialization** (usage as the main predicate combined with an infinitive)

Language contact?

What caused this grammaticalization?

- Only Kipchak–Bulgar languages (Tatar and Bashkir) have it among the Turkic languages
  - A pure innovation within Kipchak–Bulgar languages?
  - Language contact with the surrounding non-Turkic languages?
- **Volga–Kama Sprachbund**: Bashkir, Chuvash, Komi, Mari, Mordvin, Russian, Tatar, Udmurt (Johanson, 2000; Van Pareren, 2011)
- At least **Udmurt** has a similar construction

Figure: A map of the Volga–Kama region, from BBC (`https://www.bbc.com/news/magazine-29950844`)

Volga–Kama Sprachbund?

Modal nominal predicates found in Udmurt

- Udmurt < Permic < Finno–Ugric < Uralic
- Extensive language contact with Tatar (Edygarova, 2022)
- Lexical borrowing: Ud. *kijar* from Tt. *qïjar* (cucumber)
- Morphological borrowing:
  - Ud. *-tɕi* < Tt. *-čï/če* (suffix for profession)
  - Ud. *-lɨk* < Tt. *-lïq/lek* (suffix for forming an abstract noun)

(36) Udmurt[6]

Ivi-len	*kniga*	*pottɨ-nɨ*	*mɨlkɨd-ɨz*	*vań*
Ivi-GEN	book	publish-INF	mood-POSS.3	COP

'Ivi is considering publishing a book.'

A hypothesis: The grammaticalized modal nominal predicates are results of language contact in the Sprachbund?
Further cross-linguistic investigation in the Sprachbund is needed.

[6]Elicited from a native speaker

Conclusions

This presentation has...

- Identified the syntactic properties of the grammaticalized modal nominal predicates in Tatar: *isäp*, *nijät*, *röxsät*, *waqït*
  - ~~Normal noun~~, ~~Auxiliary~~, ~~Raising~~, Control
- Provided a (Minimalist) syntactic analysis for the structure of the constructions
  - Genitive argument is generated at Spec,*n*P (*isäp*, *nijät*)
  - Dative argument is generated at either Spec,HApplP (*waqït*) or Spec,LApplP (*röxsät*)
  - These also account for the different acceptabilities in passivization of the embedded clause
- Suggested a possible hypothesis for their emergence in Tatar (and Bashkir)
  - Volga–Kama Sprachbund? (cf. Udmurt)

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. BCS-2109709.
My conference travel to attend the International Conference on Turkish Linguistics was made possible by the Kellogg Institute's generous support.

Last but not least, massive thanks to my informants of Tatar and Udmurt who put up with my endless elicitation and acceptability questions!

References

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

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

Cuervo, M. C. (2020). Datives as applicatives. In *Dative constructions in Romance and beyond*. Language Science Press.

Edygarova, S. (2022). Udmurt. In Bakró-Nagy, M., Laakso, J., and Skribnik, E., editors, *The Oxford Guide to the Uralic Languages*. Oxford Academic.

Ersen-Rasch, M. I. (2009). *Tatarisch: Lehrbuch für Anfänger und Fortgeschittene*. Harrasowitz.

Farkas, D. (1988). On obligatory control. *Linguistics and Philosophy*, 11(1).

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

Hopper, P. J. and Traugott, E. C. (2003). *Grammaticalization*. Cambridge Textbooks in Linguistics. Cambridge University Press, 2nd edition.

Johanson, L. (2000). Linguistic convergence in the Volga area. *Studies in Slavic and General Linguistics*, 28:165–178.

Poppe, N. (1961). *Tatar manual: Descriptive grammar and texts with a Tatar-English glossary*. Indiana University, Bloomington.

Saykhunov, M. R., Khusainov, R. R., Ibragimov, T. I., Luutonen, J., Salimzyanov, I. F., Shaydullina, G. Y., and Khusainova, A. M. (2023). Corpus of written tatar.

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

Taguchi, C. (2022). Mermaid Constructions in Lexical Functional Grammar. In *The Proceedings of the LFG'22 Conference*.

Van Pareren, R. (2011). Areal features in the Volga-Kama region: On some non-lexical Turkic influences in Mordvin. *Studies in Slavic and General Linguistics*, 38:251–265.