

On the Sociolinguistic Origins of the term *Qazaq*: A Proposal for an Alternative Etymology of ‘Cossack’/‘Kazakh’ and an Argument for the Analytical Usefulness of Register in Historical Linguistics

Patrick C. LEWIS*

University of Mississippi, USA

Received: July 10, 2022 • Accepted: January 20, 2023

© 2023 Akadémiai Kiadó, Budapest



ABSTRACT

Although the word *qazaq* is widely agreed to be the common ancestor of both ‘Cossack’ and ‘Kazakh’, there remains uncertainty about *qazaq*’s etymology. This paper proposes that *qazaq* originated as a variant of the central Turkic etymon *qaç(g)aq* and thus derives from the Common Turkic root *qaç-* (‘flee’). The paper draws on data generated through the traditional methods of historical linguistics—the comparative method and philological analysis—in conversation with recent work in anthropological and sociolinguistics on the concept of linguistic ‘register’ to better theorize *qazaq*’s historical development and the variation in its forms and meanings.

KEYWORDS

Kazakh, Cossack, *qazaq*, Turkic historical linguistics, register

A connection between the ethnonyms Cossack and Kazakh is well established in Turkic and Slavic historical linguistics, even if the precise nature of their linguistic origin has been insufficiently explained. Although today these two terms refer to distinct ethnopolitical identities, historical linguistics have long understood that both ethnonyms are most likely derived from a common etymon *qazaq*, and they observe that this lexical item did not refer to a specific ethnic or national group but rather a generalized sociopolitical status. Beginning in the 15th century, the term be-

* Corresponding author. E-mail: pclewis@uchicago.edu

comes widely attested in many regional languages and literary traditions, reaching from Eastern Europe to Northern India, in a range of meanings from ‘unmarried’, ‘slave’, ‘vagabond’, ‘brigand’, ‘rebel’ (Doerfer 1963; Gabain 1960; Lee 2015; Pritsak 2006). However, despite widespread consensus around the *qazaq*’s sociopolitical importance to Eurasian steppe societies, scholars have been less able to establish an agreed-upon etymology for *qazaq*. Thus does Pritsak (2006: 237), in his paper on the topic, begin by noting that while most historians and etymologists of the Slavic languages had accepted ‘without any reservation, but also without proof’ that *qazaq* had a Turkic etymology, some prominent names in Turkish historical linguistics still had their doubts:

Concerning the Turkic etymology of the word Qazaq, there is no consensus among scholars...Räsänen (1969: 243), in his etymological dictionary of the Turkic languages, gives no etymology in his entry *qazaq*, and Doerfer (1967: 462–68) states flatly: ‘eine sichere Etymologie für das Wort [*qazaq*] existiert nicht.’ The same opinion was expressed by Menges (1979: 196, no. 11,12).

Although folk etymologies for *qazaq* continue to circulate and scholars have continued to debate its etymology over the past century, none of the best-known proposals—such as a connection to a Mongolian word for *qasax* (‘cart’, cf. MMon. *xəcəz*), the Common Turkic (CT) roots *qaz-* (‘dig’), *qazgan-* (‘win/gain’), or most persistently, a non-attested (and therefore hypothetical) variant **qaz-* of the Common Turkic root *kez-* (‘wander’)—have succeeded in gaining general acceptance among scholars, as evidenced by Räsänen, Dorfer and Menge’s shared skepticism of these theories.¹ A further line of inquiry has considered a potential connection between *qazaq*

¹ In her popular history of the Kazakhs, for instance, Olcott (1987: 4) begins by noting the contested origins of the word *qazaq*, listing several folk theories related to the mythology Kazakh ethnogenesis, as well as an older, more established theory linking *qazaq* to a root **qaz-* the unattested, and thus hypothetical, variant of the Common Turkic root *kez-* (‘wander’) noted above. Németh (1930)—drawing on entries for **qaz-* (‘wander’) in both Pavet de Courteille’s (1870: 405) Chagatai dictionary and Radlov’s (1893) Turkic etymological dictionary—was the first to draw an explicit etymological link between this hypothetical variant **qaz-* and *qazaq*, pointing to the Turkmen verbal root *gez-* (‘wander’). Gabain (1960: 167) further developed this theory in her influential discussion of the sociopolitical dimensions of *qazaqliq*, wherein she also argues for a relationship between *qazaq* and this hypothetical variant **qaz-* (in the meaning of ‘wander’), the Common Turkic root *qaz(gan)-* (as in ‘win’ or ‘gain’), and of hodgepodge of other historical and modern Turkic words (some of questionable etymological relationship). Pritsak (2006), more recently, offers a modified version of Gabain’s etymology, similarly tracing *qazaq* to a Turkic root *qaz-* which he links both to a common Turkic root ‘dig’ and to an earlier, contracted (and also hypothetical) variant of *qaz(g)an* (‘win’ or ‘gain’): **qaz-*. However, as noted above these connections are generally not accepted by most Turcologists today and there remains significant doubt over whether a verb like **qaz-* (as a variant of *kez-* meaning ‘wander’) ever existed. No such entry is found in Budagov (1869) or Buḥārī’s (1892) Chagatai dictionaries (except in expected meaning of ‘dig’) or Ünlü’s (2014) more recent, comprehensive Turkish-Chagatai dictionary. And Zimonyi (1990: 58), citing Róna-Tas (1982a; 1983), points out that **qaz-* in the meaning of ‘wander’ is most likely a ghost word (i.e., a word proposed by 19th and 20th-century Turkic historical linguists that is not attested in any text and more than likely never existed). Nor is there any obvious semantic connection between ‘dig’ and the meanings ascribed to *qazaq*—except in the highly imaginative connection between the pawing of a horse and an itinerant bandit drawn by Gabain (1960: 167). While the semantic connection between ‘win’ and *qazaq* is seemingly less tenuous, it remains highly speculative. The process of affixation in Turkic languages is well-established, very generalizable, and would normally preclude the deletion of an entire root syllable—i.e., *kaz(g)an* + *ak* > *kazak*—which is why Pritsak’s (2006) analysis relies on the existence of a hypothetical original verb **kaz-* in the meaning of ‘raid’ (of which *kaz(g)an* is meant to be a derivation) but no such word is attested in historical texts or in the modern Turkic languages.



and the ethnonym Khazar (see Golden 2007: 175). However, the Khazar Khaganate disappeared centuries before the first attestations of *qazaq* in the 14th century and this connection, while not disproven, remains unsubstantiated. There is a further etymological possibility, however, one first proposed more than a century ago by the Tatar public intellectual and Turcologist Hadi Atlasi (1911), albeit one largely unaddressed in more recent discussions of the question²: the proposal that *qazaq* is a variant of a Western Common Turkic etymon *qaç(g)aq* and therefore derived from a combination of the Common Turkic root *qaç-* (flee, break away from) and the CT suffix *-(g)aq*. This proposal, this paper argues, finds support in existing evidence and is the most plausible etymological explanation.

In advancing the case for an etymological connection between **QAZAQ** and **QAÇAQ**³, the paper also develops several points of sociolinguistic and sociological speculation about the period of large-scale political and social transformation that shaped life on the Pontic-Caspian steppe and in Volga-Kama region beginning with the Mongol invasions in the 13th century—a period also marked by important linguistic shifts during which the first attestation of both **QAZAQ** and **QAÇAQ** first appear in the written record, several decades apart, in the first half of the 14th century. At the same time, the paper raises several questions about the interplay between language and dialect in multiethnic and multilingual societies, and by extension, makes a case for the analytical usefulness of ‘register’ as a conceptual tool in historical linguistics. The first two sections of the paper provide brief outlines of the linguistic development of **QAÇAQ** (section I) and **QAZAQ** (section II). Section III develops the argument for the etymological origin of **QAZAQ**, proposing that: **QAZAQ** emerged as a contrasting variant of **QAÇAQ** in multilingual interaction between speakers of Oghuric and varieties of CT, together with Finno-Ugric and Slavic speaking communities on the Pontic-Caspian steppe and in the Volga-Kama region following the social transformations beginning with the Mongol invasions in the 13th century, and that its emergence was intimately connected, as Atlasi suggests, to the formation of a new Tatar identity in this region during this period. In the final section (IV), this paper advances a short case for how the concept of register provides a useful analytical framework for understanding processes of language innovation that can supplement diachronic analysis of systematic sound changes.

² This is not to suggest that everyone has forgotten or rejected Atlasi’s proposal. For example, Rustemov’s (2020: 299) recent glossary of Crimean Tatar legal terminology, published in Kazan, similarly lists ‘qazaq’ as a variant of CT *qaç(g)aq*, showing this theory already has wider acceptance among some Turcologists familiar with Atlasi’s scholarship.

³ A note on transcription: throughout the paper the author uses **QAZAQ** and **QAÇAQ** to refer, abstractly, to the entire range of attested reflex and cognate forms of each word across the Turkic languages and as borrowings into non-Turkic languages. However, when citing a concrete actually occurring variation in a given language or dialect italics are used (e.g., Turkish *kaçak*, Chagatai *qaçaq*, Chuvash *kasak*, etc.). Glosses from primary sources are indicated by single quotations (e.g., Rabghuzi’s *Qışaş al-anbiyā*’ (c. 1310) gives ‘qaçqaq’). In general, I use a standard orthography for the language in question, or an accepted form of transcription (e.g., Birnbaum’s system is used for OAT, Ottoman Turkish, Chagatai, and Old Orthography Tatar), except where it is necessary to indicate a noncorrespondence between phonetic production and orthographic representation, or where individual sound changes are analyzed, in which case Turkish conventions for representing fricatives and affricates are used. Generally, ‘ç’ = [tʃ], [tɕ], [tʂ], ‘c’ = [dʒ], [dʒ], [dʒ], ‘ş’ = [ʃ], [ɕ], [ʂ] (depending on the Turkic language) whereas ‘ts’ and ‘dz’ are introduced to represent [ts] and [dz] respectively. However, citations from secondary sources deploying other transcription systems (e.g., the Turkic Phonetic Alphabet) are quoted as originally written.



