

Cem Keskin*

Transient subordinate clauses in Balkan Turkic in its shift to Standard Average European subordination. Dialectal and historical evidence

<https://doi.org/10.1515/flin-2023-2001>

Received April 29, 2022; accepted September 3, 2022; published online February 16, 2023

Abstract: The Turkic varieties of the Balkans use two main diametrically opposed subordination strategies: (i) the Turkic model, where typical subordinate clauses are prepositive, nonfinite, contain clause-final subordinators, etc. and (ii) the Indo-European model, where typical subordinate clauses are postpositive, finite, contain clause-initial subordinators, etc. The paper observes that Balkan Turkic additionally uses several kinds of subordinate clause that allow for problematic mixtures of these two models ('X-clauses'). Spread over a spectrum between the Turkic and Indo-European extremes, X-clauses can, for instance, be prepositive but contain clause-initial subordinators. The paper, then, hypothesizes that X-clauses emerge due to uncertainties in the structural parameters of the Balkan Turkic subordination system. Such uncertainties are typical of complex systems undergoing change and arise in the present case due to the shift in Balkan Turkic away from Turkic towards Indo-European subordination.

Keywords: Balkan linguistic area; contact-induced syntactic change; subordination; systems theory; transient behavior; Turkic languages

1 Introduction

Turkic varieties spoken in the language contact setting of the Balkan sprachbund make use of two main subordination strategies:¹ the Turkic and the Indo-European

¹ As the definition of subordination, we can adopt a blend of the definitions offered in Brown and Miller (2013: 425) and Crystal (2008: 462): a type of syntactic linkage between clauses in which they have different syntactic status, one being dependent upon the other, and usually a constituent of the other. This definition subsumes both subordinate clauses proper, i.e. clauses that are embedded in

***Corresponding author: Cem Keskin**, Department of German Studies, University of Potsdam, 14469 Potsdam, Germany, E-mail: keskin@uni-potsdam.de. <https://orcid.org/0000-0002-7398-7032>

(IE). It is generally observed that while the native Turkic subordinate clauses (SC) are in marked decline, IE-type SCs represent the preponderant model among these varieties on average (e.g. Friedman 2006; Matras 2003; Matras and Tufan 2007; Menz 2001, 2006). Within the class of IE-type SCs itself, the dominant type tends to be those modeled after Standard Average European (SAE) SCs,² while Persian-type SCs usually constitute a sizeable minority. In addition to SCs that conform to the Turkic and IE models, those that mix the properties of the two in several different ways are also attested in the region with marginal frequencies and present a problem regarding the shift in the subordination system of Balkan Turkic. Let us now briefly study the two main models of subordination referred to, before we turn to the problematic hybrid clauses.

1.1 Models of subordination in Turkic

The typical SC in Turkic languages has four features which are relevant to the scope of the present study: (i) its predicate is a nonfinite verb form; (ii) it is positioned before the head noun or the matrix verb, i.e. it is prepositive; (iii) it is embedded by means of a subordinative element suffixed to its predicate; (iv) if it does involve a free subordinative element, that element is clause-final. This is the native subordination model (see e.g. Csátó and Johanson 1998: 223–224, 229–233; Johanson 1998b: 48, 57–66, 2021: 854–931). Two illustrative examples are given in (1).

- (1) a. [*Gel-diğ-in-i*] *duy-du-m.*
 come-FNOM-2SG.POSS-ACC hear-PST-1SG
 ‘I heard that you have come.’
- b. *Sen gel-dik-ten sonra*
 2SG come-FNOM-ABL after
 ‘After you have come.’

The bracketed argument clause in (1a) contains a nonfinite predicate (cf. feature [i] above), signaled by the factive nominal suffix which also functions as a subordinative element (cf. [iii]) (Göksel and Kerslake 2005: 424–426; Kornfilt 1997: 50–51). The clause

other clauses, such as complement and relative clauses, as well as cosubordinate clauses which are dependent on other clauses without being embedded in them, such as adverbial clauses.

2 The term ‘Standard Average European’ (Whorf 1944) refers in recent studies to a proposed sprachbund that includes Romance, Germanic, Balto-Slavic, the Balkan languages, etc. (see e.g. van der Auwera 2011; Haspelmath 1998, 2001). The reason for this choice of terminology is that the common SAE linguistic features that are relevant to this paper extend to Balkan Turkic as well: finite subordination strategies, clause-initial subordinators, postnominal relative clauses introduced by a relative pronoun based on an interrogative pronoun (Haspelmath 2001: 1494–1495, 1504).

is positioned before verb *duydum* ‘I heard’ (cf. [ii] above). The adverbial clause in (1b) is again a factive nominal (cf. [i] and [iii]) and also includes the postposition *sonra* ‘after’ that functions as an additional subordinative element (cf. [iv]) (Csató and Johanson 1998: 224; Göksel and Kerslake 2005: 467–485; Kornfilt 1997: 67–68).

The second subordination strategy referred to above, the IE model, is structurally antithetical to the Turkic model and can be seen in Turkic varieties that have been in intensive contact with IE languages, particularly with Persian. Although used frequently, this strategy is perceived as marked and secondary. This model makes SCs with the following characteristics possible: (i) their predicates are finite verb forms; (ii) they are positioned after the head noun or the matrix verb, i.e. they are postpositive; (iii) they are linked to the superordinate clause by means of free subordinative elements; (iv) these subordinative elements are clause-initial (see e.g. Göksel and Kerslake 2005: 409–411, 457–460, 463–465; Johanson 1998b: 65–66, 2021: 867–868, 894–899; Kerslake 2007; Kornfilt 1997: 3, 46, 60, 321–323, 439–440, 443). This strategy is exemplified in (2).

- (2) *Duy-du-m* [*ki* *gel-miş-sin*].
 hear-PST-1SG CONN come-PRF-2SG
 ‘I heard that you have come.’

The bracketed argument clause here can be contrasted with the one in (1a). It has a finite predicate (cf. feature [i] immediately above) marked in perfective and is positioned after the matrix verb *duydum* (cf. [ii]). It is introduced by the free clause-initial connector *ki* (cf. [iii] and [iv]) of Persian origin (see e.g. Göksel and Kerslake 2005: 111–112; Kornfilt 1997: 46).

These descriptions are admittedly a simplified account of the SC types in Turkic languages but should suffice for the purposes of this paper. They can be summarized using a feature decomposition approach as in Table 1, which makes the complementary relation between the two models clearer.^{3,4}

The scheme in Table 1 should not be taken as a literal formal feature analysis; it is intended as an informal and rough descriptive tool, with at least the following flaws. Some features are not necessarily always binary or their values mutually exclusive. For instance, as we will see below, the subordinator may in some cases be found in the middle of the clause or a free subordinator can be used in combination with a suffix (cf. [1b]). The scheme also contains some degree of redundancy. For instance,

3 Color coding is used in Tables 1, 3, 6, and 7 for ease of interpretation.

4 I emphasize that I do not claim nonfiniteness in SCs to be a non-Indo-European feature *in general*. Even a cursory consideration of only the English nonfinite SCs would be enough to refute that claim. The observation here is that *in Turkic languages* IE-type SCs are always finite and never nonfinite (see e.g. Johanson 2021: 866, 895, 903). So, within that context only, nonfiniteness could be seen as a non-IE feature and finiteness as an IE feature (see also Matras 2003: 75).

Table 1: The Turkic versus the IE subordinate clause models.

Features	Turkic	IE
Finiteness	–finite	+finite
Clause position	–postpositive	+postpositive
Subordinator type	–free	+free
Subordinator position	–initial	+initial

nonfinite clauses usually contain suffixed subordinators. Overlooking these shortcomings, I will be using this scheme as the main descriptive tool in the rest of the paper.

1.2 The Indo-European model of subordination in Balkan Turkic

In the light of the descriptions above, Balkan or Rumelian Turkic (RT) emerges as a divergent branch of the Turkic family, because although Turkic languages primarily make use of the native Turkic model and the IE model is marked, the latter is the dominant subordination strategy on average in RT, as mentioned at the beginning. Let me briefly define this group before I provide a description of IE-type SCs there.

From the perspective of the syntactic properties of its members, RT can be said to consist of three subgroups.⁵ The first is the West Rumelian Turkish (WRT) dialect group spoken in the disputed territory of Kosovo in Serbia, North Macedonia, and western Bulgaria (Németh 1956, 1980). The second subgroup is what I refer to as ‘North Rumelian Turkic (NRT)’ which appears to be a continuum of dialects whose southern tip is constituted by the Turkish dialects of north-eastern Bulgaria (for which I use the designation ‘Dobruja Turkish’) and whose northernmost member is Gagauz spoken mostly in Moldova (cf. Boev [1968] cited in Günşen [2012]). The group likely also includes the Turkish dialects in between, i.e. those spoken in Constanța and Tulcea counties in Romania. I will use the cover term ‘Peripheral Rumelian Turkic’ (PRT) here to refer to these two subgroups, based on their syntactic commonalities which we will see in the following sections. The third subgroup of RT, i.e. East Rumelian, is spoken in the greater Thrace region of the Balkans. In this paper, my focus will be almost exclusively on PRT and East Rumelian will mostly be ignored, as its syntax presents nothing of relevance for the purposes of the paper.

Turning now to IE-type SCs in RT, these are particularly prominent in PRT, especially in the WRT group, according to data from the Balkan Turkic Corpus

⁵ Several different classifications of RT varieties have been proposed in the literature. I refer the reader to Günşen (2012) for further information.

(Keskin et al. in preparation b). For instance, in Kosovar Turkish (KT) Turkic SCs are the smallest class, constituting only 14.8% of SCs (32 out of a sample of 216). By contrast, IE-type clauses are at 66.7% (144 out of 216).⁶ As already pointed out, IE-type SCs in RT can be split into two subtypes with different frequencies: (i) Persian-type clauses (28.7% of all SCs in KT; 62 out of 216) which can be distinguished by the Persian subordinators that introduce them (e.g. *ki* in example [2]), (ii) SAE-type clauses (38% of all SCs in KT; 82 out of 216) which are specific to Turkic varieties in contact with European languages. Below are two illustrative examples of SAE-type SCs from KT and Gagauz, respectively.

- (3) a. *O kız-lar [ne cel-i Mamuşa-ya] Türkçe konuş-ur.*
 DIST girl-PL CONN come-PROG.3SG M.-DAT Turkish speak-AOR.3SG
 ‘Those girls who are coming to Mamusha speak Turkish.’
 (Sulçevsi 2019: 240)
- b. *Anna-muş Aliksandri [ani Por on-u yense-ycek].*
 understand-EVID.3SG Alexander CONN Porus 3SG-ACC vanquish-FUT.3SG
 ‘Alexander understood that Porus will vanquish him.’
 (Moškov 1904: 52 via Özkan 2007: 177)

Typical of IE-type SCs in general, the bracketed clauses in these examples are post-positive and contain finite predicates and free clause-initial subordinative elements. The connector *ne* in (3a) is derived from the *wh*-word *ne* ‘what?’, the derivation of subordinators from *wh*-words through processes of grammaticalization being a very productive strategy in WRT as elsewhere in the SAE area. The connector *ani* in (3b) takes various forms (*angi*, *ani*, *hani*, etc.), possibly with different lexical sources, one of which is the *wh*-word *hangi* ‘which?’ (Özkan 1996: 185, 216, 267; Pokrovskaya 1964: 141; Schöniß 1995; see also Menz 1999: 67, 2001: 236).⁷

There now remain 18.5% (40 out of 216) of the SCs in the figures cited from KT to be accounted for and these constitute the group of problematic hybrid SCs referred to at the beginning. I pay special attention to these clauses in the rest of this paper.

6 These counts do not take all SC types in RT into account but focus only on the relevant types.

7 Given that Persian-type and SAE-type clauses are structurally identical (compare examples [2] and [3]), the question of why we should distinguish the two may arise. There are both grammatical and historical reasons for this. First, the depth of subordination of these two clause types are likely to be different (see Matras [2003: 72–73] and Johanson [2021: Sections 55.2.6 and 55.3.8] for diverging views). Second, Persian-type clauses have been attested in the Turkic family for a far longer period than SAE-type clauses. The former are seen in the family from the Old Uyghur period onward (9th–13th c.; Johanson 2021: 895) and, in Turkish specifically, from the Old Ottoman period onward (13th–15th c.). In other words, Persian-type clauses were already being used in Turkish when it arrived in the Balkans (Johanson 1998a: 118–119; Kerslake 1998: 181, 199). Other historical reasons will become clear in Section 10 (see fn. 26 for the rest of this discussion).

2 The structure of the paper

Up to this point, I have provided a description of the RT subordination system, which will serve as a background for the main theme of the paper. But before I move on to the main theme, let me provide some information on the content of the rest of the paper.

First and foremost, in Section 3, I outline the research problem that the present study identifies and addresses. As pointed out immediately above, this problem is constituted by a group of hybrid SCs ('X-clauses'). In that section, I also provide a summary of the answer to this problem that I develop in the rest of the paper ('the transient behavior hypothesis').

Next, in Section 4, I list the textual sources and statistical tools that I use and explain my research method.

In the subsequent two sections, I first flesh out my description of X-clauses. In Section 5, I provide a detailed description of X-clauses in modern PRT varieties using the feature decomposition approach introduced above. In Section 6, I present examples of structures that are akin to X-clauses in various languages. This shows that X-clauses are not just an oddity of PRT.

In the five sections following that, I present the case for the transient behavior hypothesis and the broader diachronic approach to the X-clauses problem that the hypothesis implements. First, in Section 7, I lay out this hypothesis, which, in a nutshell, proposes that X-clauses are 'oscillations' of sorts in the syntactic system. Next, in Section 8, I provide various suggestions as to how the notion of oscillation could be understood in the present context and discuss one implication of that construal. That discussion is one argument for the transient behavior hypothesis. Next, in Section 9, I transition from a contemporary synchronic dialectological perspective to a diachronic one and provide historical data on X-clauses. Thus, I try to establish that X-clauses are not a recent product of modern PRT dialects but have been in use for several centuries. That discussion leads up to an argument for my diachronic approach to X-clauses: in Section 10, I compare 17th and 21st-century WRT from the perspective of subordination and propose a scenario in which X-clauses fit into an account of syntactic shift in that dialect group. I present another argument in favor of the diachronic approach in Section 11 where I propose that a frequency drift has been underway from Early to Modern PRT from X-clauses that are structurally closer to Turkic SCs towards those that are more like IE-type SCs.

In Section 12, I turn away from the approach adopted in the preceding sections that focus on the formal aspects of X-clauses and take up a number of psycho- and sociolinguistic issues pertaining to the shift to SAE-type subordination in PRT and

the attendant emergence of X-clauses. First, I propose a broader contact theoretic account as a context within which the shift to SAE-type subordination and X-clauses in PRT can be considered. Then, I discuss various sociolinguistic factors that could potentially influence the use of X-clauses in PRT. Section 13 concludes the paper.

3 The research problem and the summary answer

So to repeat, in addition to Turkic and IE-type clauses, a group of problematic SCs are in use in PRT. These are attested exclusively in the two PRT groups, with differing frequency distributions. I will dub these structures ‘X-clauses’, a term that I chose to convey a sense of their indeterminate, ambivalent nature: these clauses are seemingly idiosyncratic mixtures of the features of the Turkic and the IE models in Table 1, sometimes also containing novel features that do not come from either model. Thus, they fit neither model and present a problematic pattern. Indeed, they do not even constitute a well-defined class of SC at first glance and look more like a patchwork of clause types each with a low frequency of occurrence. Below are some illustrative examples from the six types occurring at above-average frequency within this group, which account for 81.3% of their occurrences.

- (4) [*Ani sırala-dı-m*] *to urba-lar-ı giy-ē-sin.*
 CONN tell-PST-1SG DIST clothes-PL-ACC wear-AOR-2SG
 ‘You’ll put on the clothes that I told you about.’
 (Razgrad, BG; Murtaza 2016: 81)
- (5) *Çuval doqū-du-lā onlā-dan [ani zāre-ler-i quy-mā].*
 sack weave.AOR-PST-3PL 3PL-ABL CONN grain-PL-ACC put-ANOM.DAT
 ‘They weaved sacks from them to put the grains in.’
 (Razgrad, BG; Haliloğlu 2017: 214)
- (6) *Kimsä de-yär-miş [ani ki o nicä gün duuması].*
 some say-PROG-EVID.3SG CONN CONN 3SG like sun rise
 ‘Some were saying that she is like the sunrise.’
 (Chişinău, MD; Çimpoeş 1988: 5 via Özkan 2007: 157)
- (7) *Sevin-ēr-im [ani mizin ol-du deye].*
 rejoice-AOR-1SG CONN muezzin become-PST.3SG CONN
 ‘I am happy that he became a muezzin.’
 (Razgrad, BG; Murtaza 2016: 112)

- (8) *O sene [bu cade ne yap-ıl-di] çalış-i-dı-k biz orda.*
 DIST year PROX road CONN make-PASS-PST.3SG work-PROG-PST-1PL 1PL there
 ‘The year that this road was built, we were working there.’
 (Mireşh, KO; Sulçevsi 2019: 255)
- (9) *Sü-müş o [ani bizim yaşa-dı-muz] yer-i.*
 plough-EVID.3SG 3SG CONN 1PL.GEN live-OREL-1PL.POSS place-ACC
 ‘He ploughed the place where we lived.’
 (Silistra, BG; Karaşinik 2011: 230)

The SC in example (4) is almost like an IE-type SC in that it is finite and introduced by a clause-initial connector, but it is prepositive like a Turkic SC. The SC in (5) is postpositive and introduced by an initial subordinator, again like an IE-type SC, but is nonfinite (i.e. an action nominal) like a Turkic SC. The SCs in (6), (7), and (8) are postpositive and finite, conforming to the IE model, but are atypical in the subordinators that introduce them. The first has two initial subordinators (viz. *ani ki*);⁸ the second has two subordinators, one clause-initial and the other clause-final (viz. *ani ... deye*); and the last has a clause-internal subordinator (viz. *ne*). Finally, the SC in example (9) is prepositive and nonfinite like a Turkic SC but is introduced by a clause-initial subordinator like an IE-type SC.

So, in terms of the feature scheme in Table 1, the problematic nature of X-clauses stems from the fact that they allow both the Turkic (the ‘–’ pole) and the IE (the ‘+’ pole) feature values in virtually any combination that violates the complementary relation between the two models. For instance, as we have seen in examples (4–9), a given X-clause may be prepositive like a Turkic SC but at the same time contain a clause-initial connector like an IE clause, i.e. [–postpositive, +initial].

We can, then, summarize and restate this problem of RT syntax as in (10):

(10) *The X-Clauses Problem*

In Peripheral Rumelian Turkic, several kinds of SC with marginal respective frequencies are attested that fit neither the Turkic nor the IE model well and allow incompatible value combinations of the component features of these two complementary models.

As stated above, my main purpose in this paper is to put forth and lay out the X-clauses problem to the extent that currently available data allow. My investigation is guided by the following questions, among others, which will be taken up in later

⁸ Clause initial complex subordinators are frequently used in IE-type subordination both in general (e.g. *as if* in English) and in Turkic specifically (e.g. *madem(ki)* ‘since, seeing that’). However, I assume, based on Quirk et al.’s (1985: 998, 1001) observations about English complex subordinators that these connectors are non-core.

sections. What patterns do X-clauses present in PRT? Where do X-clauses fit as a group in the syntactic shift at the expense of Turkic and in favor of SAE-type subordination in PRT? Beyond modern PRT varieties, what is the typological and historical status of X-clauses?

Even though X-clauses may appear to be an ill-defined group of SCs at first glance, I will be treating them together as a cluster, perhaps even a third class of SC in addition to Turkic and IE-type clauses. I have two motivations for doing so. First, as we will see in Section 9, in Ferraguto (1611), which is a Turkish text written in what is probably Early WRT (Keskin 2023a, 2023b; Stein 2016), clauses that fit neither the Turkic nor the IE-model (i.e. my X-clauses) are exclusively introduced by the now-obsolete subordinator *sciú* (i.e. *şu*) ‘that’. This suggests that these clauses do form a class even though they might appear quite different from one another on the surface. Second, as I will show in due course, subjecting them to a unified treatment affords us interesting insights into the emergence of SAE-type SCs in PRT and the diachronic shift in their favor at the expense of Turkic and Persian-type SCs.

Now, the preceding remarks in this section may have given the reader the impression that the X-clauses problem is essentially a classification problem: if X-clauses are neither Turkic nor IE-type clauses, what type of clause are they? This is not my intent. To begin with, there are 13 different subtypes of X-clause – as we will see in Section 5, which would mean there are 13 different classification problems, not a single ‘X-clauses problem’. Far more importantly, treating X-clauses as individual structures without investigating the patterns that the class as a whole conceals would be missing the forest for the trees. As we will see in due time, X-clauses present a complex pattern, whose real extent and nature we are only beginning to reveal and understand – a pattern that lies hidden in what was hitherto likely thought to be mere noise in the data.

Other researchers may, of course, feel justified to approach each type of X-clause individually from their respective theoretical frameworks. However, irrespective of one’s preference in this regard, it is important that we establish the existence of these aberrant SCs in PRT, as they have not yet been recognized for what they are and studied with any systematicity. They are only briefly and partially mentioned in a few works, notably Friedman (2006: 39), Kılıç (2018: 34–35, 68–69), Matras (2003: 73–79; see Section 12.1 for a discussion of that study), Menz (1999: 105–107), and Stein (2016: 165–167) without noting their extraordinary character and aiming to reveal the patterns that underlie them. The identification and detailing of the X-clauses problem is, then, the first contribution of this paper to scholarly literature.

In addition to identifying the X-clauses problem, I also aim to offer a solution to it which explores a diachronic outlook on the diversity of SCs in PRT: once again, less briefly, I hypothesize that X-clauses are manifestations of the shift from Turkic to

SAE-type subordination in PRT syntax. This shift causes oscillations in the values of the features of the PRT subordination system, and the various combinations of feature values yield the various subtypes of X-clause seen in (4–9). This is the transient behavior hypothesis, and the presentation of this potential solution is the second scholarly contribution of the paper.⁹

The X-clauses problem and the transient behavior hypothesis are not just relevant for PRT. As work in progress suggests, they provide insights into patterns emerging in heritage Turkish in contact with German and English (Keskin et al. in preparation a) and may provide at least a partial model for understanding the finer aspects of contact-induced syntactic change in Turkish in general. That is where there is a theoretical lacuna as explained in Section 12.1. Also, there are reasons to think that this model can be applied, *mutatis mutandis*, to other contact situations: pieces of this problematic phenomenon in clause combining were observed independently by other researchers in unrelated contact situations, such as in Laz with Turkish contact (Demirok and Öztürk 2022) and in Indo-Aryan languages with Dravidian contact (Bayer 2001).

4 Textual sources, method, and statistics

The data used in this investigation come from the following modern and historical texts.

- 1) Dialect texts¹⁰
 - a) Kosovar Turkish: collected and published by Sulçevsi (2019: 192–261) representing the modern West Rumelian varieties.
 - b) North Rumelian
 - i) Dobruja Turkish: collected and published by Gülvodina (2018: 136–204), Güneş (2009: 123–195), Haliloğlu (2017: 163–231), Karaşinik (2011: 167–250), and Murtaza (2016: 60–319).
 - ii) Gagauz: selected from numerous sources and published by Özkan (2007: 100–178).
- 2) Historical texts
 - a) Pietro Ferraguto's *Grammatica turchesca* dated 1611 published by Bombaci (1940: 222–236)

⁹ Due to lack of space, I will not consider a synchronic account: the possibility that X-clauses may constitute a possible clause type in their own right and are not just an intermediate stage in syntactic change.

¹⁰ The material from Kosovo and north-eastern Bulgaria are transcripts of spontaneous oral productions of informants (e.g. personal accounts, accounts of local traditions). The Gagauz texts in Özkan (2007) are fairy tales in an orate register that has been edited.

- b) Jacob Papas’ letter from around 1484–1486 published by Brendemoen (1980: 228)
- c) Yusof’s letter also from around 1484–1486 and published by Brendemoen (1980: 230–231)

The frequency counts of SCs in KT cited in the text are based on the Balkan Turkic Corpus (Keskin et al. in preparation b). The KT texts in this corpus were culled from Sulçevsi (2019: 192–261). The survey of X-clauses in KT was based on all the texts in Sulçevsi (2019) and was done by means of the software #LancsBox (Brezina et al. 2018). For this survey, SCs introduced by *ne* were identified, as this is the most common SAE-type subordinator there (Stein 2016: 165–166) and were analyzed for the features in Table 1. The findings were then schematized as in Table 3. The same approach was used for the survey of X-clauses in NRT, with the difference that in this case the focus was on clauses introduced by *ani*, the most common SAE-type subordinator in that group (cf. Menz 1999: 67).¹¹ Ferraguto’s work was also analyzed using #LancsBox.

Two separate tools were used for the statistical tests carried out on the frequency counts: (i) Lancaster Stats Tools online that runs R code (Brezina 2018), and (ii) Real Statistics Resource Pack for Microsoft Excel (Zaiontz 2020). Percentages were calculated using MS Excel.

5 Patterns of X-clause in Peripheral Rumelian

As pointed out in Section 4, for my survey of X-clauses in PRT, I first identified the SCs that are introduced by the subordinators *ne* (in KT representing WRT) and *ani* (in NRT). As shown in Table 2, this sample totaled 326 SCs, out of which 91 fit neither the Turkic nor the IE models, i.e. were X-clauses.

Table 2: Place of X-clauses in PRT.