I: AN OUTLINE OF THE LINGUISTIC HISTORY OF QAÇAQ

The Common Turkic verbal root *qaç-* ‘flee’ is known from the earliest Turkic texts and attested in both Runic and Old Uyghur inscriptions from the 8th century (Clauson 1972: 589–90; Blagova *et al.* 1997: 340). The earliest known attestation of the derived nominal form **QAÇAQ** occurs several centuries later, in Rabghuzi’s *Qışaş al-anbiyā*’ (c. 1310), an Eastern Turkish version of *The Story of the Prophets*, where it appears as ‘qaçğaq’ in the meanings of ‘fugitive’ or ‘slave’ (Boeschoten *et al.* 1995: 712; Erdal 1991: 392). The *Nişanyan Sözlük* provides a comparable attestation for ‘qaç-qaç’ from another 14th-century Old Anatolian Turkish (OAT) translation of the *Story of the Prophets* by an anonymous author. The word is unambiguously derived from a combination of the verbal root *qaç-* with the CT suffix *-gaq*. In Old Turkic, as Erdal (1991: 391–392) notes, this suffix has three primary functions: forming (1) agentive and (2) instrumental nouns, and (3) forming the names of ailments; ‘qaçğaq’ and ‘qaçqaq’ are of the first type and have close semantic parallels with earlier attested words such as the Karakhanid military terms ‘tut-gak’ (‘mounted patrol’) and ‘tur-gak’⁴ (‘daytime sentry’).

It is unclear from this evidence whether *qaçğaq* and *qaçqaq* derive from an older Turkic etymon that first becomes attested in the Middle Turkic period, or whether they represent a later linguistic innovation among a subset of Turkic speakers that subsequently spread across a wider geography. Unlike its rough semantic equivalent **QAÇQIN**—a word that also means ‘fugitive’ and is attested earlier in Karakhanid Turkic literature such as Mahmud Kashgari’s *Dīwān Lughāt al-Turk* (1074) (Clauson 1972: 592)—**QAÇAQ** is not attested in all branches of Common Turkic. While, as outlined below, modern reflexes of **QAÇAQ** are almost universally present in the Western and Southeastern branches of CT (where they frequently coexist with modern reflexes of **QAÇQIN**), **QAÇAQ**’s reflexes are absent in the Northeastern branch of CT, where only reflexes of **QAÇQIN** are found or the root has disappeared entirely.⁵ This suggests that **QAÇAQ** potentially originated as a Middle Turkic neologism, albeit one based on existing principles of Old Turkic word-formation. The emergence of Khwarezmian Turkic and Old Anatolian Turkish (OAT) as distinct Turkic literary standards in the 13th and 14th centuries is understood by some scholars to mark the transition between the ‘Older’ and ‘Middle’ periods (see Johanson 1998: 84–87), in which Kipchak and Oghuz influences became more pronounced in the written record, and so, **QAÇAQ**’s relatively simultaneous appearance in both Khwarezmian Turkic and OAT in the 14th century is noteworthy.⁶ However, we cannot eliminate the possibility that **QAÇAQ** was more widely in use at

⁴ This term is already attested in Uyghur inscriptions (Erdal 1991: 392–393).

⁵ There are exceptions to this rule. Reflexes of both **QAÇAQ** and **QAÇQIN** exist in Siberian Tatar—as *қауақ* (qatsaq) and *қаққы* (qatsqı) respectively (Pyrerka & Tereshchenko 1948: 14–15)—but the former was likely a borrowing from Kipchak following its emergence as a regional lingua franca beginning in the 13th century (see section III). On the other hand, I have not been able to find a reflex of **QAÇAQ** attested in dictionaries of Karakalpak or Kyrgyz—the easternmost members of the SW branch of Common Turkic.

⁶ Khwarezmian Turkic was a literary idiom that was centered in the Amu Darya River Delta south of the Aral Sea that gradually supplanted the former Karakhanid literary standard across much of Central Asia following the Mongol invasions of the 13th century, incorporating, among other innovations, new influences from Kipchak and Oghuz (Györfi 2014). Old Anatolian developed in the Oghuz *beyliks* that were established following the Turkic expansion into Asia Minor beginning in the 11th century. By the 13th century, OAT had emerged as a distinct literary standard, subsequently developing as the basis for the Ottoman literary language. Like Khwarezmian, OAT adapted established literary conventions of Eastern Karakhanid Turkic to better reflect the (Western Oghuz) dialects spoken in Anatolia, incorporating a large number of words from both Western Common Turkic (Kipchak and Oghuz) and other regional languages like Persian and Greek (Johanson 1998).



an earlier date and subsequently disappeared in some dialects of CT. What this evidence appears to confirm is that the use of **QAÇAQ**—in the form of Middle Turkic *qaçgaq/qaçqaq*—was already widespread by this period, appearing in texts from Khwarazm to Anatolia.

A further question of analytical importance for the argument of this paper thus concerns the timing and distribution of the loss of the *g* in the derivational suffix *-gaq* and the corresponding change *qaçaq < qaçgaq*. In Chagatai, the Eastern Turkic literary successor of Khwarezmian, *qaçgaq* appears to have persisted as a literary form long after this *g* was lost in many regional dialects of spoken Turkic present where Chagatai was used as the literary standard. For example, Abu al-Ghazi Bahadur's uses 'qaçgaq' in his *Şecere-i Terakime* (1659) (Pavet de Courteille 1870: 394), while a century later, Mirza Mehdi Khan's *Sanglax* (1759)—a Persian-language guide to the (Chagatai) Turkic language—only gives 'qaçaq' (Clauson 1960: 66, 267v). This textual juxtaposition might give the impression that the loss of *g* in the *-gaq* suffix was a development that did not occur in Chagatai until the 17th or 18th century, but the precise timing of loss in spoken registers, as outlined below, was likely obscured by literary conventions that sometimes afforded preference to traditional spellings and archaic forms of words over subsequent shifts in pronunciation.⁷ In fact, *qaçgaq* was never entirely replaced by *qaçaq* in early modern Chagatai, but rather came to coexist with the latter as a contrasting, archaic form. Thus Buḥārī's 19th-century Chagatai-Ottoman Dictionary, *Lügat-i Çağatayi Ve Türki-yi Osmani* (1892: 214), lists 'qaçgaq' alongside the newer colloquial variant 'qaçaq'—both of which exist as reflexes in modern Uzbek, i.e., 'qochqoq' and 'qochoq' respectively (Butayev and Irisqulov 2008). In both Chagatai and Uzbek, moreover, these variants seem to function as contrasting forms with slight variation in meaning—a problem of 'enregisterment' addressed in the final section of the paper.

Like in the case of Chagatai, an analysis of the timing of the change *-aq < -gaq* in Western CT is complicated by the fact that *qaç(q)aq* is very poorly attested in Middle Ottoman and Kipchak texts. In Ottoman, various reflexes and derived forms of **QAÇQIN** (e.g., *qaçğın*, *qaçkun*, *qaçkın*, *qaçğuncu*, *qaçkuncu*)—already the established literary variant in Eastern Karakhanid Turkic—appear to predominate during the Middle Period.⁸ Beyond the 14th-century attestation of 'qaçqaq' in OAT noted in the *Nişanyan Sözlük*, the earliest known attestation of **QAÇAQ** in Ottoman occurs in Meninski's *Thesaurus Linguarum Orientalium* (1680: 3632) as 'qaçaq' with the 'ğ' already absent (see Tietze's etymological entry for 'qaçqaq' and other related derived nouns of *qaç*). Unlike in Chagatai, however, the older form (i.e., *qaçgaq* or *qaçqaq*) appears to have been largely lost by the early modern Ottoman period (if it was ever even widely in use). It is not attested in any of the best-known 19th-century Ottoman dictionaries and does not survive in modern Turkish or Azeri (where its reflexes are *kaçak* and *qaçaq* respectively). In the more limited number of Middle Kipchak sources, the picture is even less clear, and to my knowledge, there is no documented attestation of a word like *qaçgaq* or *qaçqaq* in the Mamluk Kipchak literature or the *Codex Cumanicus*.⁹

⁷ Clauson (1972: vii) frames the problem when introducing the organizational framework for his Turkic etymological dictionary thus: 'There is, of course, no means of discovering exactly how Turkish was pronounced between 900 and 1200 years ago; the alphabets employed are all to a greater or less degree unscientific; the most that can be achieved is a broad system of transcription which will give an approximate idea of the pronunciation.' He adds several pages later that '...the Turkish texts are written in a variety of alphabets, all more or less ambiguous, and it is often impossible to determine the correct transcription of a number of words; moreover, some words were pronounced slightly differently in different languages' (Clauson 1972: x).

⁸ *Tarama Sözlüğü* 1974: 2157–2159; Tietze 2016: 34–35.

⁹ Of course, I would be very interested to be acquainted with them if they exist.



However, even in the absence of robust textual attestations from this earlier period, two further considerations—in addition to the 14th-century attestations of its etymon discussed above—allow us to argue that 1) **QAÇAQ** was likely already in use across a wide geography by the time of the first attestations of its etymon in the 14th-century; and that 2) in that many dialects of spoken Kipchak and Oghuz the *-aq* < *-gaq* change was already underway by this period. The first derives from contemporary evidence, namely that the loss of the *g* in modern reflexes of **QAÇAQ** occurs across the Western (Kipchak and Oghuz) and Southeastern branches of CT—and appears to be universal in western CT (cf. Blagova *et al.* 1997: 342). Some modern standards, like Uzbek, preserve both the older literary form and the reduced form **QAÇAQ**. Both *gaçgak* and *gaçak*, for example, are found in modern dictionaries of Eastern Oghuz Turkmen (Khamzaev 1962). In the Kipchak languages, as in Western Oghuz (i.e., Turkish and Azeri), however, the reduced form is predominant. In the modern Kipchak languages, for example, one encounters *качак* (*qaçaq*) in Volga Tatar, *қасақ* (*qasaq*) in Bashkir, *қашақ* (*qaşaq*) in Kazakh, *къачакъ* (*qaçaq*) in Crimean Tatar, and *къачакъ* (*qaçaq*) in Kumyk.¹⁰ Importantly, all of these reflexes of **QAÇAQ** show expected phonetic divergence from common Western Turkic that are evident in a shift in the production of the CT affricate *ç* that began in the early Middle Period (see section III) and are therefore unlikely the results of more recent borrowing between the different languages of the NW branch of CT.