	All <i>ne/ani</i> -cl.	Relative freq. (per 10,000)	X-clauses	Relative freq. (per 10,000)	Ratios
KT	113	54.75	30	14.54	26.5%
NRT	213	14.19	61	4.06	28.6%
Tallies	326	19.09	91	5.33	27.9%

¹¹ The reason for collecting examples from two sets of clauses (viz. one set introduced by *ne* and the other by *ani*) was to have data from both dialect groups that make up PRT, since the connector *ne* is exclusive to WRT, while *ani* is only seen in NRT. Otherwise, both sets of clauses are typical SAE-type clauses.

X-clauses have widely differing frequencies in these two groups of PRT dialects, ultimately due to the differences in the frequencies of SAE-type clauses in each. These figures can be found in the relative frequency columns in Table 2. Despite these differences in frequency, the ratios of X-clauses among *ne/ani*-clauses in the two dialect groups are not too different from the average of 27.9% (see under ‘Ratios’). However, there may be large differences (not shown in Table 2) between the varieties that make up the two dialect groups. For instance, in NRT, while X-clauses constitute a mere 4.5% of *ani*-clauses in Gagauz, they are as high as 58.8% on average in Dobruja Turkish.^{12,13}

For the feature analysis done on the identified *ne/ani*-clauses based on the SC features in Table 1, I first jettisoned 27 headless relatives, as the question of where the relative clause is positioned with respect to the head noun (i.e. the clause position feature) is not applicable to headless relatives and creates ambiguities when sorting them into one of the X-clause subtypes we will presently see. The remaining 64 SCs could be grouped into 13 subtypes based on their properties, presenting a complex pattern of which we examine the main features below. During this exposition, I will be referring to Table 3 which shows the 13 subtypes and their properties in schematic form.¹⁴

5.1 The 13 subtypes and their distribution

The 13 subtypes of X-clause attested in PRT (plus the Turkic and IE-type SCs) are indicated in the top row of Table 3. In my sample, most subtypes (i.e. X1–X6, X8, X11,

¹² This may be either due to regional differences or the registers of the texts used as sources. As already pointed out, the Gagauz texts in Özkan (2007) are in an orate register that has been edited, while texts from Dobruja Turkish are spontaneous productions.

¹³ Another perspective that may be of relevance is the distribution of X-clauses, relative to Turkic and IE-type clauses, among argument, relative, and adverbial clauses. Here, the data from KT reveal an asymmetry. Relative and adverbial clauses have the highest percentage of X-clauses (25 and 19%, respectively) and are mostly of the IE type (68 and 70%, respectively). Argument clauses, by contrast, have minimal amounts of X-clauses (5%) and a more balanced distribution into Turkic versus IE-type clauses (40 vs 55%, respectively). I will not pursue this line of investigation further in this paper but see Section 12.1 for the distribution of argument, relative, and adverbial clauses into different subtypes of X-clause.

¹⁴ The meanings of the feature values that appear for the first time are as follows. [±free]: combination of suffixed and free subordinators, [±init]: combination of clause-initial and clause-final free subordinators (i.e. a circumclausal subordinator), [+init²]: double clause-initial subordinator, [•init]: clause-internal subordinator, [+•init]: combination of clause-initial and clause-internal subordinators.

Table 3: Subtypes and properties of X-clauses attested in PRT.

Clause types Properties	Turkic	X1 ⁿ	X2 ⁿ	X3 ⁿ	X4 ⁿ	X5 ⁿ	X6 ⁿ	X7 ^k	X8 ⁿ	X9 ^{n,k}	X10 ^{n,k}	X11 ⁿ	X12 ⁿ	X13 ^{n,k}	IE
Finiteness	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin	–fin
Clause position	–post	–post	–post	–post	–post	–post	–post	–post	–post	–post	–post	–post	–post	–post	–post
Subordinator type	–free	±free	±free	±free	±free	±free	±free	±free	±free	±free	±free	±free	±free	±free	±free
Subordinator position	–init	±init	±init ²	±init	±init	±init ²	±init	±init	±init ²	±init	±init	±init	±init	±init ²	±init
Degree of IE-ness	0%	25%	31%	38%	50%	56%	63%	63%	69%	75%	88%	88%	91%	94%	100%
Frequency	N/A	4	1	5*	1	1	7*	3	1	14*	5*	9*	1	12*	N/A

and X12) are exclusive to NRT (marked by an ^{nb}); only X7 is restricted to KT (^{dk}). The remaining three subtypes (i.e. X9, X10, and X13) are seen in both KT and NRT (^{nk}).

From this perspective, NRT could be seen as a hotbed of X-clauses, as 12 different subtypes are in use there as opposed to four in KT, even though X-clauses have a higher relative frequency in KT. Within NRT, Dobruja Turkish stands out, as X-clauses constitute 58.8% of SCs there on average, as already mentioned (in stark contrast to Gagauz where their ratio is just 4.5% and to KT where they reach 26.5%). Also, the Gagauz data contain only four subtypes of X-clause (i.e. X3, X9, X11, and X13; not indicated in Table 3), while Dobruja Turkish has 12 subtypes (i.e. all except X7).

5.2 Degree of Indo-Europeanness

The 13 subtypes shown in Table 3 are ranked horizontally based on an ‘Indo-Europeanness’ score that was calculated for each and is shown in the penultimate row. I reckoned this rough measure by first assigning a numerical value to each SC feature value. Each fully Turkic feature (i.e. [–fin, –post, –free, –init]) was assigned a numerical value of zero and each fully IE feature (i.e. [+fin, +post, +free, +init]) a value of one. The feature values [±free] and [±init] which are combinations of Turkic and IE values and [•init] which is between the two poles of Turkic versus IE were assigned a numerical value of 0.5. The remaining two feature values were more difficult to assign numerical values. [+init²] was assigned a value closer to IE (i.e. 0.75), given that it is almost fully IE. Finally, [+•init] was assigned the value of 0.625 by taking the median of 0.75 and 0.5. The score for each X-clause subtype was then tallied up, divided by the maximum possible score of four, and multiplied by 100.

What does a given subtype’s Indo-Europeanness score say about the subtype? This score simply expresses as a percentage value, how much a given subtype descriptively resembles an IE-type SC from the perspective of the four SC features. It implies nothing about where the subtype stands historically on the way from the Turkic to the IE model. In other words, it is not intended to reveal or represent a grammaticalization cline or any such diachronic tendency.

5.3 Frequencies

As the bottom row of Table 3 shows, six subtypes have above-average absolute frequencies (marked with “*”; average = 4.9) and account for 81.3% of all X-clauses in

PRT (52 out of 64).¹⁵ Of these six subtypes, three (i.e. X9, X10, and X13) are seen in both WRT and NRT, and three (i.e. X3, X6, and X11) are restricted to NRT. In contrast, five subtypes (i.e. X2, X4, X5, X8, and X12) are attested only once each, all in NRT. Their single occurrence may cast doubt on their existence as subtypes, however at least some of the below-average frequencies across the table (as well as the above-average frequencies) are sure to increase with a larger sample. For instance, subtype X4 is recorded by Menz (1999: 105–107) as a common SC type in Gagauz, despite its single occurrence in my sample. This diversity of subtypes and low frequencies makes for a very fragmented picture. We will return to subtype frequencies and Indo-Europeanness scores in Section 11 where we exploit their interplay to derive an argument in support of the diachronic explanatory framework adopted in Sections 7–11.

5.4 Subordinate clause features

We now turn to the structural features of X-clauses (i.e. finiteness, clause position, and subordinator type and position). These are given in rows two to five of Table 3.

Finiteness and clause position are two neatly binary features, and both show a slight overall tendency to be of the IE-type: of all the subtypes, 54% (seven out of 13) are either finite or postpositive as opposed to the remaining 46% (six out of 13) that are nonfinite or prepositive. Subordinator type and position present an increasingly more complex pattern.

Regarding subordinator type, in 46% of subtypes, suffixal and free subordinative elements are used in conjunction, which is only seen in NRT. This use of subordinators in combination is seen in many standard Turkic adverbial clauses as well (cf. [1b]), so in a sense there is little that is out of the ordinary in this fact. However, the difference in the case of X-clauses is that this feature combination is seen in relatives and argument clauses as well. Also, the free subordinators in Turkic adverbial clauses are always clause-final, whereas this sample of X-clauses reveals several subordinators appearing in a non-final position, which is atypical.

With respect to subordinator position, there are five possibilities. Double clause-initial subordinators are the most common type by 30.8% in terms of how many X-clause subtypes they occur in. They are mostly attested in NRT (two examples from KT vs 14 examples from NRT). Next, by 23.1% are circumclausal subordinators, which are exclusive to NRT (with 15 occurrences in the source texts). Clause-initial subordinators, the only unmarked kind of subordinator in Table 3, are equally common

¹⁵ There might be a power law governing this distribution given that the top 39% of subtypes in terms of frequency account for 81% all X-clauses in PRT (Stavros Skopeteas, p.c.).

and are seen much more in NRT (26 occurrences) than in KT (three occurrences). Free clause-internal subordinators, seen in 15.4% of X-clause subtypes, are more common in KT than in NRT (seven vs three occurrences, respectively).¹⁶ Finally, the combination of a clause-initial and a clause-internal connector is seen once in NRT.

5.5 Statistical analyses

The apparent complexity in Table 3 obscures the equal distribution of SC feature values over the 13 subtypes. The differences between the competing feature values for each SC feature in terms of how many X-clause subtypes they occur in were found to be statistically nonsignificant. Sign tests comparing [–fin] versus [+fin], [–post] versus [+post], and [±free] versus [+free] all returned $p = 1$. The differences between the five subordinator positions were also nonsignificant according to a chi-squared goodness of fit test done manually ($\chi^2(4) = 2$, $p = 0.74$) and an Anderson-Darling goodness of fit test ($p = 0.51$). Finally, all four features considered, there was no statistically significant difference between IE-type and non-IE-type features (Mann-Whitney U (or Wilcoxon rank-sum) test: $U/W = 7$; $p = 0.88$). And there was no difference between the four features in this regard (Pearson's Chi-squared test: $\chi^2(3) = 3.71$; $p = 0.29$).

6 X-clauses from a typological perspective

Given the significant percentage distribution of X-clauses in PRT, questions arise as to whether they are attested in other languages, where they fit typologically, etc. One cross-linguistic survey that helps place X-clauses in a typological context to some extent is Dryer's (2013) study on (adverbial) clause types as part of WALS (Dryer and Haspelmath 2013), which covers a range of SC types comparable in structure to those examined here. First, Dryer's data show that at least some of the atypical features and feature combinations observed in X-clauses are attested in languages outside the Turkic family and the Balkan sprachbund (and presumably not necessarily in a contact setting). This suggests that they are a more common feature of language in general and not an isolated oddity of PRT (see also the works cited at the end of Section 3). Second, according to Dryer's data, X-clause-like structures are not seen in other members of the Turkic family than PRT or in other languages of the Balkan sprachbund. This is suggestive for my diachronic approach (Sections 7–11): X-clauses

¹⁶ When we take headless relatives into account as well, this asymmetry becomes much more marked: 25 versus three examples.

are neither a native Turkic feature nor a syntactic borrowing from the majority languages of the Balkans but emerge from the dynamics of syntactic change, probably through language contact.

According to Dryer (2013), the pattern that I refer to here as the IE model is typologically the most frequent one out of a total of five possibilities. It is used as the dominant model in 60.4% of the languages in the WALS sample (398 out of 659), meaning that the other four patterns are fairly marginal cross-linguistically. In 14% of the languages of the sample (93 out of 659), several patterns co-occur. In some languages among these, such as Miya (Chadic, Nigeria), one of the co-occurring patterns will be dominant, as is the case in KT, for instance, where the IE-type clauses make up 66.7% of all SCs (see Section 1.2).

In languages with more than one adverbial clause type, a combination of types is sometimes used in the same clause, which is a characteristic feature of X-clauses. For instance, in example (11) from Bwe Karen (Tibeto-Burman, Myanmar), a clause-initial/internal (Dryer's first/third type) and a clause-final subordinator (Dryer's second type) are used in combination.

- (11) *Nə-dé* *ɔ* *khalé*
 2_{SG}-if be.at if
 'If you stay.'
 (Henderson 1997: 78 via Dryer 2013)

This pattern can be compared to the combination of clause-initial and clause-final connectors in subtypes X1, X4, and X11 (see example [7]), and the combination of clause-internal and clause-initial connectors in subtype X12.

Another combination, seen in Majang (Surmic, Ethiopia), involves a free subordinative element at the beginning of the clause, a suffix on the verb, and a clitic at the end of the clause, as in example (12):

- (12) *Agutucee=ko* *tolay* *doko-du* *ogol=ku*
 because=_{PST} Tolay bring-reason mead=reason
 'Because Tolay brought mead.'
 (Unseth 1989: 117 via Dryer 2013)

This is similar to subtypes X3 and X6 (see example [5]) with clause-initial connectors and nonfinite predicates which bear a subordinative suffix.