The second point is that, even without a greater number of attestations of **QAÇAQ**, we can derive a better sense of the timing of the change *-aq* < *-gaq* by looking at textual attestations of other words formed by the same derivational suffix, where the evidence points to this change being already underway in some dialects of CT at an early date. Tietze's (2016: 35) etymological dictionary, for instance, compares the Turkish 'kaçak' to Turkish 'çolak' (meaning 'one-armed' or 'maimed'). But Clauson (1972: 419) provides no attestation of a form like 'çol-gak' in pre-13th-century Turkic and the word is already attested without a 'g' as early as the 8th-century, so this comparison is of questionable value. A better comparison can perhaps be made with the evolution of forms like **BIÇ(G)AQ**, **YAT(G)AQ**, **SAÇ(G)AQ**—each of which has an older Turkic attestation where it is spelled with the full *-gaq* suffix. The word **BIÇ(G)AQ** is derived from the CT root *biç-* ('cut') and developed the meanings of 'something cut off', 'segment', and eventually 'corner'. It is attested as 'buçgaq' in 11th-century Karakhanid Turkic but the word appears both with and without the *g* in both Kipchak and Chagatai texts from the 14th-century (Clauson 1972: 294).¹¹ The development of the word **YAT(G)AQ**, meaning a 'resting place' or 'bed'—from *yat-* ('lie down')—exhibits a similar pattern. It is attested in 11th-century Karakhanid Turkic as 'yatgaq', but it is also attested in Ottoman texts from the 14th and 15th centuries as 'yataq' (Clauson 1972: 888), suggesting the loss of *g* was already underway in some Oghuz dialects spoken in Anatolia and the Balkans by this period. The word **SAÇ(G)AQ**, finally, provides the closest phonetic parallel to **QAÇAQ** (cf. modern Turkish *kaçak* and *saçak*) of these three. It is derived from the CT root *saç-* ('scatter') and has developed various modern meanings such as 'tassel' or 'fringe'. Like the other two examples, it is attested in 11th-century Karakhanid Turkic as 'saçgaq', but a reduced form 'saçaq'¹² is already attested in both the late-13th-century 'Interpreters' Book'—the first or

¹⁰ See, respectively, TTAS 2017; Akhmerov *et al.* 1958; Kaliev *et al.* 2005; Aqtay and Jankowski 2015; Bamatov 1960.

¹¹ The word underwent early labial assimilation to produce *biçgaq* < *buçgaq*. Additionally, the *g* was lost and the affricate has undergone *c* < *ç* voicing in many Western Common Turkic reflexes (Clauson 1972: 294).

¹² The original gloss in the CC is 'sazac' (see Drimba 2000).



‘Italian’ part of the *Codex Cumanicus* (Ligeti 1981)—and a Mamluk Kipchak text from the 14th century, as well as in 15th-century Chagatai texts (Clauson 1972: 796), and likewise appears in the early 15th-century Ottoman medical treatise *Yâdigâr* by İbn-i Şerif (Nişanyan 2022). All of this suggests that the *-aq < -gaq* change was part of an ongoing, gradual process that was already widespread by the time of the first attestations of ‘qaçğaq’ and ‘qaçcaq’ in the 14th century. Taken together with evidence about the distribution of **QAÇAQ**’s reflexes in the modern Turkic languages, it is reasonable to speculate the reduced form *qaçaq* already existed as a variant of *qaçğaq* in many spoken dialects of Western CT by this time.¹³

In addition to a diachronic overview of **QAÇAQ**’s phonetic evolution in Common Turkic, the argument of this paper also requires a brief consideration of the different senses in which the word has been used and a corresponding effort to draw distinctions between those meanings that are older and therefore potentially more generalizable, and those which are latter semantic innovations belonging to a particular region or period. In Eastern Turkish, as noted above, ‘qaçğaq’ is first attested in the early 14th century when it was used in a meaning like ‘fugitive’, ‘liable to flee’, or ‘slave’ (Boeschoten *et al.* 1995: 712; Erdal 1991: 392). The meaning ‘fugitive’ is also ascribed to ‘qaçğaq’ by Pavet de Courteille (1870: 394) in his reading of Abu al-Ghazi Bahadur (1659) *Genealogy of the Turkmens*; and Mirza Mehdi Khan (1759) similarly glosses ‘qaçağ’ with the Persian ‘fugitive’ or ‘runaway’ (‘firārī’) (Clauson 1960: 66, 267v). Shaw’s (1878: 140) dictionary of Kashgari Turkic gives ‘qaçcaq’ (transcribed as ‘qûch-qûq’) in the meaning of ‘given to running away’; and Buḥārī’s (1892: 214) Chagatai-Ottoman dictionary defines ‘qaçağ’ with the Ottoman and Persian glosses ‘fugitive’ (‘firārī’; ‘qaçqin’) and ‘escaped slave’ (‘girizende gûlâm, bende’ and ‘abd girizpa’).

As in the case of the analysis of its phonetic evolution above, an evaluation of **QAÇAQ**’s semantic development is complicated by a gap in attestations spanning nearly three centuries. But the consistency of usage between its first attestation in the early 14th century and the modern-day suggests a high degree of semantic continuity. Its modern reflexes in Uzbek and Uyghur convey a comparable range of meanings, from ‘runaway’, ‘fugitive’, ‘deserter’, to ‘refugee’ (Butayev and Irisqulov 2008). The older sense of an ‘escaped slave’, on the other hand, appears to have been largely lost with the abolition of slavery as a formal institution in Central Asia throughout the 19th and early 20th centuries.

The late re-appearance of **QAÇAQ** into the Ottoman textual record is similarly curious—with the attestation in Meninski’s *Thesaurus Linguarum Orientalium* (1680), likewise coming more than three centuries after the first (and last) known attestation of ‘qaçcaq’ in OAT described in the *Nişanyan Sözlük*. Also curious are the different meanings ascribed to ‘qaçaq’ in Ottoman dictionaries over time. The word appears to have been used in the original sense of ‘fugitive’ or ‘runaway’ when first attested in OAT in the mid-14th century (Nişanyan 2022). But Meninski (1680: 3632) glosses ‘qaçaq’ in the meaning of ‘retreat’ (‘fuga’) and, in the verbal compounds ‘qaçaq vermek, qaçaq göstermek’, ‘to feign retreat’ (‘simulare fugam’). However, it is also possible that this was a secondary, military meaning¹⁴ that did not replace, but co-existed alongside an older, colloquial meaning of ‘fugitive’. And that this meaning was largely omitted in Ottoman writing, where

¹³ It is conceivable, for example, that early spellings might have been deliberately modeled on older Karakhanid words like *tutgaq* and *turgaq* and not on contemporaneous colloquial pronunciations.

¹⁴ In early modern Chagatai, for instance, the verb ‘qaçmaq’ (قاچماق) and the derived form ‘qaçiqmaq’ (قاچيقماق) had the secondary meanings of ‘waging a fighting retreat’ (‘harb ve firar etmek’ ‘çekinub savaşmaq’, ‘qaçub savaşmaq’) (Buḥārī 1892: 214)—a common tactic among mounted steppe armies well-known since at least the time of the Mongols.



qaçğun (and its variants) remained as the predominant literary form well into the early modern period.¹⁵ This hypothesis is supported by the widespread inclusion of *qaçaq* in late 19th and early 20th-century Ottoman dictionaries, where it is generally glossed in both the primary meaning of ‘fugitive’ and the secondary meaning of ‘contraband goods’.¹⁶ While the secondary gloss denoting ‘contraband’ likely points to a more modern semantic innovation¹⁷, its gloss as ‘fugitive’ indicates that **QAÇAQ**’s original sense was likely preserved among many Turkish speakers (and speakers of other languages) in Ottoman lands.

Several potential factors might account for **QAÇAQ**’s curious development in the Ottoman textual record: its early attestation of **QAÇAQ** (as ‘qaçgaq’) in 14th-century OAT, its subsequent absence from the textual record of Anatolian Turkish and Ottoman for several centuries, followed by its reappearance in the 17th-century (in a specialized military meaning in a foreign dictionary), and finally, its widespread attestation in 19th-century Ottoman literature, exemplified by its inclusion in many major 19th-century Ottoman dictionaries in the same primary meaning of ‘fugitive’ or ‘flight’ in which it was used in its first attestations six centuries earlier and which it continued to be used across most other parts of the Turkic speaking world.¹⁸ One potential factor for **QAÇAQ**’s reappearance in Ottoman literature in its original meaning in the late-19th century might have stemmed from broader trends within Ottoman literature over the previous two centuries to bring the literary language in closer alignment with vernacular speech registers (Leezenberg 2016). Another potential explanation derives from the importance of the word acquired in giving definition to new popular resistance movements that arose to challenge state and colonial power across the empire during the 19th century—further evidence that **QAÇAQ**’s widespread use in colloquial language had already spread beyond Turkic speakers by this period. The word survives in the Kipchak Turkic languages of the Caucasus, and by the 19th century at the latest, and probably earlier, the word had passed into many regional languages, including the Northeast Caucasian languages of Daghestan (such as Avar, Lak, Bezhta, and Karata) and Georgian, where it is used in the meaning of ‘bandit’ and ‘outlaw’, or as a synonym for the pan-Caucasian term ‘abrek’ (Bobrovnikov 2007: 247, 264; Gould 2016: 252; Grant 2009: 83). The word had a comparable meaning for (Western Oghuz) Azeri speakers to the south, and following Russian expansion into the region during the Caucasian War (1817–1864), the term also became applied to guerrillas who fought against Russian occupation, as exemplified by the Azeri folk hero Qaçaq Nəbi (1854–1896), as well as members of a subsequent proto-nationalist movement (the Gachag Movement) that mobilized against Russian, and later, Soviet rule (Ukudeeva-Freeman 2003: 107). The word exhibits a similar pattern of use in the languages of the Ottoman Balkans—e.g. Albanian *kaçak*, Serbian *kačaci*, Greek *κατσάκ* (*katsak*)—where it originally served as a generalized gloss for local bandits and outlaws (with a secondary meaning denoting ‘smuggled goods’), before emerging in

¹⁵ Also noteworthy is the form *qaçici* which appears to have been formed from the nominal suffix *-ici* (cf. Modern Turkish *seçici*, *geçici*, *biçici*, *et al.*) but this word does not survive in modern Turkish (except sometimes as a last name). It is attested both by Molino in *Dittionario della lingua italiana turchesca* (1641) who gives ‘kacigi’ in the meaning of ‘fugace’ (‘fugitive’) (p. 156); and by Meninski (1680) as ‘qaçici’ (transcribed as ‘kaçigi’) who glosses it with both formal Latin ‘fugax’ and vernacular Latin ‘fugace’ (p. 624). Like **QAÇAQ**, *qaçici* appears to have been a colloquial variant of *qaçğun*, which would explain its appearance in two 17th century foreign dictionaries but its poor attestation in Ottoman texts.

¹⁶ See, for, instance, Kelekian 1911: 922; Sami 1899: 1017.

¹⁷ A semantic innovation that likely developed by analogy with the older literary formulation of ‘gümrükten kaçırılan mal’.

¹⁸ See, for instance, Hindoğlu 1838: 348; Faik 1896: 123; Vefik 1889: 594; and Sami 1899: 1017.



the late 19th Century as a more specific designation for proto-nationalist Albanian rebels of the Kachak Movement who opposed Ottoman rule in Albania, and later, the annexation of Albanian lands by the Kingdom of Serbia following the First and Second Balkan Wars (Gulyás and Csüllög 2015: 229).

Of course, one cannot simply assume that a word's modern semantic range reflects its use historically. In the case of **QAÇAQ**, for instance, the secondary meaning of 'smuggled [goods]' or 'smuggler' (*kaçakçı* in modern Turkish) was a later, albeit widespread innovation that was taken up across the entirety of Ottoman territory and spread east into Iran and Central Asia.¹⁹ But this meaning is poorly attested in the modern reflexes of Northeast and Southeast branches of CT.²⁰ Later semantic innovations, however, do not preclude historical continuity. On the contrary, in cases like that of **QAÇAQ**, where an older meaning of the word is suggested in its first attestations and likewise preserved as the primary meaning in modern usage across a wide geography and perhaps several dozen modern languages, it becomes entirely reasonable to assume a degree of semantic continuity across time and space. And, while subsequent research might reveal further attestations and thereby produce a clearer view of this history, in the case of **QAÇAQ** there is already sufficient evidence to assert that it was continually used in a sense akin to 'fugitive' or 'run away' across much of Turkic speaking world in the seven centuries between its first attestation and the present day.

The argument of this paper begins with the observation that the range of historical meanings given to **QAÇAQ** has very close parallels with the range of meanings that became ascribed to **QAZAQ** over the same period and across an overlapping geography. Like **QAÇAQ**, as noted above, **QAZAQ** is likewise attested for the first time in the early to mid-1300s, roughly a century after the Mongol invasions that precipitated widespread dislocation and completely reshaped political and sociolinguistic conditions on the Eurasian steppe. Like **QAÇAQ**, too, **QAZAQ** designated an amorphous class of 'fugitives', 'outlaws', and 'slaves'. This paper suggests that it was in the wake of these migrations, and the language shifts that resulted, that the word **QAZAQ** first emerged as a contrasting, regionally dominant Tatar form of the Turkic **QAÇAQ** that subsequently spread out from the Pontic-Caspian beginning in the 14th century. The details of this argument are outlined in the third section of the paper, but first, it is necessary to outline the linguistic history of **QAZAQ**—a history that begins, appropriately for this paper's argument, in the same region and period under discussion.

II: AN OUTLINE OF THE LINGUISTIC HISTORY OF QAZAQ

The first attestation of **QAZAQ** is disputed, but in all cases can be dated to the early to mid-14th century and traced to a speech community centered on the vicinity of the Pontic-Caspian steppe. Some Turcologists have proposed that the Cuman gloss 'cofac' (in the meaning of 'guard'²¹) that occurs in the first part of the *Codex Cumanicus*—a trilingual Italio-Latin/Persian/Cuman dictionary produced by Italian traders operating on or around the Crimea and dated to the late

¹⁹ **QAÇAQ** is widely attested in this meaning, for example, in Azeri and modern Persian.

²⁰ In the modern Volga-Kipchak languages such as Kazan Tatar and Bashkir, for example **QAÇAQ** seems not to be used to describe 'contraband' or 'smuggling' but is only deployed in the older sense of a 'fugitive' or 'political refugee' (TTAS: 187).

²¹ 'The glosses are 'guayta' in Vernacular Latin and 'naobat' in Persian (Drima 2000; Kuun 1880).



13th or early 14th centuries—is an attestation of **QAZAQ** (Kuun 1880; Pritsak 2006), although the feasibility of this connection is justifiably questioned by others (Doerfer 1963: 466–468; Drimba 2000: 221; Urban 2017: 758–760). Another potential attestation occurs in a Greek text, also from the Crimea and dated to 1308, where the designation of ‘καζάκων’ (kazakon) is used to describe a group of bandits, but this attestation has not been verified and nearly a century passes between this attestation and the next regional attestation of **QAZAQ** in a meaning like ‘bandit’ or ‘outlaw’ (Urban 2017: 761).

The first undisputed attestations of **QAZAQ** occur in 14th-century Mamluk Kipchak texts from Egypt and the Levant.²² This is significant because—like the contested attestations from the Crimea noted above—it also suggests that the word originated among Kipchaks and other Turkic-speaking peoples living on the Pontic-Caspian steppe, then the primary target of the slave trade linking the Golden Horde to Mamluk Egypt through Genoese merchants in the Black Sea and Mediterranean. As Urban (2017: 762) observes, however, the early Kipchak attestations of ‘qazaq’ in the Mamluk texts are glossed in Arabic in the meaning of ‘unmarried, solitary, alone’ (Ar. ‘mujarrad’)—and in the compound ‘qazaq başlı’—in the meaning of ‘unmarried’ or ‘bachelor’ (Ar. ‘âzib’) and not in the meaning of ‘bandit’ (see discussion in section III). The difference in meaning found in these early attestations and later uses of **QAZAQ** on the Pontic-Caspian steppe are noteworthy, as both Lee (2015) and Urban (2017) observe, raising questions about the common origin of a word that appears to have denoted both a social or kinship status on the one hand, and to have described a brigand or outlaw on the other. But these different meanings are not necessarily suggestive of distinct etymological origins. Rather, they can be read as revealing of different social perspectives and interpretations. After all, as discussed in the next section, the social positions and qualities of ‘unmarried men’, ‘wandering vagabonds’, and ‘outlaws’ can come to overlap in certain sociopolitical contexts.

The interplay of these different meanings becomes clear when we consider **QAZAQ**’s near-simultaneous diffusion north and west into the territories of Muscovy, the Kingdom of Poland, and the Grand Duchy of Lithuania, and southeast into post-Timurid Central Asia beginning in the 15th century (Lee 2015)—a pattern of diffusion that is also highly suggestive of **QAZAQ**’s origin in the vicinity of the Pontic-Caspian steppe and the Volga-Kama regions. In the earliest Muscovite sources from the mid-15th century, the term *казак* (kazak) was used to describe Tatar rebels who sided with Muscovy against the Khanate of Kazan, the military retainers of émigré Tatar princes hosted by Muscovy, individual Tatar exiles or runaways, as well as to brigands (of any ethnic background) who attacked Muscovite traveling diplomats and merchants, and by the turn of the 16th century, the autonomous communities of Don and Zaporizhian Cossacks living on the Pontic steppe (Lee 2015: 44–45). **QAZAQ** also appears in Polish-Lithuanian sources beginning in the mid-15th century, in the Ukrainian form *kozak*, where it was deployed in various meanings that Lee (2015: 43–44) suggests likely reflected Tatar usage, including: ‘fugitives, expellees, and robbers’, ‘propertyless people’, ‘raiding’ and ‘guerrilla warfare’, and ‘slave’, ‘vagrant’, or ‘freebooter’. Therefore, while **QAZAQ** was used in these sources in a sense like ‘bandit’ or ‘outlaw’, the word was also deployed in a wider range of meanings.