Another clause type included in the WALS sample comparable to X-clauses is one where the adverbial subordinator appears inside the clause, exactly as in subtypes X7 and X10. These clauses are seen as the dominant pattern in just 1% (eight out of 659) of the languages in the sample. In (13) is an illustrative example from Nkore-Kiga (Bantu, Uganda):

- (13) *Wa-kami obu y-aa-tuuriza enjojo*
 Mr.-rabbit when 3SG-PST-challenge elephant
 ‘When Brother Rabbit challenged the elephant.’
 (Taylor 1985: 27)

Here, the subordinator *obu* ‘when’ is positioned between the subject and the finite verb of the adverbial clause (see also Taylor 1985: 169). This example can be directly compared to the example in (14) of the subtype X7 from KT:

- (14) *Sora [bayram açan col-i] kolk-ay-sin sabale.*
 then Eid when come-PROG.3SG get.up-PROG-2SG by.morning
 ‘Then, when the Eid comes, you get up by morning.’
 (Pristina, KO; Sulçevsi 2019: 200)

Here, as with (13), the subordinator (viz. *açan* ‘when’) is positioned between the subject and the finite verb of the adverbial clause in brackets (see also [8] that exemplifies a relative clause of the subtype X10 again from KT).

7 The transient behavior hypothesis

Beginning with this section, I move on to the solution that I propose in this paper for the X-clauses problem and the arguments in favor of this solution. As the main component of my proposal, I hypothesize that X-clauses are manifestations of ‘transient behavior’ in the shift from Turkic to SAE-type subordination in PRT syntax.¹⁷ Transient behavior, a concept borrowed from systems science, can be described as follows. When a system Σ in a steady state S_1 encounters an impact from the outside (i.e. a disturbance), Σ may adapt by shifting to a new steady state S_2 , if the disturbance is long-lasting, which implies a new normal condition. During the shift from S_1 to S_2 , Σ is said to exhibit transient behavior, an unsteady transition that typically involves oscillations in the values of Σ ’s parameters (e.g. temperature, voltage) (Marshall 1978: 73–74; Mobus and Kalton 2014: 241–246, 375–385). This may be represented as in Figure 1.

We can translate this description into the present context as follows. As Σ we have the subordination system of PRT, and S_1 and S_2 are Turkic and SAE-type subordination, respectively. The parameters of the subordination system,

¹⁷ The idea that the atypical SCs in RT are connected to language change is also found in Kılıç (2018: 34–35). The author observes, in passing, that while the general tendency for finite relative clauses in RT is to be postpositive, some are prepositive. This unexpected positioning for a finite relative clause is due to these exceptional clauses being “a kind of transitional stage structure” that have preserved the dependent-head order typical of Turkish.

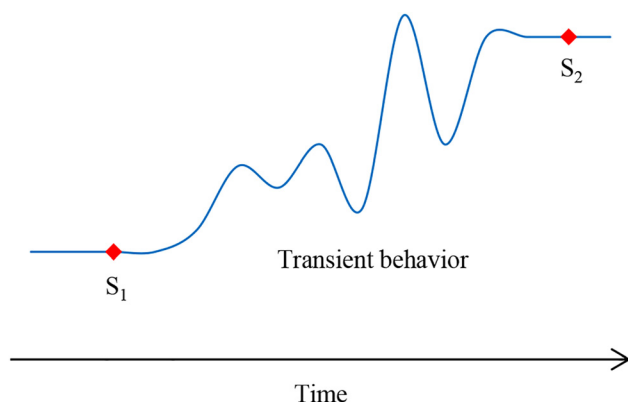


Figure 1: Transient behavior.

according to my formulation in Table 1, are the SC features of finiteness, clause position, and subordinator type and position, each taking a minimum of two values. Now, in the case of PRT, the majority languages of the Balkan sprachbund constitute a long-lasting disturbance that implies a new normal condition which is the SAE-type subordination seen in those majority languages. As the PRT subordination system shifts to SAE-type subordination in adaptation to long-term language contact, it shows transient behavior in the form of oscillations in the values of the four SC features. An oscillation (in feature value) can be defined as uncertainty and variation (of feature value) for now, and a more detailed discussion of the term can be found in Section 8. As these uncertain and variable SC feature values combine, they bring about interference patterns of sorts, which are the several kinds of X-clause attested in PRT. The preceding remarks are summed up in the statement in (15).

(15) *The Transient Behavior Hypothesis*

X-clauses in Peripheral Rumelian Turkic are an outcome of the transient behavior exhibited by the PRT subordination system in its shift from the Turkic towards the SAE subordination model.¹⁸ This transient behavior involves oscillations in the values of the parameters of the PRT subordination system, and the various combinations of parameter values yield the various X-clause subtypes.

¹⁸ Note that the shift of PRT subordination away from Persian-type towards SAE-type subordinate clauses also appears to play a modest role in the generation of X-clauses. That effect will be touched upon in Section 10.

The transient behavior hypothesis could be contrasted with an alternative that may be called the ‘smooth transition hypothesis’. Under this alternative hypothesis, PRT subordination system would shift from Turkic to SAE-type subordination without showing any transient behavior in the form of oscillations. In more technical terms, while transient behavior follows a sine wave pattern, as shown in Figure 1, smooth transition would follow a sigmoid wave pattern (also called an ‘S-curve’). Presumably, this scenario would have different types of X-clause ordered chronologically, involving a stepwise transformation of Turkic SCs to SAE-type SCs, and the earlier types would be phased out in favor of the later ones, with few overlaps in their lifecycles. This stepwise transformation would presumably rely on a diachronic interpretation of the horizontal ordering of subtypes according to their degree of Indo-Europeanness in Table 3, and it would be a point in its favor if SCs did follow that trajectory (i.e. first changing into X1, then into X2 and so on) or a comparable one. Yet, the fact that the 13 subtypes of that hypothetical trajectory are attested in the same time period, rather than previous subtypes being phased out in favor of later ones, argues against this hypothesis. For this reason, I will not develop this second scenario in any detail.

In the light of my unified treatment of X-clauses, the following question may arise: are we inferring a single trajectory of change from what are essentially different corpora (i.e. texts from different varieties and historical periods [see Section 4]), which may, in fact, be reflecting different trajectories?

One answer to this question is that, as a corollary of the transient behavior hypothesis, there is no single well-defined trajectory as such. First, in the overall X-clause pattern, all SC feature values are equally distributed over the 13 subtypes, as we saw in Section 5.5. Furthermore, looking at this pattern in more depth, we will see in Section 8.2 that there are no correlations between feature values in Table 3. In other words, the PRT subordination system is not biased towards any class of feature values (Turkic, IE, or otherwise) or combinations of feature values when producing the 13 subtypes. In short, there appears not to be an X-clause *pattern* as such – it is all random. A consequence of this lack of trajectory or randomness would seem to be that not all subtypes need be attested in all varieties. In fact, any convergences in subtypes between varieties are likely just lucky coincidences. The only expected convergence between different varieties under the transient behavior hypothesis, then, is that they all go through oscillatory transitions.

The randomness of X-clause feature values and their combinations should, however, be distinguished from the pattern produced by the frequencies of subtypes, i.e. the bottom row of Table 3. The frequency distributions of the 13 subtypes point to a frequency drift that favors subtypes with higher Indo-Europeanness scores, as discussed in Section 11. The analysis in that section does carry a risk of generalizing across corpora that should perhaps not be compared. However, given that those

corpora are all from varieties that are changing in the direction of SAE as observed by numerous independent studies (see e.g. Keskin 2023a, 2023b and the references therein), a frequency drift favoring the IE template is not unwarranted.

As for convergences between different historical periods, this is probably different than convergences between regional varieties. Convergence through time could be expected, as later historical stages of development of a given variety would normally carry at least some features of its previous historical stages. And therein lies again a risk of generalizing across incompatible corpora – this time greater than in the case of frequency drift.

8 Oscillations

As summed up in (15), I hypothesize X-clauses to be an outcome of the transient behavior exhibited by the PRT subordination system in its shift from the Turkic towards the SAE subordination model. Put differently, X-clauses are the various combinations of the SC feature values of the PRT subordination system, and these feature values show oscillations. But how should the term oscillation be understood in the present context? Take, for instance, the SC feature of clause position in Tables 1 and 3 and its two values [+post] and [–post]. We do not observe that a complement clause, say, is to the left of the matrix verb at one point in time and to its right at the next and then once more to its left, unlike what one would presumably expect in view of the literal meaning of the term oscillation.¹⁹ That being the case, we need to look for plausible approximations of literal oscillations or potential manifestations in the SC feature values of X-clauses of a more abstract notion of oscillation. Below, I first propose various alternatives, then discuss an implication of the idea that SC feature values of X-clauses oscillate, which corroborates this conceptualization.

8.1 Various interpretations of oscillation

One interpretation of oscillation in the domain of X-clauses could be as follows. There are several sets of clause types in Table 3 that differ in a single feature value (i.e. are minimal sets), mostly subordinator position. For instance, subtype X9 and IE-type subordination differ only in clause position, the former being prepositive and the latter postpositive. In other words, it is as if a finite SC with a free clause-initial

¹⁹ More accurately, it may well be that a given SC feature value *does* change from one point in time to the next and then back to its previous value, however the timespan that would be involved in such a scenario is probably rather large, i.e. decades or centuries (see e.g. Heine and Kuteva 2005), and no data is available yet to show that.

subordinator (see the relevant feature values of X9 and IE-type subordination in the table) were appearing on either side of the head, i.e. oscillating in its syntactic position. This seems to be the most literal interpretation of oscillation. The other minimal sets found in Table 3 are listed in Table 4.

Second, the values of all four SC features are distributed equally over the 13 subtypes of X-clause, as I pointed out in Section 5.5. In other words, as a looser interpretation of oscillation, SC features appear to variously acquire any value available for a given feature in a balanced manner, i.e. oscillate between those values.

Third, the value of the subordinator position feature in subtypes X7 and X10 has a value in between the Turkic and the IE values (i.e. [•init] for clause-internal connectors), as if captured moving between those two extreme values. Subordinator type feature could also be said to display such behavior if cliticized subordinators (i.e. those in between free vs suffixed subordinators) were discovered in a larger sample. In fact, we will see one potential instance in (17b).

Finally, the value of a given feature in some subtypes can include both the Turkic and the IE values simultaneously (e.g. [±init] in the case of circumclausal connectors of X1, X4, and X11), appearing like no particular value has been set for that feature such that its value can be realized at both extremes at the same time.

8.2 Independence of feature values

The possibility that the values of the four component features of the PRT subordination system all show oscillations as part of its transient behavior suggests another possible feature of the transient phase. The values of the four component features are probably independent of each other in the sense that there does not seem to be a theoretical reason for clause position to be a function of finiteness value, for instance. Thus, as feature values oscillate between the two extremes (i.e. Turkic vs IE) until

Table 4: Minimal sets of X-clause.

Minimal set	Differing in
{X1, X2, X3}	Subordinator position
{X10, X11, X12, X13, IE}	Subordinator position
{X4, X5, X6}	Subordinator position
{X7, X8, X9}	Subordinator position
{X1, X4}	Clause position
{X7, X10}	Clause position

they reach the new steady state S_2 , it would be perfectly possible, perhaps even natural for them to do so independently of each other in an uncoordinated fashion at least for a period of time until they reach S_2 . That would indeed seem to be the case in the pattern presented in Table 3, as explained below.

According to a Spearman’s correlation test, there were no correlations between the feature values in Table 3 except between finiteness and subordinator type. For this test, I first adapted the method used for calculating the Indo-Europeanness score from Table 3, by assigning ranks to the individual feature values instead of the numerical values used previously, since Spearman’s correlation is a non-parametric test that works with ranks: IE feature values = 7; [+init²] = 6; [+•init] = 5; [±free, •init, ±init] = 4; Turkic feature values = 1. This ranking scheme might appear lopsided in the sense that ranks 2 and 3 are not used. This is because I wanted to place the feature values [±free, •init, ±init] in the middle of the Turkic and IE extremes, and there were no features that fit ranks 2 and 3. Also, I judged [+init²] to be closer to IE than [+•init], as clause initial complex subordinators (e.g. *as if*) are frequently used in SAE. Using this approach, I transformed rows two to five in Table 3 into a table of feature rankings. Then, I ran Spearman’s correlations on all pairs of feature rankings for all 13 subtypes and the Turkic and IE-type clauses: finiteness versus clause position, finiteness versus subordinator type, etc. The results are given in Table 5.

First, finiteness and subordinator type were strongly correlated. However, there was a weak correlation between finiteness and clause position that was nonsignificant, and a slightly stronger correlation between clause position and subordinator type that also failed to reach significance.

More explicitly, (i-a) finite clauses contain free subordinators and (ii-a) tend to be postpositive, and (iii-a) postpositive clauses tend to contain free subordinators. Conversely, (i-b) nonfinite clauses contain both suffixes and free subordinators and (ii-b) tend to be prepositive, and (iii-b) prepositive clauses tend to contain both suffixes and free subordinators. These tendencies hint at the links between the

Table 5: Spearman’s correlations between subordinate clause features.