Doerfer (1963: 464) makes a similar observation, noting that when ‘the word appears for the first time in Russia in the second half of the 15th Century... [*qazaq*] did not have a purely military

²² The oldest uncontested attestation occurs as ‘qāzāq’ in the *Kitāb Majmū‘ Tarjumān Turkī wa-‘Ajamī wa-Muḡalī wa-Fārsī* which Ermers (2017: 32) dates to 1342.



meaning but also denoted families and dispossessed individuals regardless of whether or not they were nomadic or engaged in banditry' (author's translation). Some *qazaqs*, as Doerfer (1963: 464) observes, lived as 'free laborers' rather than as serfs. However, many such persons were likely 'free' in Marx's 'double sense': they were free of feudal obligations and clan or tribal affiliations, and consequently, were also free of any legal claim to grazing land or movable property, and were thus forced to survive, if not by banditry or begging, then by temporary wage work or semi-permanent servitude in foreign communities. This earlier meaning appears to have been preserved in some modern senses of **QAZAQ** in languages across a wide geography, from the Caucasus to the Urals. Both modern Kumyk *Къазакъ* (*qazaq*) and Komi *казак* (*kazak*) each have the double modern meanings of both 'Cossack' on the one hand, and 'farm worker' or 'seasonal labor' on the other, a connection further attested in the Komi words *казакавны* (*kazakavny*) meaning 'seasonal labor' and *казакӧдны* (*kazakedny*) meaning 'forced labor' (Beznosikova *et al.* 2003)—'farmhand' also being a secondary meaning of *казак* (*qazaq*) in Russian that is attested as early as the late 16th century (Urban 2017: 791). In the 17th and 18th-century court registers of the Crimean Khanate, moreover, the word *qazaq* could simply denote a 'slave' (Królikowska-Jedlińska 2018: 74); and in modern Karachay, a *kazak* not only denotes a 'tramp' or 'vagabond', but also retains an older sense referring to 'the lowest class of purchased slaves in the social structure of old Karachay-Malkar' society (Takvul 2000: 251).²³ Originally, therefore, the term *qazaq* seems to have designated a more generalized form of social personhood that, over time, became associated with different modes of livelihood or legal status (e.g., itinerant wage labor, banditry, mercenary service, enslavement). By the end of the 16th century, **QAZAQ** was also used to refer to the autonomous political communities, like the Don, Tatar, and Zaporozhian Cossacks, that had formed from diverse groups of refugees, political fugitives, and escaped serfs inhabiting the shatter zone between the Crimean Khanate to the south, Muscovy and Poland-Lithuania to the north and west, and the disintegrating Golden Horde to the east—a region of Pontic steppe north of the Black and Azov Seas that came to be known as the 'Wild Fields' in Polish, Lithuanian, and Russian sources beginning in the 16th century (Lee 2015: 74–78).

An analogous process of diffusion occurred toward the southeast, into post-Timurid Central Asia, where **QAZAQ** is similarly first attested in the Persian and Chagatai texts over the 15th century. By the following century, **QAZAQ** had become established as a multilingual gloss for those engaged in political vagabondage, or *qazaqliq*, which Lee (2015) argues was driven in part by the institutionalization of this practice across the region in the wake of the consecutive disintegration of the Mongol and Timurid Empires. This new institution, as Lee (2015: 48–50) describes, was divided into three stages: 1) a flight or separation from one's tribe or polity; 2) a period of political brigandage on the frontier intended to gather further resources and manpower; and 3) (in successful cases) the establishment of political alliances and the eventual contestation of dynastic power. Babur, the founder of Mughal Dynasty in Northern India, later described his own period of youthful *qazaqliq* in his early 16th-century autobiography, the *Baburnama*, and post-Timurid 16th-century narratives from the same period also retroactively refer to Timur's period of political exile in his youth as a period of *qazaqliq*, although the term was not yet widely adopted in Timur's lifetime and he is not referred to as such in contemporary sources, suggesting that the term was only popularized in post-Timurid Central Asia over the 15th century (Lee 2015: 71). Like on the

²³ Like **QAÇAQ**, **QAZAQ** was borrowed widely across the Caucasus, as variants of **QAZAQ** have passed into many regional languages, including Georgian, the Dagistanian languages, and Avar (Golden 2011: 257–258)



Pontic steppe, moreover, **QAZAQ** also became used to identify new steppe polities whose origins lay in intradynastic civil war and political separation from an existing state. One of the best-known of these ethnopolitical formations emerged in the second half of the 15th century, following a series of rebellions against the Uzbek Khanate. This conflict eventually divided the Uzbek tribes between the Qazaq Uzbeks and the Shaybanids, laying the foundations for the emergence of the Kazakh Khanate.

The word *qazaq* is also attested in Old Anatolian Turkish, albeit in a distinct, colloquial sense of ‘traveling companion’ or ‘friend’ (Kanar 2011: 435)—a meaning I elucidate further in the next section. However, this was likely a borrowing from neighboring Turkic-speakers and does not appear to have survived in Ottoman. The Ottomans came into contact with the institution of *qazaqlıq* as they expanded their territories to the northern shores of the Black Sea (Veinstein 1999). The Ottomans had established themselves as a major player in the politics of the Pontic steppe by the end of the 15th century, following conquests in the southern Crimea beginning in the 1470s and their alliance with the Crimean Khanate (subsequently an Ottoman vassal state), in whose domestic and regional politics the various Cossack bands occupying the Pontic steppe played an important role. In this context, military alliances with or against various **QAZAQ** groups became an important feature of Ottoman strategy in this region, as seen in the example of the Ottoman’s punitive military expedition against the Ioan Voda, the rebellious Voivode of Moldavia, and his Zaporozhian Cossack allies in 1574 (Berindei 1972). Ottoman *qazaq* is already attested in a contemporary account by the Ottoman author Remmal Hoca of the life of Sahib I Giray (Khan of both Crimea and (briefly) Kazan in the early 16th century) (Lee 2015: 177). And in the account of his travels in the Caucasus, Russia, and Crimea the following century (between 1640–1670), Evliya Çelebi frequently returns to the topic of the different ‘qazaq’ political factions and military contingents active around the Crimea and the northern Black Sea littoral from Moldavia to the Caucasus.²⁴ Despite the Ottomans’ obvious familiarity with the term, however, the meaning of *qazaq* appears to have remained limited to a designation for Cossack groups north of the Black Sea, and unlike in post-Timurid Central Asia, never came into use as a generalized term for bandits or outlaws across the whole of Ottoman territory.

As outlined in the previous section, however, by the 19th century at the latest (and likely for some time earlier in colloquial language) another term designating an analogous class of outlaws, rebels, and fugitives was in use across the Ottoman Empire and beyond: the word *qaçaq*. However, whether the close semantic parallels between **QAÇAQ** and **QAZAQ**, both historical and modern, are a product of historical coincidence (i.e., an instance of false cognates) or whether they reflect a common etymological origin, remains to be established. So far this paper has shown that **QAÇAQ** first appears as ‘qaçgaq’ and ‘qaçqaq’ in the early 14th century, and has been consistently attested in the meaning of ‘fugitive’ until the present day; and it has further demonstrated that the sound change that produced *-aq* < *-gaq* was already underway among at least some Kipchak and Oghuz speakers even prior to its first attestation. It has also shown that **QAZAQ** likewise first becomes attested in 14th-century Mamluk Kipchak texts, albeit in a seemingly different set of valences, before emerging over the following century as a generalized, multilingual term for a new class of refugees, slaves, and outlaws across Eastern Europe and Central Asia, in a pattern of diffusion that strongly suggests an origin somewhere in the vicinity of Pontic-Caspian steppe. To complete the argument for the common etymology of **QAÇAQ** and **QAZAQ** it is still necessary to provide

²⁴ Çelebi brings up the topic dozens in recounting his multiple trips to the region, see Volumes 2-3, 5, and 7.



both 1) a plausible account of the sound change that produced **QAZAQ** < **QAÇAQ** in the period and region under consideration; and 2) a more robust philological analysis, drawn from historical and contemporary evidence and supported by current scholarship on this region and period, that places the semantic development of both terms within a coherent historical and sociolinguistic narrative—the twin tasks to which this paper turns its attention in the next section.

III: THE CASE FOR THE COMMON ETYMOLOGY OF QAÇAQ AND QAZAQ

Atlasi advances the argument for the common etymology of **QAÇAQ** and **QAZAQ** only briefly, at the beginning of a section on the origins of the Cossacks in his book on the history of Siberia, *Seber Tarihi* (1911: 46), noting that ‘the word *qazaq* is Turkish and its origin are the words *qaçaq*, *qaçgaq*, having the meaning of *qaçqin* (‘fugitive’).²⁵ Atlasi subsequently explains that **QAZAQ** was originally a Tatar word that was applied to refugees who lived on the borderland between Russia and the Tatars (i.e., between Muscovy and the Tatar-controlled Volga and Crimea regions) and was later borrowed by Russian officials as well as these refugees themselves as a term of self-designation. As evidence Atlasi cites Buḥārî’s (1892) Ottoman-Chagatai dictionary, comparing the latter’s glosses for ‘qaçağ’ (see section I) and ‘qazaqci’ (‘thief’, ‘bandit’, ‘highway robber’, *et al.*).²⁶ But Atlasi’s intuition of an etymological connection between the two words was probably derived less from his knowledge of the use of these words in literary Chagatai and more from what one can assume was his intimate familiarity with the meanings and different pronunciations of both lexemes in the multilingual society of greater Tatarstan. One recent comprehensive dictionary of modern Kazan Tatar, for instance, defines ‘качак’ (*qaçaq*) as 1) ‘one who has left their country in flight from enemies, etc.’; 2) ‘in the state of flight’; and 3) ‘fugitive’.²⁷ In defining ‘казак’ (*qazaq*), in turn, the same dictionary, gives the following primary meaning: ‘a *qaçaq* or *irekli keşe* (‘free man’) who escaped from serfdom to the periphery of the Russian state during the 15th–17th centuries (Don, Ural (‘Yayıq’), or Zaporozhian).’²⁸ Beyond simply the semantic parallels already noted above, however, Atlasi was likely also keenly aware of how distinctions in the phonological systems of languages spoken in the vicinity of greater Kazan —distinctions not only between the different dialects of Volga Kipchak but also Oghuric Turkic, Eastern Slavic, and the Finno-Ugric languages—rendered such a transformation possible.

In addition to transforming political and social life on the Pontic-Caspian steppe, the Mongol invasions of the 13th century set off a wave of migrations that altered the social composition and sociolinguistic realities of the Volga-Kama region, as Sinor (1999), Golden (1992) and Zimonyi (1992) outline: Before the Mongol invasions, the southern region of the Pontic-Caspian was controlled by a tribal federation of predominantly Kipchaks and Cumans, while the Volga-Kama region to the north was under the control of the Bulgar Kingdom whose nobility was primarily Oghuric Turkic speaking and Muslim. However, the sociopolitical situation changed dramati-

²⁵ ‘Qazaq süze terekçe bolub anıñ asılı qaçqin meğnesenı bolgan qaçaq, qaçgaq kälimälarıdır.’