	Finiteness	Clause position	Subordinator type	Subordinator position
Finiteness		$r_s = 0.196, p = 0.48$	$r_s = 0.976, p < 0.001$	$r_s = 0.032, p = 0.9$
Clause position	$r_s = 0.196, p = 0.48$		$r_s = 0.244, p = 0.38$	$r = 0.032, p = 0.9$
Subordinator type	$r_s = 0.976, p < 0.001$	$r_s = 0.244, p = 0.38$		$r_s = 0.132, p = 0.64$
Subordinator position	$r_s = 0.032, p = 0.9$	$r_s = 0.032, p = 0.9$	$r_s = 0.132, p = 0.64$	

features values for IE and Turkic subordination, respectively. There were only negligible, nonsignificant correlations between the other pairs of features.

9 X-clauses in Early Peripheral Rumelian

I now turn to historical data on X-clauses. Here, a particular category of text is of special importance: the so-called ‘transcription texts’ which are texts composed in Turkish during the Ottoman period by westerners and non-Muslim subjects of the Ottoman Empire in non-Ottoman scripts. Before I present any data from transcription texts, however, the question of how reliable they are should be addressed.

Whether transcription texts are reliable sources for historical linguistic studies of Turkish is a legitimate question given the following background (see e.g. Hazai 1990: 64–67; Stein 2016: 161–162). Turkology of the 1960s and the beginning of the 1970s was the scene of a debate focusing on the question of whether the Turkish observed in transcription texts (particularly Georgievits [1544] published by Heffening [1942] and Illésházy [1668] published by Németh [1970]) could be taken to consistently reflect any particular Turkish variety. Németh’s (1968, 1970) view was that these texts were representative of the Balkan dialects of Turkish. The opposing position was Kissling’s (1968) claim that the texts simply contained idiosyncratic linguistic mixtures and reflected an at best imperfectly learned Turkish superimposed on a Balkan substrate. The debate was settled in favor of the former position, and a body of scholarly work emerged that makes use of transcription texts as consistent sources of data for historical linguistic studies of Turkish (see e.g. Csató et al. 2016).

To return, now, to the main line of discussion, three transcription texts from the 15th and 17th centuries, which I examine below, contain expressions showing patterns which characterize X-clauses and provide evidence that X-clauses are not a new occurrence in PRT. By the late 15th century, when we first see X-clauses in historical texts along with SAE-type SCs, varieties of Turkish had been spoken in the Balkan sprachbund for about a hundred years (Artun 2013: xiv, xix; Johanson 2021: 132–133) and contact-induced syntactic changes seem to have already been underway (see e.g. Keskin 2022).

The first transcription text that I analyze is Pietro Ferraguto’s work in Latin script titled *Grammatica turchesca* (henceforth ‘*Grammatica*’) dated 1611 (published by Bombaci [1940]). It contains a dialogue titled *Dialogo tra un Turco et un Christiano* that is 1692 words long in Turkish, exemplifying the use of Turkish, from which my examples come (Stein 2014). The *Grammatica* is a rather clear example of RT, more precisely of Early WRT, as shown by Keskin (2023a, 2023b) and Stein (2016). This text

is also by far the richer historical source of examples of X-clauses, as well as the other SC types attested in WRT, compared to the other two texts.

The other two texts are two short letters (i) by an Italian priest called Jacob Papas (lit. ‘Jacob the Priest’), written in Latin script and sent to the Ottoman Sultan Bayezid II sometime in 1484–1486 (91 words, published by Brendemoen [1980]), and (ii) by a certain Yusof, written in Greek script and sent to Cem Sultan, a claimant to the Ottoman throne again sometime in 1484–1486 (131 words, also published by Brendemoen [1980]). Neither Jacob Papas’ nor Yusof’s letters have been treated as examples of RT by the Turkological literature – probably because they have not drawn much attention, however, they could be included in the same sphere of Turkish–SAE contacts along with my other sources, given their background. Indeed, Jacob Papas’ text even contains what appears to be an SAE-type SC introduced by *neia* (cf. *ne* in WRT) as connector – occurring twice in the text, its second occurrence as a proclitic in an X-clause (see [17b]) – which should justify its treatment as a PRT text: *bil-mis ol-(a)sun* [*neia* *men cul-un Jacob fran papas*] (know-PRF be-2SG.OPT CONN 1SG servant-2SG.POSS Jacob European priest) ‘May you be informed that I am Jacob Papas, your servant’. I will make use of these two texts as secondary sources as they contain a small number of X-clauses.

The *Grammatica* presents a rich and varied picture. Even then, X-clauses in this text present a unified class to some degree in that they are introduced exclusively by the subordinator *sciú* except for one example introduced by *neredé* ‘where’.²⁰ As pointed out in Section 3, this justifies to a certain degree the treatment of X-clauses as a class of SC even though the data from modern PRT do not point to this. In (16) I give one example each from the three subtypes of X-clause attested in the text and the single example with *neredé*.^{21,22,23}

- (16) a. *Sciú* *ollará* *uerdicnís* *rísg*
 [*şu* *ollar-a* *ver-dig-iniz*] *rızk*
 CONN 3PL-DAT give-OREL-2PL.POSS livelihood
 ‘The livelihood that you give them.’ (X3)
 (Bombaci 1940: 230)

²⁰ Note, though, that not all *sciú*-clauses are X-clauses. Five out of the 20 examples introduced by *sciú* that the text contains are regular SAE clauses.

²¹ I will not give examples of the SAE-type SCs in the *Grammatica* due to lack of space. They are essentially the same as the SAE-type SCs in KT (i.e. [3a]).

²² The text underneath the original is my normalization of the authors’ orthography. I use the Modern Standard Turkish alphabet as much as possible.

²³ In some of these examples, *sciú* may look like a medial demonstrative rather than a connector, especially when it precedes a relative clause as in (16a). This is because the examples are taken out of their context. In most of these seemingly unclear cases the demonstrative interpretation is ruled out by the context.

- b. *Júg hólmassun sciú sanghá [...] surdigumdén.*
Yük ol-ma-sun [şu saña [...] sur-diğ-um-den].
 burden be-NEG-OPT.3SG CONN 2SG.DAT ask-FNOM-1SG.POSS-ABL
 ‘I would not want to overburden you by asking you.’ (X6)
 (Bombaci 1940: 222)
- c. *Sciú bén doumísciúm uelaiét*
[şu ben doğ-miş-um] vilayet
 CONN 1SG be.born-PRF-1SG province
 ‘The province that I was born in.’ (X9)
 (Bombaci 1940: 224)
- d. *Neredé doumíscsén uelaiét*
[neredé doğ-miş-sen] vilayet
 CONN be.born-PRF-2SG province
 ‘The province where I was born.’ (X9)
 (Bombaci 1940: 223)

The SC in (16a) is prepositive and has a nonfinite verb form (i.e. an object relative) as predicate. The object relative suffix also functions as a subordinative element. Additionally, the clause contains a free clause-initial connector. Thus, it is of the subtype X3. The SC in (16b) differs in its position from the previous one and is consequently of the subtype X6. The SCs in (16c) and (16d) differ from the first example in the finiteness of their predicates. They are, then, of the subtype X9. Additionally, (16d) contains a subordinator different from the others (i.e. *neredé*).

As for Jacob Papas’ and Yusof’s letters, they contain a total of just three examples of X-clauses (in [17] and [18], respectively), not unusual given how short these texts are. These examples fit two patterns we have encountered in Table 3, namely subtypes X10 and X12.

- (17) a. *Jacob fran papas savoya tarfinda chim gondarmis edin*
Jacob Frenk papaz [Savoya tarafında kim gönder-miş i-di-n]
 J. European priest S. towards CONN send-PRF COP-PST-2SG
 ‘Jacob the European priest whom you had sent towards Savoy.’ (X10)
- b. *Paulu coli ila chim besirgan nedi*
Paolo da Colle ile [kim bezirgan n-i-di]
 P. with CONN merchant CONN-COP-PST.3SG
 ‘With Paolo da Colle who was a merchant.’ (X12)
 (Brendemoen 1980: 228)

- (18) *Venedi gördüm gaurum bayramna ta kim gemi bulam.*
Venediğ-i gör-dü-m [gâvur-un bayram-ın-a ta kim gemi bul-am].
 V.-ACC see-PST-1SG infidel-GEN holiday-3SG.POSS-DAT CONN vessel find-OPT.1SG
 ‘I have been to Venice in order that I may find a vessel till the infidel’s holiday.’ (X10)
 (Brendemoen 1980: 230)

In all three examples, the SCs are finite, postpositive, and contain free subordinative elements. They differ, however, in the positions of their subordinators. In examples (17a) and (18), the subordinators are clause-internal, which, in conjunction with the other SC features, makes these clauses instances of the subtype X10. In (17b), by contrast, we seem to have a combination of clause-initial and clause-internal connectors, the trademark of subtype X12. Here, I interpret the prothetic *n-* on the copula as a proclitic form of the connector *ne* typical of WRT, occurring as *neia* earlier in the text.²⁴

The immediately preceding descriptions are summarized in Table 6 with some additional information.

To sum up, five X-clause subtypes seen in Modern PRT can already be identified in historical texts, which amounts to about 38.5% of the subtypes attested today.²⁵ All of these subtypes have survived to be joined by a further eight.

10 A comparison of the *Grammatica* with Modern Kosovar Turkish

With the historical perspective adopted in the previous section, we have seen that X-clauses have been present in PRT syntax since at least the 15th century, that their general properties have remained consistent throughout that period, and that they have diversified into further subtypes up until the modern era. We now change the angle of this diachronic outlook and focus on the percentage distributions of different clause types (i.e. Turkic, Persian-type, SAE-type, and X-clauses) in two different historical periods, namely Early WRT as reflected in the *Grammatica* and Modern WRT as reflected in KT, shown in Table 7. This comparison suggests a scenario consistent with X-clauses being products of the diachronic shift from Turkic to SAE-type subordination in PRT syntax.

²⁴ This prothetic *n-* could, thus, be taken as a cliticized subordinator which would be an intermediate value between free and suffixed subordinators, as speculated in Section 8.1.

²⁵ The *Grammatica* contains a total of 18 X-clauses, seven of which are headless relatives. As before, they have been excluded from this analysis.

Table 6: Subtypes of X-clause attested in Ferraguto's, Jacob Papas', and Yusof's texts.

Clause types Properties	Turkic	X3 ⁿ	X6 ⁿ	X9 ^{n,k}	X10 ^{n,k}	X12 ⁿ	IE
Finiteness	–fin	–fin	–fin	+fin	+fin	+fin	+fin
Clause position	–post	–post	+post	–post	+post	+post	+post
Subordinator type	–free	±free	±free	+free	+free	+free	+free
Subordinator position	–init	+init	+init	+init	+init	+init	+init
Degree of IE-ness	0%	38%	63%	75%	88%	91%	100%
Frequency	N/A	6	2	3	2	1	N/A

Table 7: Comparison of the distribution of clause types in Early and Modern WRT.

Clause type Period	Turkic	X-clause	IE-type		
			Persian-type	SAE-type	IE total
17th c. (<i>Grammatica</i>)	25.3%	12%	50.7%	12%	62.7%
21st c. (Modern KT)	14.8%	18.5%	28.7%	38%	66.7%

According to these data, Turkish SCs in WRT decreased by 10.5% from the 17th century till the 21st. By contrast, there was a mere 4% increase in this period in the ratio of IE-type SCs (cf. ‘IE total’). Within the class of IE-type SCs, however, a major reshuffling took place: there was a 26% increase in the ratio of SAE-type clauses and a 22% decrease in Persian-type clauses. This is a clear picture of the shift in favor of SAE-type clauses and at the expense of the Turkish and Persian-type SCs which brought about the present-day distribution of clause types, as mentioned in Section 1. Finally, and most importantly for our purposes, X-clauses increased by 6.5%. What can this 6.5% increase be attributed to? Is it due to the Turkish-to-SAE shift, as I have proposed till now? What is the contribution of the Persian-to-SAE shift, if any?

We can probably safely speculate that the decrease in Persian-type clauses almost directly translated into the increase in SAE clauses, i.e. with little impact on the ratio of X-clauses, as the Persian-to-SAE shift presumably only involves replacing Persian subordinative elements with newly created SAE-type ones. Also, recall that X-clauses are mostly mixtures of Turkish and IE SC feature values which means that their structures are largely inconsistent with the Persian-to-SAE route due to the presence of Turkish features values. In other words, it is unlikely that the Persian-to-SAE shift contributed to the 6.5% increase in X-clauses to any significant degree. There is, however, evidence of a modest influence of this particular shift on the ratio

of X-clauses. My sample of X-clauses includes seven examples exclusively from NRT that contain the double clause-initial subordinator *ani ki* whose second member is the Persian connector *ki* (see example [6]). These examples possibly reflect a pre-terminal stage of the process of replacing the Persian subordinative element and five of them belong to subtype X13, a subtype with no Turkic features. They constitute 13.5% of the X-clauses in NRT or 10.9% of all X-clauses in my sample.

Now, the WRT data (from both the *Grammatica* and Modern KT) contain no comparable examples and show a notable presence of Turkic feature values in X-clauses, which makes the contribution of the Persian-to-SAE shift to X-clauses improbable. Still, assuming that this influence was also present in WRT but that it could not be detected due to a sampling issue, suppose that 13.5% of the 6.5% increase in X-clauses in WRT (i.e. 0.9%) can be attributed to the Persian-to-SAE shift. The remaining approximately 5.6% increase in X-clauses would then be due to the Turkic-to-SAE shift. In other words, that much of the decrease in Turkic SCs would have gone to X-clauses. The remaining 4.9% decrease in Turkic would have fed into SAE-type SCs without us observing its effects on the percentage distribution of X-clauses, with the remaining 21.1% of the increase in SAE coming from the decrease in Persian-type clauses.

Connecting the observations above into a thread from Turkic to X-clauses to SAE, we could conclude that the shift in the WRT subordination system from Turkic to SAE-type clauses was facilitated by X-clauses. During this process, WRT grammar funneled clauses from the former to the latter through X-clauses, metaphorically speaking, and continues to do so in the present day.²⁶

11 Frequency drift

Observations on the interplay of the Indo-Europeanness scores and frequencies of subtypes, in conjunction with a comparison of Early and Modern PRT data, afford an additional argument for the view that X-clauses are a product of diachronic change. In Section 5.2 I pointed out that the Indo-Europeanness score is not intended as a potential indicator for diachronic change, however in conjunction with subtype frequency it does have implications in that regard.

²⁶ Returning to the discussion in fn. 7, further reasons for distinguishing Persian-type and SAE-type SCs must have become clear at this point. The distinction is central to recognizing the two points of origin for the shift to SAE-type subordination (viz. Turkic and Persian-type SCs). Also, distinguishing Persian-type and SAE-type SCs is crucial for being able to identify the redistribution in the IE class of SC from the 17th to the 21st centuries.

First, in Modern PRT there seems to be a relationship between the Indo-Europeanness score and the frequency of a subtype, since frequencies tend to increase in tandem with Indo-Europeanness scores in Table 3 as we move away from the Turkic towards the IE column. This is shown graphically in Figure 2.

A one-tailed Pearson's correlation test showed that the medium-to-large positive correlation between the Indo-Europeanness scores of subtypes and their frequencies was marginally nonsignificant ($r = 0.434$, $p = 0.07$), which is likely to become significant with a larger sample. These observations suggest a frequency drift (cf. e.g. Laitinen 2012; Leech et al. 2009: 270) towards the IE template, towards SAE-type subordination to be more precise. When we now go back to the Early PRT data in Table 6, we see that there is a very strong *negative* correlation between Indo-Europeanness scores and frequencies, as shown by a one-tailed Pearson's correlation test ($r = -0.891$, $p = 0.021$). This contrast between Early and Modern PRT data is consistent with (and perhaps even expected in the light of) the fact that the historical texts reflect PRT in the early phase of the shift towards SAE-type subordination. As SAE-type subordinate clauses have yet to gain preponderance at that early stage, we catch a glimpse of X-clause subtypes before their frequency drift towards SAE-type subordination has become discernable.

It should be pointed out that this drift is not predetermined in any way. Now, as observed in Sections 7 and 8, the pattern created by X-clause feature values and their combinations in Table 3 appears to be random. When this randomness is brought together with the frequency drift pattern, their synthesis points to the possibility that SC types in the contact situation that PRT has been in were subject to some sort of Darwinian process as follows (cf. Croft 2000). When Turkish came in contact with SAE

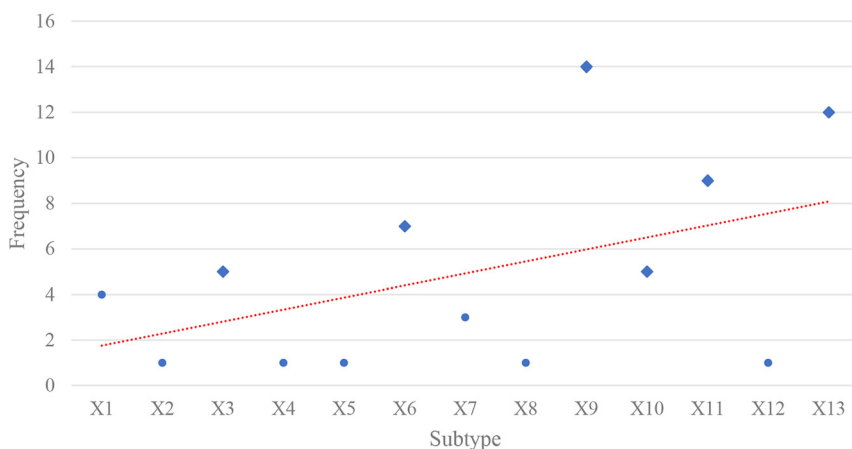


Figure 2: Frequency drift in X-clauses.

languages, several new SC types were generated randomly. Some of these new SC types were selected by the linguistic environment of the Balkans and grew in frequency, while others were not and remained at low frequencies or shrank in frequency and perhaps disappeared. In other words, there were two complementary and mutually independent processes at play through which the linguistic system evolved: one, SC feature combination (that produced subtypes), and two, subtype selection (and their resultant propagation). We can refer to this proposal as the ‘selectionist hypothesis’.

As a final side note, the frequency fluctuations in Figure 2 from one subtype to the next would argue against the smooth transition hypothesis or at least interpreting this ordering of subtypes in its favor. If this arrangement of subtypes also reflected a diachronic sequence, one would expect earlier subtypes to consistently have lower frequencies than later subtypes as part of the smooth transition to SAE-type subordination and such fluctuations would be ruled out.

12 Contact theoretic and sociolinguistic perspectives

In this section, I move away from the approach to X-clauses taken in the preceding sections that emphasized the formal aspects of X-clauses and take up some psycho- and sociolinguistic issues surrounding the shift to SAE-type subordination in PRT and the attendant emergence of X-clauses. In Section 12.1, I propose a broader contact theoretic account as a context within which the shift to SAE-type subordination and X-clauses can be viewed. Section 12.2 describes various sociolinguistic factors and tests their potential effects on the use of X-clauses by means of two logistic regression models.

12.1 Birth of Rumelian Turkic

Through which contact-related processes did RT shift to SAE-type subordination? Or more generally and informally, how did RT come to be the way it is? The short answer is that RT (particularly the PRT varieties that have undergone substantial syntactic changes) is the product of imperfect learning as the speakers of SAE languages shifted to Turkish as their primary language – a case of substratum influence.

The starting point for this answer is a prediction by Thomason and Kaufman (1988: 113–114; see also Thomason 2001: 80). Now, according to these authors, typically, the features transferred from speakers’ primary languages into their

secondary languages in language shift situations are phonological and syntactic; influences on the vocabulary are minimal at best (Thomason 2001: 80; Thomason and Kaufman 1988: 38–39). The prediction to be derived from this observation is the following: if there are significant contact-induced structural changes, but few or no loanwords in a language, then these changes must have come about by means of imperfect learning of that language as target language during language shift (and not through the borrowing of structural features).

When we examine early transcription texts in the light of this prediction (e.g. Ferraguto 1611; Georgievits 1544; Herbinus 1675; Illésházy 1668), we observe the following: the texts contain few loanwords (only in non-basic vocabulary) or none at all from the majority SAE languages of the Balkans (see e.g. Rocchi 2011), but show significant syntactic changes (e.g. beginning of the shift to SAE-type subordination and to VO order; see e.g. Keskin 2023a, 2023b), whose most likely source is contact with those languages. These observations, thus, point to the following scenario (cf. Johanson 2021: 182; Thomason 2001: 75–76; Thomason and Kaufman 1988: 39).

Parts of the SAE-speaking population of the Balkans under Ottoman rule learn Turkish as a non-primary language, failing to learn some of its features (e.g. Turkic SCs). These groups are then integrated into the original Turkish-speaking community in the Balkans, a prestigious minority in the region. Presumably, the shifting groups are large as compared to the original Turkish-speaking community in the region, and the language of the shifting groups and of the original Turkish-speaking groups amalgamate, and the shifting groups' transferred features become fixed in the Turkish spoken in the region.

Why, then, were some features of Turkish not learned by the shifting SAE speakers? One answer is that this is because the features in question are 'marked' (Thomason 2001: 65; Thomason and Kaufman 1988: 51), i.e. because, in a nutshell, they pose challenges to learners due to their "relative productive and perceptual difficulty" (Thomason and Kaufman 1988: 26). Thus, marked features are often removed during language contact (Thomason 2001: 65). More precisely, the non-acquisition of these features results in simplificatory replacements (cf. Thomason and Kaufman 1988: 129). In the case of Turkish, Turkic SCs are among the marked features, and they are replaced by SAE-type SCs. One piece of evidence for this comes from Slobin (1986) who argues that, due to general psycholinguistic processing principles, Turkic relative clauses are acquired later by Turkish speakers and are less frequent in their discourse, as compared to IE relative clauses in English; they are also often replaced in contact situations.

An alternative to the markedness (and the resultant simplificatory replacement) explanation or perhaps even the broader imperfect learning approach to substratum influence is the 'convergent development' account in Matras (2003) – a study on the language-internal mechanisms involved in contact-induced syntactic change in

Macedonian Turkish, a member of the WRT group. According to Matras (2003: 63–64), these mechanisms “are triggered by the pressure to syncretize sentence planning operations among congruent languages leading to convergence of abstract structures and patterns of sentence arrangement, though no replication of actual linguistic material from the contact language is involved”.

We can bring together Matras’ proposal with the language shift account as follows: in the Balkan sprachbund, multilingual SAE speakers in the process of shifting to Turkish as their primary language streamlined “the mental operations involved in planning the utterance and expressing relations between its individual propositional units” (Matras 2003: 69) among their languages, resulting in Turkish converging in structure (e.g. SAE-type SCs) with other Balkan languages. In other words, there was no replacement per se of Turkic SCs with SAE-type SCs due to the former’s markedness. That is to say, the SAE substratum influence on RT is not because of imperfect learning but due to convergent development.

Now, the foregoing proposal provides the reason why Turkish should shift from one subordination strategy to the other (i.e. for why an adaptive change is triggered when Σ is exposed to an external impact) and contributes to some degree to the discussion of the structural patterns produced in this process. However, as it is, it cannot address the phenomenon of X-clauses. For instance, Matras (2003) observes that both relatives and adverbials in Macedonian Turkish involve the reanalysis of interrogatives, and some of the structures he analyses are X-clauses. However, the theoretical framework adopted cannot address the following puzzle: as relative and adverbial clauses are being refashioned from interrogatives, they emerge from the process in the form of several different X-clause subtypes as well as in the form of SAE-type SCs. Why? Indeed, each clause type in PRT (i.e. argument, relative, or adverbial) can be in the form of several subtypes, as shown in Table 8.

Furthermore, neither the simplificatory replacement nor the convergent development accounts work for X-clauses as there are no comparable structures in Balkan languages to be used as replacements or to converge with. As X-clauses are not part of the Turkic inventory either, they must have arisen due to language-internal mechanisms of change triggered by that contact situation other than the processes of replacement or convergence (see also Section 6).

Table 8: Distribution of clause types into subtypes.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
Argument cl.	–	–	–	–	+	+	+	–	+	+	+	–	+
Relatives	–	+	+	–	–	–	+	–	+	+	–	–	+
Adverbials	+	–	+	+	–	+	+	–	+	+	+	+	+

This is where my study of X-clauses comes into play to fill in the lacuna, and the interface between contact theoretic approaches and my approach is the selectionist scenario at the end of Section 11: the process of simplificatory replacement due to markedness or the pressure to syncretize sentence planning operations drives the random generation of various new SC types (in the form of X-clauses and SAE-type SCs), and some of these new SC types are selected by the linguistic environment.²⁷

12.2 Sociolinguistic variables

Let us now move on to the present-day sociolinguistic setting in which RT finds itself in contact with SAE languages. The question we will ultimately try to address is whether any sociolinguistic variables have an explanatory potential for the X-clauses problem.²⁸

12.2.1 Status

On the whole, it is clear that Turkish does not have in the present day the prestige that it enjoyed in the Balkans during the Ottoman period, as the balance of power shifted in favor of the majority languages of the area after the ethnic groups under Ottoman rule began to break away in the 19th century.

12.2.2 Speech community size

Since the 19th century, through waves of emigration mostly to Turkey, RT speakers who were already in the minority across the Balkans except in eastern Bulgaria, declined drastically in number and became marginalized. Today, all RT varieties are endangered to varying degrees (see e.g. Moseley and Nicolas 2010: 25). The number of their speakers is at around 1% of the population in Kosovo, 3% in North Macedonia, 5% in Moldova, and 9% in Bulgaria. In Kosovo and Macedonia Turkish speech communities are fragmented, making up about 0.06% on average of the population of the numerous municipalities in which RT is spoken in Kosovo, and 2% in Macedonia. In Bulgaria and Moldova the situation is somewhat different. In Kardzhali and Razgrad

²⁷ This broader contact theoretic account should not be interpreted as the only pathway for the emergence of X-clauses. That is an empirical matter which remains to be investigated in future work: under which contact-related dynamics do X-clauses emerge?