²⁶ Atlasi (1911: 46) lists ‘hırsız; harâmi; quttâ-i tarik; oğru; rāhzan, çapaval (?)’ (cf. Buḥārî 1892: 219). Atlasi omits ‘qaltabân’ (‘coward’) from his list and corrects Buḥārî’s spelling of ‘hırsıs’.

²⁷ ‘1) Дошманнардан һ.б. качып, иленйортын ташлап киткән кеше. 2) Качып йөрү хәле. 3) качкын’ (TTAS: 187).

²⁸ ‘XV–XVII йөзләрдә крепостной изүдән рус дәүләте чикләренә (Дон, Жак, Запорожье) качып киткән ирекле кеше, качак’ (TTAS: 44–45).



ly beginning in the early 1200s. A Mongol army moving north through the Caucasus defeated an alliance of Kipchak-Cumans and Russian princes at the Battle of the Kalka River (1223), briefly occupying the southeastern portion of the Pontic steppe and raiding as far north as Novgorod and Bulgar, before being recalled into Asia. The Mongols returned to the region a decade and a half later, following the death of Genghis Khan and the division of his empire between his sons. In 1237, Batu Khan—son of Jochi, grandson of Genghis, and leader of the recently constituted Golden Horde—dispatched an army under the leadership of the Mongol general Möngke to assert Mongol control of the Pontic-Caspian steppe. Möngke's army broke the last vestiges of Kipchak resistance north of the Black Sea before moving north to destroy the Bulgar kingdom the same year. Following his consolidation of political power, Batu's army subjugated all of the peoples of Bulgar—including Oghuric Turks (Bulgars), Mordvins, and Bashkirs—before advancing west, sacking Kiev in 1240 and advancing as far west as Poland and Pannonia by 1241. The Mongol conquests produced extensive social upheaval, redrawing ethnopolitical and linguistic boundaries in the region, as Golden (2009: 112) describes:

The Mongols, particularly concerned about resistance from other steppe peoples, dealt harshly with them. Thus, in the campaign of 1221, conquered Qipchaq-Qangli were denationalized by being forced to cut their hair according to Mongol fashion and then distributed among the Mongol military formations. The survivors of these wars, Mongolic and Turkic, were scattered among the Chinggisid armies of Inner Asia subsequently giving rise to clans or retribalized groupings bearing these names among post-conquest Turkic peoples. A similar fate awaited lesser tribes (including Mongols) that were melded into larger military units of diverse ethnic and tribal origins. Their names (e.g. Jalayir, Barulas) became sub-unit or clan names. The dispersal and the allotment of the tribesmen (and others) to the personal armies of Chinggisid princes became the norm. Polyglot units consisting of Uighurs, Qarluqs, Türkmen, Kāshgharis and Kuchians' were not unknown.

The intermixing of detribalized people into new sociopolitical units was closely linked to larger linguistic shifts among the conquered peoples of the Mongol Empire. In Volga Bulgaria, for example, Golden (2009: 114–115) describes a process of 'Qipchaqization' that occurred in the wake of the Mongol expansion, with some Oghuric-speaking groups moving northward and intermixing with Finno-Ugric-speaking communities, creating the foundation for Chuvash ethnogenesis, while local Bulgar elites became bilingual in Oghuric and Kipchak Turkic, only gradually shifting to Kipchak between the 14th and mid-16th centuries. This process of ethnolinguistic mixing eventually produced the Volga Tatars of Kazan and Astrakhan and the Siberian Tatars further east. 'A similar process of Qipchaqization', Golden (2009: 116) observes, 'took place among the neighboring Bashkir (Bashqort) people, a union of complex ethnic origins deriving from Oghuric, Common Turkic, Finno-Ugric, and Iranian elements.'

These sociolinguistic shifts were likewise reflected in phonetic transformations of the northern Kipchak languages, whose phonological systems were influenced by close contact with Oghuric Turkic and the Finno-Ugric languages (Baitchura 1975; Róna-Tas 1982b; Johanson 2000). This is seen, as Róna-Tas (1982b: 435–439) describes, in a shift in the Common Turkic affricate ζ among the northern (NW and NE) Turkic languages from western Russia to Southern Siberia, where the changes $\zeta < \zeta$ (e.g., Tuvan, Kazan Tatar, and the northern dialects of Kazakh, Nogai, Karakalpak), $ts < \zeta$ (e.g., Misher Tatar and several dialects of Siberian Tatar), and $s < \zeta$ (e.g.,



Bashkir and Khakas) are observed. Although the precise timing of these shifts is difficult to determine, and in any case likely occurred gradually and at different rates among different speech communities, Róna-Tas argues with reference to Mongolian loanwords that these shifts began no later than the 13th–14th centuries.

In the case of Bashkir, for example, the shift in the affricate was accompanied by a breakdown in the historical phonological opposition between voiced/unvoiced fricative pair *z/s* that existed in Common Turkic and a corresponding shift in the articulation of the voiced fricative *z* > *ž*.²⁹ Consequently, the modern Bashkir reflexes of **QAČAQ** and **QAZAQ** are *қасақ* (*qasaq*) and *қазақ* (*qadaq*) respectively, with the word *қазак* (*kazak*)—adopted to distinguish between the modern meanings of ‘Cossak’ (/kazak/) and ‘Kazakh’ (/qadaq/)—being a later borrowing from Russian. These developments are notable because they show evidence of two phonetic transformations in the Common Turkic affricate—spirantization and post-alveolar fronting (i.e., *ç* > *ts* > *s* or *ç* > *ʃ* > *s*)—required to account for the change **QAZAQ** < **QAČAQ**. However, the systematic sound shifts observed in the northern Kipchak languages do not exhibit voicing (i.e., *s* > *z*) and therefore cannot by themselves account for this change. It is precisely owing to the absence of such a *Lautgesetz* among the Kipchak languages that Pelliot (1949: 217–222)—one of the few European Turcologists to have considered the possibility of an etymological connection between **QAČAQ** and **QAZAQ**—casts doubt on this possibility while also acknowledging the otherwise noticeable semantic and phonetic parallels between the two words.

However, there remains another possibility, namely that a transformation like *qazaq* < *qasaq* occurred not within the Kipchak languages themselves but reflected a process of mutual borrowing and interchange among Kipchak speakers and neighboring speech communities. The Oghuric Turkic languages spoken in the vicinity of the upper Volga region, notably, did not have a phonological system that distinguished between voiced and unvoiced consonants.³⁰ However, one of the distinguishing features of the development of Oghuric languages following the Mongol invasions, as documented by Agyagási's (2019: 187) analysis of the development of Early Middle and Late Middle Chuvash (13th–18th centuries), is the secondary voicing of intervocalic voiceless consonants—a ‘lengthy process’ that ‘gradually [affected] the different consonants and covered the whole Middle Chuvash period.’ Under such conditions, it is entirely conceivable that a northern Kipchak variant of **QAČAQ**—such as Bashkir *qasaq* or Mishar Tatar *qatsaq*—could have been borrowed into Middle Chuvash (and the other Middle Oghuric dialects) in a form like **qasaq/xasax*. However, owing to the secondary voicing of intervocalic consonants, the phonetic production of this borrowing by Oghuric speakers would have been perceived by speakers of other regional languages (Kipchak Turkic, Eastern Slavic, Udmurt, Mari, *et al.*) as [q:zaq]—since their phonological systems continued to draw contrasts between voiced and unvoiced fricatives.

In this proposal, **QAZAQ** first emerged as an Oghuric-influenced variant of **QAČAQ** among speakers of different dialects of Kipchak, Finno-Ugric, and East Slavic spoken in the upper Volga-Kama region. In its origin, therefore, **QAZAQ** was—as Atlasi argues—a ‘Tatar’ word that appeared alongside the emergence of a new Tatar identity that formed from the different ethnolinguistic groupings living in the Volga region in the century following the Mongol conquests

²⁹ Today *z* only occurs in modern Bashkir in loanwords (Poppe 1964: 10), while Common Turkic *s* has merged in some environments with *h*, cf. Bashkir *hat-* (‘sell’) with CT *sat-*, or has been preserved as *s* and subsequently merged with the CT affricate, cf. Bashkir *ses* (‘hair’) with CT *saç*.

³⁰ It is also worth noting here, for example, that Middle Mongolian did not distinguish between voiced and unvoiced fricatives, and consequently, **QAZAQ** was borrowed into the language as *xasag* (i.e., with an unvoiced *s*).



(Schamiloglu 1990). Our knowledge of the sociolinguistic situation in this region and period is derived from a limited number of direct observations in contemporaneous sources, as well as from efforts to reconstruct the history of language contact through an analysis of philological and modern linguistic data. This knowledge is buttressed by our understanding of local political developments and their bearing on the longer-term trajectory of language shift in the region. As discussed above, two major trends can clearly be deduced from the historical record: 1) the breaking up and intermixing of formerly distinct tribal groups and, consequently, a new situation of language contact and convergence; and 2) the replacement of Volga Bulgar by Kipchak as both the regional *lingua franca* and the language of state sometime between the creation of the Golden Horde and the loss of its western territories to the newly independent Crimean and Kazan Khanates two centuries later. This raises the question of why an Oghuric-influenced **QAZAQ** would emerge as the dominant regional variant of Kipchak **QAČAQ**, especially when evidence suggests that of primary direction of lexical influence in the region was from Kipchak into Oghuric and Finno-Ugric (Johanson 2000)? Although we cannot be certain, the available information is sufficiently robust to develop several lines of speculation. Two points are important here: the first concerns the local status of Oghuric Turkic in the former territories of Volga Bulgaria, and the second concerns the phonological parameters of Volga Kipchak as it coalesced in this period. Although affected by the shift toward Kipchak, it appears that elements of the Bulgar elites—the ruling aristocracy, merchants, and ulema—retained some degree of their former prestige and influence following the Mongol conquests; and archaeological evidence suggests that Volga Bulgarian persisted as written language alongside Kipchak in the region well into the 14th century (Khalikov & Muhametshin 1977; Hakimzjanov 1986). This is potentially explained by what Golden (1992: 298) notes was the growing importance of Islam to political and military life in the Golden Horde—first under first Berke (r. 1257–1266) and then again under Özbek (r. 1313–1341)—and the strategy of Mongol rulers to cultivate the loyalty of Muslim ulema and traders among their subject populations. ‘As long as the local religious institutions did not become focal points of resistance,’ Golden (1992: 297) notes, ‘the Mongols were prepared to not only support them, but grant them important privileges.’ Given that substantial portion of the local Muslim clergy and bureaucracy was likely initially Bulgar-speaking or multilingual, some residual linguistic influence is expected. Beyond the urban centers, moreover, the Middle Oghuric dialects spoken by non-Muslim communities on the right bank of the Volga retained a dominant status vis-a-vis Finno-Ugric and Kipchak, as is evidenced by both Chuvash’s survival and its significant influence on Western Mari (Johanson 2000).