²⁸ The information in this section is gathered from the following sources and the references therein unless otherwise indicated: Bolver (2014: 30–34), Dönmez (2012: 13–17), Haliloğlu (2017: 31–34), Özkan (1996: 24–25), Sulçevsi (2019: 15–40), Svanberg (2011), and Tufan (2001: 78–79).

provinces in eastern Bulgaria 50% or more of the population speaks Turkish, and in Silistra, Targovishte, and Shumen provinces around 35% on average. In many municipalities in these provinces, Turkish is spoken by the majority of the population. In Moldova, Gagauz speakers mostly live in the autonomous region of Gagauz Yeri, where they constitute around 84% of the population (percentages are based on data from: Istrati [2017]; Kotzeva [2011]; Simovski [2022]; Zabërgja et al. [2013]).

12.2.3 Bi-/multilingualism

Bi-/multilingualism is the norm among RT speakers and interethnic marriages are common in Kosovo, North Macedonia, and Moldova. Based on data from Sulçevsi (2019: 192–261), one can estimate that 53% of KT speakers are trilingual in Turkish, Serbo-Croatio-Bosnian and Albanian, about 20% are bilingual in Turkish and Serbo-Croatio-Bosnian, and the remaining 27% are trilingual variously in Albanian, Serbo-Croatio-Bosnian, and Turkish.

I have no explicit, direct data on NRT speakers. I glean from my sources that bilingualism in Bulgaria and Moldova is also widespread, but the informants seem to be less competent in the majority languages than KT speakers. I assume that all the informants from Bulgaria speak Bulgarian as a second language. The Gagauz informants speak Romanian and/or Russian. Note, however, that the Gagauz began to settle in Bessarabia in the 18th century, before which the SAE language that they were in contact with was Bulgarian.

12.2.4 Official status and institutional support

In all the above-mentioned countries and territories, RT is today under constitutional protection and, except in Bulgaria, has some degree of official status. As part of their language rights, RT speakers can in principle receive education in their native varieties, but practical hurdles make this mostly impossible except in Gagauz Yeri. Also, ambivalent official attitudes, unofficial sanctions against RT speakers, and general official resistance are commonly reported, which seems to reflect the negative attitudes RT has been subjected to since the end of the Ottoman period. These attitudes were particularly extreme as part of the assimilationist policies in Bulgaria until the establishment of the democratic regime.

12.2.5 Use in daily life

RT is mostly spoken in the speakers' private lives, but it is also widely used in the media, cultural institutions, organizations, political parties, etc. As pointed out

above, it is also used in education but to a much more limited extent, with Gagauz enjoying an exceptional status in this regard.

12.2.6 Logistic regression with sociolinguistic variables

I explored the potential effects of various sociolinguistic factors on the use of X-clauses in PRT with two logistic regression models. These models showed none of the sociolinguistic variables taken into consideration to be of explanatory significance. Note, however, that the information on these variables provided by my sources (see Section 4) is rather patchy and limited. Consequently, these statistical models should be considered preliminary explorations.

For the first model, I coded all the examples used in this study (see Section 4) for whether they contained X-clauses ('yes' and 'no'); this was the outcome variable. In addition, I also coded the explanatory variables (or predictors) which could potentially have an effect on the use of X-clauses according to sociolinguistic literature (e.g. Llamas et al. 2007) and on which my sources provided information. These were (i) primary language (Turkish or Albanian),²⁹ (ii) secondary languages ('Albanian and others' or 'Serbo-Croatio-Bosnian and others'), (iii) age (numerical values), (iv) level of education (primary, secondary, or university), (v) gender (female or male), and (vi) speech community size (percentage of Turkish speakers in the dialect locale from which the example came). Since information on these variables were available only for KT, the first model essentially tested the effect of the listed predictors on X-clause use by KT speakers only. The model revealed that none of the explanatory variables were significant predictors of X-clause use ($LL = 11.76$, $p = 0.068$), and it had acceptable classification properties (C -index = 0.71).

For the second model, I again coded all the examples used for this study for whether they contained X-clauses, as the outcome variable. The explanatory variables in this model were restricted to (i) primary language (Turkish or Albanian), (ii) secondary language(s) ('Albanian and others', 'Serbo-Croatio-Bosnian and others', or Bulgarian), and (iii) gender (female or male). Unlike the first model, this model with its fewer predictors could cover Dobruja Turkish in addition to KT, thanks to the availability of information. The model showed secondary language to be a significant predictor of X-clause use ($LL = 19.75$, $p = 0.0002$), however it had unacceptable classification properties (C -index = 0.68): PRT speakers who are bilingual in Turkish and Bulgarian produce significantly more X-clauses than those who are bilingual in Turkish and Serbian, Albanian, etc. While the former produce a balanced percentage

29 Of the informants of my sources cited in Section 4, 94% reported that they speak Turkish as their native language, and 6% said their native language is Albanian. Of the informants who produced X-clauses 97% percent were speakers of Turkish and 3% of Albanian.

of X-clauses versus SAE-type clauses (54 vs 46%), the latter produce lower percentages of X-clauses versus SAE-type clauses (26 vs 74% on average). This test result essentially echoes the observation in Section 5.1 that Dobruja Turkish is a hotbed of X-clauses and does not afford any new insights. I defer the question of why Dobruja Turkish produces such a high percentage of X-clauses to a later study.

13 Conclusion

The Turkic varieties of the Balkans make use of two main subordination strategies that have diametrically opposed structural properties: the native Turkic model, which is in marked decline, and the Indo-European model, which is the preponderant model on average. In addition, Peripheral Rumelian Turkic, a subgroup of Balkan Turkic, makes use of several kinds of subordinate clause ('X-clauses') that do not fit the Turkic and Indo-European models well and allow for atypical mixtures of these two complementary models. Structurally, X-clauses can be said to be spread out over a spectrum between the Turkic and the Indo-European extremes. The first purpose of this paper has been to lay out the phenomenon of X-clauses as a well-defined research problem ('the X-clauses problem'). As the second task of the paper, I put forward a hypothesis ('the transient behavior hypothesis') whereby X-clauses come about due to uncertainties in the values of the structural parameters of the Peripheral Rumelian subordination system ('oscillations'). Such oscillations are typical of complex systems undergoing change and arise in the present case due to the shift in Peripheral Rumelian away from Turkic towards Indo-European subordinate clauses, more precisely towards Standard Average European-type subordinate clauses within the Indo-European class. I presented three arguments for the transient behavior hypothesis and the general diachronic approach that it is embedded in. The first argument involved showing how the structural parameters of X-clauses with seemingly unset, in-between, or contradictory values could be interpreted as oscillations. My second argument was of a more general nature and focused on the differences in the percentage distributions of different clause types (i.e. Turkic, Persian-type, Standard Average European-type, and X-clauses) in 17th century versus 21st century data, which suggest that the shift from Turkic to Standard Average European-type clauses was facilitated by X-clauses. Third, as another general argument for a diachronic approach to X-clauses, I showed that between Early and Modern Peripheral Rumelian there appears to have been a frequency drift from X-clauses that are structurally closer to Turkic subordinate clauses towards those that are more like Standard Average European subordinate clauses. Thus, X-clauses look like a bridge between the two diametrically opposed models from this perspective as

well. Finally, I proposed that the shift to Standard Average European-type clauses and the attendant emergence of X-clauses took place due to language-internal processes in a context of language shift by Standard Average European speakers to Turkish.

List of abbreviations

1	first person
2	second person
3	third person
ABL	ablative
ACC	accusative
ANOM	action nominal
AOR	aorist
CONN	connector
COP	copula
DAT	dative
DIST	distal
EVID	evidential
FNOM	factive nominal
FUT	future
GEN	genitive
NEG	negation
OPT	optative
OREL	object relative
PASS	passive
PL	plural
POSS	possessive
PRF	perfect
PROG	progressive
PROX	proximal
PST	past
SG	singular

Acknowledgements: First and foremost, I thank Christoph Schroeder for his unwavering encouragement and support, as well as his comments, suggestions, exhortations, etc. I also thank the *Folia Linguistica Historica* Editorial Board, two anonymous reviewers, and the following colleagues (in alphabetical order): Daria Alfimova, Ümit Atlamaz, Metin Bağrıaçık, Fatih Bayram, Ömer Demirok, Lale Diklitaş, Marcel Erdal, Zeynep Güler, Annette Herkenrath, Hans Henrich Hock, Kateryna Iefremenko, Matthias Kappler, Jaklin Kornfilt, Kirill Kozhanov, Tanja Kupisch, Yaron Matras, Astrid Menz, Ehtiram Muzaffarli, Nil Özkul, Maria Petrou, Julian Rentzsch, Sergey Say, Ilja Seržant, Taner Sezer, Dmitri Sitchinava, Stavros

Skopeteas, Barbara Sonnenhauser, Heidi Stein, Peter Svenonius, Deniz Tat, and Süleyman Yüceer. The usual disclaimers apply.

Research funding: The academic research that constitutes the foundation of this article was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant 313607803, as part of the projects *Head directionality change in Turkic in contact situations: A diachronic comparison between heritage Turkish and Balkan Turkic* (project number: SCHR 1261/3-1) and *Clause combining in Balkan Turkic: Pathways and stages of contact-induced grammaticalization* (project number: SCHR 1261/4-1) within the Research Unit *Emerging Grammars in Language Contact Situations: A Comparative Approach* (FOR 2537).

References

- Artun, Erman. 2013. Geçmişten günümüze kültürel değişim ve gelişim sürecinde Balkanlarda Türkçe [Turkish in the Balkans in its historical process of cultural change and development]. In Adem Balaban & Bünyamin Çağlayan (eds.), *Uluslararası dil ve edebiyat çalışmaları konferansı "Balkanlarda Türkçe" bildiri kitabı* [Proceedings of the international conference on Turkish in the Balkans], vol. 1, x–xx. Tirana: Universiteti "Hëna e Plotë" (BEDËR) Press.
- Auwers, Johan van der. 2011. Standard Average European. In Bernd Kortmann & Johan van der Auwers (eds.), *The languages and linguistics of Europe: A comprehensive guide*, vol. 1, 291–306. Berlin & New York: Mouton De Gruyter.
- Bayer, Josef. 2001. Two grammars in one: Sentential complements and complementizers in Bengali and other South Asian languages. In Rajendra Singh (ed.), *The yearbook of South Asian languages and linguistics*, 11–36. New Delhi: Sage.
- Boev, Emil. 1968. Bulgaristan Türk diyalektolojisiyle ilgili çalışmalar [Works on Turkish dialectology in Bulgaria]. In *11. Türk Dil Kurultayında okunan bilimsel bildiriler 1966* [Proceedings of the 11th Turkish Language Congress 1966], 171–178. Ankara: Türk Tarih Kurumu.
- Bolver, Mert. 2014. *Üsküp'te yaşayan Türklerin dil durumu* [The linguistic situation of the Turks in Skopje]. Samsun: Ondokuz Mayıs University MA thesis.
- Bombaci, Alessio. 1940. Padre Pietro Ferraguto e la sua grammatica turca (1611). *Annali del Reale Istituto Superiore Orientale di Napoli, Nuova Serie* 1. 205–236.
- Brendemoen, Bernt. 1980. Labiyal ünlü uyumunun gelişmesi üzerine bazı notlar [Some notes on the development of labial vowel harmony]. *Türkiyat Mecmuası* 19. 223–240.
- Brezina, Vaclav. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Brezina, Vaclav, Matt Timperley & Tony McEnery. 2018. #LancsBox (version 4.5). <http://corpora.lancs.ac.uk/lancsbox> (accessed 15 June 2020).
- Brown, Keith & Jim Miller. 2013. *The Cambridge dictionary of linguistics*. Cambridge: Cambridge University Press.
- Çımpoș, Lăbuș S. 1988. *Literatura okumakları 9–10 klaslar için* [Literature readings for 9th and 10th grades]. Chișinău.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.
- Crystal, David. 2008. *A dictionary of linguistics and phonetics*, 6th edn. Malden, MA: Blackwell.