At the same time, the shift in the Volga Kipchak affricate and the resulting variation in its articulation among the regional Kipchak dialects meant that the standard literary representation of the affricate diverged significantly from local pronunciations. In fact, there is seeming less of a difference between the pronunciation of Bashkir *qasaq* [qa:saq] and Chuvash *kasak* [ka:zak], than between either word and literary *qačaq* [qa:tʃaq]—the standard form as it developed Khawazmian and Chagatai, and later in Kazan Tatar—and it could be argued that **QAZAQ** is as much a Volga Kipchak variant of **QAČAQ** as Oghuric. It is therefore conceivable that the reborrowing of **QAZAQ** by Kipchak-speakers in the historical territories of Volga Bulgaria was precipitated by its simultaneous uptake by other language communities in the region as well as an instance of sociolinguistic differentiation from Kipchak speakers further south. At any rate, borrowings from Middle Oghuric into Kipchak are not unknown (Agyagási 2002).



A comparable process of mutual influence and borrowing, notably, is outlined by Keenan (1980) in his analysis of the etymology of ‘Kazan.’ Keenan argues that the toponym Kazan was likely derived from a hypothetical East Iranian word like **qadzanj* (meaning ‘crooked place’ or ‘bend’ [in the Volga River]) that was subsequently borrowed into Oghuric Turkic in a form like **Xasanj/Xosanj*³¹ (cf. modern Chuvash *Xusan*). Owing to intervocalic voicing of consonants in Middle Oghuric dialects, however, the name subsequently passed into the Kipchak languages as *Qazan* and into Udmurt and Mari as *Kuzon* and *Ozon* respectively (i.e., with a voiced fricative *z*). In the case of **QAČAQ**, this paper proposes that the influence was bidirectional: a northern Kipchak variant of **QAČAQ** (like *qasaq* or *qatsaq*) was first borrowed into Oghuric languages where it subsequently underwent intervocalic voicing. In this theory, **QAZAQ** emerged as the regionally predominant, multilingual Tatar variant of **QAČAQ** that was subsequently borrowed into regional languages and reborrowed into the Kipchak Turkic dialects as a kind of *Rückwanderer*. If correct, this proposal might also account for the lack of attestations of **QAČAQ** in Middle Kipchak, since *qazaq* likely served as the primary local literary form alongside distinct dialectal reflexes of **QAČAQ** in spoken Volga Kipchak, e.g., *qačaq*, *qaşaq*, *qatsaq*, *qasaq* (all of which are attested in the modern Kipchak languages of the region).

This proposal also finds support in the evidence from modern Chuvash—the only extant Oghuric language—where **QAZAQ** was borrowed on at least two distinct occasions. We can distinguish between the timing of these borrowings based on their contrasting phonetic structures in the contemporary language: the older borrowing—modern Chuvash *хусак* (*xusax*), from Middle Chuvash **xasax*³²—exhibits expected phonetic transformations from Volga Bulgar and Middle Chuvash, while these changes are not encountered in the later borrowing *касак* (*kasak*), suggesting that they were already completed when the word entered the language. Both words, importantly, can be used in the colloquial meaning of ‘Cossack.’³³ But beyond this shared valence, both words also have different primary meanings and therefore are not synonymous in the modern language.

Modern Chuvash *xusax* denotes a 1) ‘unmarried man,’ ‘bachelor,’ or ‘single person’ (‘холостяк,’ ‘холостой,’ ‘одинокий,’ ‘холостяком’) (Fedotov 1996); and it was borrowed in this meaning into both Mari (i.e., *ozak* ‘unmarried man’) and the Kipchak languages of the Pontic-Caspian steppe, as evidenced in the 14th-century Mamluk Kipchak glosses noted above—as well as its continued use in this and a range of (semi)-related meanings in some dialects of modern Turkish.³⁴ Modern Chuvash *kasak*, in contrast, denotes both more specifically a ‘Cossack’ (‘казак’), and more generally a ‘fugitive’ (‘беглец’) and ‘vagrant’ (‘бродяга’); and the verbal compound ‘*касакра şüre*’ (‘касакра сүре,’ gloss: *kasak*-LOC move) means to ‘wander,’ ‘loiter,’ or ‘go about aimlessly’ (‘бродяжничать,’ ‘слоняться,’ ‘шляться прост’) (Skvortsov 1985). Modern Chuvash *kasak* can be further connected to the modern Chuvash word *каскайн* (*kaskän*) which is used in a wide range of meanings including a ‘vagrant’ (‘бродяга’), ‘fugitive’ (‘беглец’), ‘runaway’ (‘бежать’),

³¹ Kenaan posits **Xosanj* but the change *o < a* was a development during the Middle Oghuric period and seems to have occurred only after it was borrowed in Kipchak. In either case the basic argument about voicing remains unchanged.

³² See, by way of comparison, Agyagási’s (2019: 118) analysis of the development of Turco-Mongolian *qaraq* ‘bandit’: *qaraq* > **xarax* (Middle Chuvash) > *xurax* (modern Chuvash).

³³ However, the later borrowing seems to be more common in academic and official registers.

³⁴ See Derleme Sözlüğü 1993: 2709–2710. Since in Ottoman Turkish, as noted above, *kazak* did not develop into a generalized term for a kinship status or type of social person, the preservation of this meaning in some dialects of Turkish is potentially explained by later migrations of Turkic speakers from the Caucasus and Crimea to Anatolia.



‘уберать’), ‘rogue’ (‘пройдоха’), ‘libertine’ (‘распутник’; ‘развратник’), someone who is ‘debauched’ (‘распутный’; ‘развратный’), or ‘straying’ (‘отбившийся’)—e.g., ‘kaskān jytā’ (‘stray dog’)—and to denote a person or animal that is ‘restless’ (‘непоседливый’) or ‘perky’ (‘бойкий’); and the compound ‘kaskānra šüre’ (‘касқанра шыре’, gloss: kaskān-LOC move) means to ‘wander about doing nothing’ (‘шататься без дела’) (Skvortsov 1985; Fedotov 1996; Egorov 1964). This same word was also borrowed into Udmurt as *кас̄кы* (kas̄ki) and Mari as *кас̄ке* (kaske) in the meaning of ‘fugitive’ (‘беглец’). Owing to their clear phonetic and semantic correspondence, moreover, both Chuvash *kasak* and *kaskān* can also be confidently identified as borrowings of the local vernacular Kipchak reflexes of *qaçaq* and *qaçqi(n)*³⁵ respectively. This is significant because it points to a clear link between **QAZAQ** (modern Chuvash *qasaq*) and regional reflexes of **QAÇAQ**.

Here it is useful to provide a brief overview of the proposed timeline of these borrowings. While contact between Oghuric and Kipchak speakers in the Upper Volga region likely predated the Mongol invasions, this contact intensified after the Mongol invasions in the 13th century, the same period to which Róna-Tas (1982b) dates the shift in the CT affricate in the north Kipchak languages. This is when a local Kipchak reflex of **QAÇAQ** was likely first borrowed into Volga Bulgar Oghuric in a form like **qa(t)saq*, through whose mediation the word passed back into Kipchak (and other regional languages) as *qazaq*. In addition to secondary voicing, the word underwent two additional (and expected) phonetic developments in Chuvash over the following two centuries: the spiritization of *q* (i.e., *q > x*), and later, a shift in the articulation of the vowel (*a > u/o*³⁶). The process of spiritization was largely completed by the end of the Volga Bulgar period in 1200s, and the vowel shift occurred in Middle Oghuric following contact with Mari migrants to the Volga region in the aftermath of the Mongol invasions (Agyagási 2019). If one accepts that a Kipchak reflex of **QAÇAQ** was the origin of modern Chuvash *xusax*, then this borrowing can be placed within a narrow window of time around the 13th century. Indeed, when one compares the reflexes of **QAZAQ** in modern Chuvash, Mari, and Volga Tatar with those of **KAZAN**—(C) *xusax* (M) *ozak* (VT) *qazaq* : (C) *Xusan* (M) *Ozon* (VT) *Qazan*—the close parallels in the development of both sets of reflexes are apparent, suggesting that they were both borrowed from Volga Bulgar into Mari and Kipchak around the same time (i.e., around the 13th or 14th century when both groups first migrated to the region and Kazan first emerged as important political and economic center in the region).³⁷

A local Kipchak reflex of **QAÇAQ** was then subsequently borrowed a second time into Chuvash in the form *kasak*, perhaps as early as the end of the Early Middle Chuvash Period (15th century), and likely during the same period that Chuvash borrowed the related form *kaskān*. Unlike *xusax*, the form *kasak* did not undergo spiritization or a vowel shift, although secondary voicing of *s > z* still occurs, and the word is sometimes rendered with a ‘z’ in modern Chuvash orthography (likely under the influence of Russian spelling). It is probable that the semantic distinctions between these two forms—*xusax* and *kasak*—were not yet as rigid in this earlier period, as evidenced by the fact that both are still used in the sense of ‘Cossack’ in the modern language. However, it is clear that over time the older form retained a meaning like ‘single’ or ‘unmarried’ (as evident in

³⁵ Fedotov (1996: 235), citing Egorov (1964), also links Chuvash *kaskān* to Kipchak *qaçqi(n)*. Compare with Kazan Tatar reflexes *качкын* (qaçqin) and *качкы* (qaçqi) and Bashkir reflexes *касқын* (qaskən) and *касқы* (qaska).