- Csató, Éva Á., Astrid Menz & Fikret Turan (eds.). 2016. *Spoken Ottoman in mediator texts*. Wiesbaden: Harrassowitz.
- Csató, Éva Á. & Lars Johanson. 1998. Turkish. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, 203–235. London & New York: Routledge.
- Demirok, Ömer & Balkız Öztürk. 2022. The emergence of clausal nominalizations in Laz. Paper presented at *The international workshop on Formal Approaches to Contact in/with Turkish*. Tromsø: UiT The Arctic University of Norway.
- Dönmez, Özlem D. 2012. *Kuzeydoğu Bulgaristan Silistre ili Dulovo/Akkadınlar ilçesi Türk ağızları* [The Turkish dialects of Dulovo municipality in Silistra province of north-eastern Bulgaria]. Ankara: Gazi Kitabevi.
- Dryer, Matthew S. 2013. Order of adverbial subordinator and clause. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures online*, Ch. 94. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/94> (accessed 3 December 2021).
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Ferraguto, Pietro. 1611. *Grammatica turchesca* [Turkish grammar]. Unpublished manuscript. Napoli.
- Friedman, Victor A. 2006. West Rumelian Turkish in Macedonia and adjacent areas. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 27–45. Wiesbaden: Harrassowitz.
- Georgievits, Bartholomaeus. 1544. *De Turcarum ritu et caeremoniis* [On the ritual and ceremonies of the Turks]. Antwerp: Gregorius Bontius.
- Göksel, Aslı & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London & New York: Routledge.
- Gülvodina, Gonca. 2018. *Tirgovişte (Eski Cuma) ili ve yöresi ağızlar* [The dialects of Targovishte province and the environs]. Edirne: Trakya University MA thesis.
- Güneş, Nurcihan. 2009. *Kuzeydoğu Bulgaristan'da Çerkovna köyü ve çevresi Türk ağızları* [The Turkish dialects on Cherkovna and the environs in north-eastern Bulgaria]. Malatya: İnönü University MA thesis.
- Günşen, Ahmet. 2012. Balkan Türk ağızlarının tasnifleri üzerine bir değerlendirme [An evaluation of the classifications of Balkan Turkish dialects]. *Turkish Studies* 7(4). 111–129.
- Haliloğlu, Neşe. 2017. *Zavet köyü Türk ağızı* [The Turkish dialect of Zavet]. Kırşehir: Ahi Evran University MA thesis.
- Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20. 271–287.
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2, 1492–1510. Berlin & New York: Mouton de Gruyter.
- Hazai, György. 1990. Die Denkmäler des Osmanisch-Türkisch in nicht-arabischen Schriften. In György Hazai (ed.), *Handbuch der türkischen Sprachwissenschaft*, 63–73. Budapest: Akadémiai Kiadó.
- Heffening, Willi. 1942. *Die türkischen Transkriptionstexte des Bartholomaeus Georgievits aus den Jahren 1544–1548*. Leipzig: Brockhaus.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Henderson, Eugénie J. A. 1997. *Bwe Karen dictionary: With texts and English-Karen word list*. London: School of Oriental and African Studies, University of London.
- Herbinius, Johannes. 1675. *Horae turcico-catecheticae* [Turkish cathetical hours]. Gdansk: David-Fridericus Rhetius.
- Illésházy, Miklós. 1668. *Dictionarium turcico-latinum* [Turkish–Latin dictionary]. Unpublished manuscript. Vienna.

- Istrati, Valentina. 2017. *Key results of the 2014 Population and Housing Census*. Chişinău: National Bureau of Statistics of the Republic of Moldova. <https://statistica.gov.md/newsview.php?l=en&id=5583&idc=168> (accessed 4 August 2022).
- Johanson, Lars. 1998a. The history of Turkic. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, 81–125. London & New York: Routledge.
- Johanson, Lars. 1998b. The structure of Turkic. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, 31–66. London & New York: Routledge.
- Johanson, Lars. 2021. *Turkic* (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Karaşinik, Bahtışen. 2011. *Silistre (Silistra) ili ve yöresi ağızları* [The dialects of Silistra province and the environs]. Edirne: Trakya University MA thesis.
- Kerslake, Celia. 1998. Ottoman Turkish. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, 179–202. London & New York: Routledge.
- Kerslake, Celia. 2007. Alternative subordination strategies in Turkish. In Jochen Rehbein, Christiane Hohenstein & Lukas Pietsch (eds.), *Connectivity in grammar and discourse* (Hamburg Studies on Multilingualism 5), 231–258. Amsterdam & Philadelphia: John Benjamins.
- Keskin, Cem. 2022. Erken Batı Rumeli Türkçesiyle Yunus Emre: Georgius de Hungaria'nın transkripsiyon metinlerindeki Batı Rumeli Türkçesi özellikleri [Yunus Emre in Early West Rumelian Turkish: West Rumelian Turkish features in Georgius de Hungaria's transcription texts]. In Abdullah Esen, Ömer S. Güler, Özlem Turan, Sümeyye Sarı & İbrahim B. Durak (eds.), *Dost bağının meyveleri: Nurettin Albayrak hatıra kitabı* [Fruits of the garden of friends: In memory of Nurettin Albayrak], 295–324. İstanbul: Türk Edebiyatı Vakfı Yayınları.
- Keskin, Cem. 2023a. On the directionality of the Balkan Turkic verb phrase. In Jaklin Kornfilt (ed.), *Theoretical Studies on Turkic Languages*. [Special issue]. *Languages* 8(1). Available at: <https://www.mdpi.com/2018894>.
- Keskin, Cem. 2023b. Rumelian Turkish features in Pietro Ferraguto's *Grammatica turchesca* (1611). *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 173(1). 115–143.
- Keskin, Cem, Kateryna Iefremenko, Jaklin Kornfilt & Christoph Schroeder. In preparation a. *Aspects of clause combining in Turkish language contacts*. University of Potsdam and Syracuse University.
- Keskin, Cem, Taner Sezer, Christoph Schroeder, Lale Diklitaş, Zeynep Güler, Nil Özkul, Süleyman Yüceer, Ayşe N. Güler & Ebru Özgenç. In preparation b. *Balkan Turkic corpus*. University of Potsdam.
- Kılıç, Sedef. 2018. *Rumeli ağızlarında ilgi cümleleri* [Relative clauses in Rumelian dialects]. Ankara: Hacettepe University MA thesis.
- Kissling, Hans J. 1968. Bemerkungen zu einigen Transkriptionstexten. *Zeitschrift für Balkanologie* 6. 119–127.
- Kornfilt, Jaklin. 1997. *Turkish*. London & New York: Routledge.
- Kotzeva, Mariana. 2011. *Census 2011*. Sofia: Bulgarian National Statistical Institute. <https://nsi.bg/census2011/indexen.php> (accessed 4 August 2022).
- Laitinen, Mikko. 2012. Typological hierarchies and frequency drifts in the history of English. In Terttu Nevalainen & Elizabeth C. Traugott (eds.), *The Oxford handbook of the history of English*, 633–642. Oxford: Oxford University Press.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicolas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.
- Llomas, Carmen, Louise Mullany & Peter Stockwell (eds.). 2007. *The Routledge companion to sociolinguistics*. London & New York: Routledge.
- Marshall, Samuel A. 1978. *Introduction to control theory*. London: Macmillan Education UK.
- Matras, Yaron. 2003. Layers of convergent syntax in Macedonian Turkish. *Mediterranean Language Review* 15(4). 63–86.

- Matras, Yaron & Şirin Tufan. 2007. Grammatical borrowing in Macedonian Turkish. In Yaron Matras & Jeanette Sakel (ed.), *Grammatical borrowing in cross-linguistic perspective*, 215–227. Berlin & New York: Mouton de Gruyter.
- Menz, Astrid. 1999. *Gagausische Syntax: Eine Studie zum kontaktinduzierten Sprachwandel* (Turcologica 41). Wiesbaden: Harrassowitz.
- Menz, Astrid. 2001. Gagauz right-branching propositions introduced by the element *ani*. *Turkic Languages* 5(2), 234–244.
- Menz, Astrid. 2006. On complex sentences in Gagauz. In Hendrik Boeschoten & Lars Johanson (eds.), *Turkic languages in contact*, 139–151. Wiesbaden: Harrassowitz.
- Mobus, George E. & Michael C. Kalton. 2014. *Principles of systems science* (Springer Complexity). New York, NY: Springer.
- Moseley, Christopher & Alexandre Nicolas (eds.). 2010. *Atlas of the world's languages in danger: Maps*. Paris: UNESCO Publishing.
- Moškov, Valentin A. 1904. Narèčija bessarabskich gagauzov [The Gagauz dialect of Bessarabia]. In Vasilij V. Radlov (ed.), *Proben der Volksliteratur der türkischen Stämme*, vol. 10. St. Petersburg: Commissioners of the Imperial Academy of Sciences.
- Murtaza, Durgul M. 2016. *Razgrad ili Kubrat ilçesi Alevi-Bektaşî köyleri ağızı* [The dialect of the Alevi-Bektashi villages in Razgrad province Kubrat county]. Edirne: Trakya University MA thesis.
- Németh, Gyula. 1956. *Zur Einteilung der türkischen Mundarten Bulgariens*. Sofia: Bulgarische Akademie der Wissenschaften.
- Németh, Gyula. 1968. Die türkische Sprache des Bartholomaeus Georgievits. *Acta Linguistica Academiae Scientiarum Hungaricae* 18(3–4), 263–271.
- Németh, Gyula. 1970. *Die türkische Sprache in Ungarn im siebzehnten Jahrhundert* (Bibliotheca Orientalis Hungarica 13). Budapest: Akadémia Kiadó.
- Németh, Gyula. 1980. Bulgaristan Türk ağızlarının sınıflandırılması üzerine [On the classification of Turkish dialects of Bulgaria]. *Türk Dili Araştırmaları Yıllığı-Belleten* 1981, 113–167.
- Özkan, Nevzat. 1996. *Gagavuz Türkçesi grameri* [Gagauz grammar]. Ankara: Atatürk Kültür, Dil ve Tarih Yüksek Kurumu.
- Özkan, Nevzat. 2007. *Gagavuz destanları* [Gagauz legends]. Ankara: Atatürk Kültür, Dil ve Tarih Yüksek Kurumu.
- Pokrovskaya, Ljudmila A. 1964. *Grammatika gagauzkogo jazyka: Fonetika i morfologija* [A grammar of the Gagauz language: Phonetics and morphology]. Moscow: Nauka.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Rocchi, Luciano. 2011. Turkish historical lexicography. *Lexicographica* 27, 195–220.
- Schönig, Claus. 1995. **qa:ño* und Konsorten. In Marcel Erdal & Semih Tezcan (eds.), *Beläk Bitig: Sprachstudien für Gerhard Doerfer zum 75. Geburtstag* (Turcologica 23), 177–187. Wiesbaden: Harrassowitz.
- Simovski, Apostol. 2022. *Census 2021*. Skopje: Republic of North Macedonia State Statistical Office. https://www.stat.gov.mk/PrikaziSoopstenie_en.aspx?rbtxt=146 (accessed 4 August 2022).
- Slobin, Dan I. 1986. The acquisition and use of relative clauses in Turkic and Indo-European languages. In Dan I. Slobin & Karl Zimmer (eds.), *Studies in Turkish linguistics*, 273–294. Amsterdam & Philadelphia: John Benjamins.
- Stein, Heidi. 2014. Ferraguto, Pietro. In Kate Fleet, Gudrun Krämer, Denis Matringe, John Nawas & Everett Rowson (eds.), *Encyclopaedia of Islam, THREE*. Leiden: Brill. https://doi.org/10.1163/1573-3912_ei3_COM_27103 (accessed 23 February 2021).

- Stein, Heidi. 2016. The dialogue between a Turk and a Christian in the *Grammatica turchesca* of Pietro Ferraguto (1611). Syntactical features. In Éva Á. Csató, Astrid Menz & Fikret Turan (eds.), *Spoken Ottoman in mediator texts*, 161–171. Wiesbaden: Harrassowitz.
- Sulçevsi, İsa. 2019. *Kosova Türkçesi ve fiil yapıları* [Kosovar Turkish and its verb forms]. Istanbul: Istanbul University dissertation.
- Svanberg, Ingvar. 2011. Gagauz. In Jeffrey E. Cole (ed.), *Ethnic groups of Europe: An encyclopedia*, 159–162. Santa Barbara, CA: ABC-CLIO.
- Taylor, Charles. 1985. *Nkore-Kiga*. London: Croom Helm.
- Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tufan, Şirin. 2001. *On word order in Gostivar Turkish*. Istanbul: Boğaziçi University MA thesis.
- Unseth, Peter. 1989. Sketch of Majang syntax. In Lionel M. Bender (ed.), *Topics in Nilo-Saharan linguistics*, 97–127. Hamburg: Helmut Buske Verlag.
- Whorf, Benjamin L. 1944. The relation of habitual thought and behavior to language. *ETC: A Review of General Semantics* 1(4). 197–215.
- Zabërgja, Sabri, Kadri Sojeva, Avni Kastrati, Zymer Maxhari, Rrahman Tara, Dren Gashi & Hysni Ferizi. 2013. *Population by mother tongue, sex and municipality 2011*. Pristina: Kosovo Agency of Statistics. https://askdata.rks-gov.net/pxweb/en/ASKdata/ASKdata__Census%20population__Census%202011__3%20By%20Municipalities/tab%205%206.px/ (accessed 4 August 2022).
- Zaiontz, Charles. 2020. *Real statistics using Excel* (version 8.3.1). <https://www.real-statistics.com/> (accessed 20 April 2020).