³⁶ Fedotov (1996) provides both the standard spelling ‘хысак’ and the alternative forms ‘хосак’ and ‘хосак’. In the case of word final *q* the sound change *q > x* was not universal across all of the territory of Middle Oghuric (Agyagási 2019: 73).

³⁷ For a discussion of the Mari migration see Agyagási’s (2019) appendix.



both modern Chuvash *xusax* and Mari *ozak*) while the later borrowings came to possess a meaning like ‘fugitive’ or ‘vagrant’.

When considering the linguistic data from Chuvash and Mari, the problem of **QAZAQ**’s etymology thus appears in a new light. As outlined above, there is ample reason to believe that *kasak*—the modern Chuvash word for **QAZAQ**—was borrowed from a local Kipchak reflex of *qa-çaq*. But what of the earlier Chuvash borrowing, *xusax*, and its meaning of ‘unmarried man’? One possibility is that **QAZAQ** derived from two distinct sources which merged as homonyms into a single word in the process of passing through Oghuric into Middle Kipchak, in which case the timing of the borrowing of *xusax* discussed above would require revision. Even in this case, however, the argument for the etymological connection between **QAZAQ** and **QAÇAQ** would remain largely unchanged, since it was primarily in the meaning of **QAÇAQ** (i.e., ‘fugitive’ or ‘runaway’)—and not ‘unmarried man’—that **QAZAQ** was borrowed into various regional languages spoken in the vicinity of Greater Kazan, and subsequently spread across Eastern Europe and Central Asia. However, this explanation does not account for the use of both modern Chuvash *xusax* and *kasak* in the same meaning of ‘Cossack’, and it still leaves open the question of the origin of the earlier form, as no other source for the word *xusax* is currently forthcoming.³⁸ This paper argues, however, that both Chuvash words are derived from the same Kipchak-Turkic etymon **QAÇAQ**, and that the difference in the meanings—an analogous difference in meaning to the one encountered between the 14th-century Mamluk Kipchak glosses of *qazaq* as ‘single’ or ‘unmarried man’ and its subsequent 15th-century attestations in Eastern Europe and Central Asia as ‘runaway’, ‘bandit’, ‘exiled prince’, and ‘slave’—is not the product of different etymological origins but the sociopolitical conditions under which they were first borrowed and used. In developing this argument, it further seeks to construct a plausible narrative that can account for the semantic evolution of the reflexes of **QAÇAQ** and its contrasting variant **QAZAQ**, here also recognizing that while such a narrative can be grounded in historical and philological evidence it must ultimately remain speculative, and therefore open to contestation and reinterpretation.

It is helpful to begin this discussion by considering what precise social status or form of personhood was intended by the label **QAÇAQ** among Kipchak speakers in Medieval Cumania. It is observed that among some of the modern descendants of Common Turkic the verbal root *qaç-*, in addition to the literal meaning of ‘flee’ or ‘run away’, has a secondary meaning of ‘break away’ [from one group] and/or ‘seek refuge’ [with another]. In his Ottoman-Chagatai dictionary, for example, Buḥārī (1892: 214) lists the following meanings for Chagatai *qaçmaq*: ‘fight and retreat’

³⁸ Collinder’s (1955: 15) etymological dictionary of the Finno-Ugric languages contains an entry for a word meaning, inter alia, ‘bridegroom’, ‘young man’, and ‘servant’ that is evidenced in, among other forms, Hungarian (dialects) *hős*, Mari *kače*, Mansi *kuš/kuč/kuš/huuš* and the Samoyedic languages *kač/koč/keeče*. It seems worth considering if, by its virtue of its phonetic and semantic parallels, if there is some etymological connection. Could this word have influenced the semantic development of Chuvash *xusax* in a manner analogous to how the meanings of the Turkish *gavur* (‘non-Muslim’), from the Persian *gabr* (‘Zoroastrian’), is understood to have been influenced in by a perceived similarity to the Arabic *kāfir* (‘infidel’)? Has a process of mutual semantic influence occurred, for example, in the development of Mari *kače* (‘bridegroom’, ‘boyfriend’, ‘fiancé’, ‘suitor, admirer’, ‘young man, boy, guy’, ‘bachelor, unmarried man’) and its apparent borrowings *ozak* (‘unmarried, single (man)’), *казак* (‘Cossack’), *казак* (‘Kazakh’), and *каське* (‘deserter, runaway’) (Mari-English Dictionary 2022)? Collinder himself claims the origins of the Finno-Ugric root is obscure. Could it itself be a borrowing, perhaps from an East-Iranian word via Oghuric like Keenan’s etymology of *Kazan*, cf. Iron Ossetian *къайаг* (*kayag*) meaning ‘couple; groom; bride’ (‘пара; жених; невеста’) from a root meaning ‘pair’ (Guriev 1993)? Lacking the knowledge to evaluate these questions, I defer to those with greater expertise in Finno-Ugric historical linguistics to either disregard them or consider them further.



(‘harb ve firar etmek’); ‘pull back and wage war’ (‘çekinub savaşmaq’); ‘be saved’ (‘qurtulmaq’); ‘avoid or refuse’ (‘ictināb ve istinkāf etmek’); and ‘be rescued’ (‘hālāş bulmaq’). And he similarly glosses the related Chagatai derived verb ‘qaçıqmaq’ with the meanings ‘flee and wage war’ (‘qaçub savaşmaq’); ‘become scattered and destitute’ (‘parākende ve perişān olmaq’); ‘be pushed off to one side’ (‘bir yana çarpılmaq’); and ‘turn away from’ or ‘break off relations with’ (‘yüz çevirmek’). In modern Kazan Tatar, similarly, the verbal root *qaç-* designates, inter alia, an ‘escape from a place of danger or captivity, ‘hiding’, ‘keeping out of sight’, ‘avoiding contact with others’, ‘being lost or extinct’, and ‘refusal’ or ‘avoidance’ (TTAS: 187–188).³⁹ A comparable set of meanings is expressed in Modern Turkish by the derived form *kaçınmak* (‘avoid’, ‘refuse’, ‘stand aloof from’). In some of the Turkic languages, moreover, the verb *qaç* can also denote the act of a girl or woman who leaves her family to live with the family of her husband. In Altai, for example, *кач* (*kaç*), in addition to meaning 1) ‘run away’ and 2) ‘avoid’, has an additional meaning of ‘get married’ (‘выходить замуж’) (Chumakaev 2018). And in modern Turkish, *kaç* can take on the meaning of ‘elope’ in some contexts (e.g., in the expressions *evden kaçmak* and *kocaya kaçmak*). In all of the above examples, *qaç* conveys the sense of the breaking of social relationships and the making of new ones.

More than simply a ‘fugitive’ or ‘runaway’, therefore, this paper speculates that in 13th and 14th-century Turkic steppe societies, **QAÇAQ** was used as a more generalized term applied to an individual or group who lived outside the normal system of clan kinship and tribal affiliation, whether as outcasts living on the margins of society or as refugees living among a foreign group. It further speculates that this status became more widespread and politically significant in the period of widespread social upheaval that followed the Mongol invasions of the 13th century, during which old tribal and clan structures were destroyed, the political salience of traditional kinship relationships were diminished, and new sociopolitical identities first emerged (Golden 2009: 118). For instance, in his well-known account of his travel to the Golden Horde between 1245–1247, in the immediate aftermath of the Mongol occupation of the Pontic-Caspian steppe, the Franciscan friar and Papal diplomat Giovanni da Pian del Carpine describes the social devastation he witnessed across the steppes of southern Russia, where conquered peoples were subject to forced relocation, conscription into the Mongol army, or deportation to the slave markets of the Crimea. Those living outside of recognized communal structures, including destitute populations, as well as unmarried adult men and women, were especially vulnerable to impressment or enslavement by Mongol officials, as Carpine’s account makes clear:

When we were in Russia, a Saracen was sent to us, and it was said he was from Cuyuc Khan and Bati, and this officer took one boy from whomever had three, as was later told to us; and any man who did not have a wife he led away; and he did the same with women who did not have legitimate husbands. He likewise deported paupers and those who begged for their food. (Hildinger’s translation 1996: 80).⁴⁰

³⁹ Among other meanings the dictionary’s authors list ‘Хәуеф-хәтәрле урыннан, әсирлектән, кешеләрдән һ.б. яшерен рәвештә рөхсәтсез китү, ычкыну, котылу’; ‘Яшеренү’; ‘күренмәскә тырышып, берәр жиргә посу’; ‘Күренмәскә, очрашмаска тырышу; читтәрәк, кырылап йөрү, читләтәп йөрү, аралашмау’; Югалу, бетү, юкка чыгу; Баш тарту; читтә калу, читләшү’.

⁴⁰ The Latin reads: ‘Nam cum essemus in Russia, missus fuit Saracenorum ex parte Cuynthcan vt dicebatur & Bati : & praefectus ille a quolibet homine qui habebat tres pueros vnum accipiebat : & quicumque viri non habebant vxores, illos deducebant, & aciebant demulieribus etiam illud idem que viros legitimos non habebant. Pauperes etiam qui mendicando suum victum quasrebant similiter deportabant’ (Carpine 1903: 66).



Here we encounter the beginning of a plausible explanation for the semantic development of **QAZAQ** as evident in both the different meanings ascribed to it in earlier 14th-century Mamluk attestations ('solitary', 'unmarried') and subsequent 15th-century usage ('runaway'), as well as the corresponding semantic differences between modern Chuvash *xusax* and *kasak*. We cannot suppose that the Mongols, who have been accused of every conceivable injustice save for theological dogmatism, were especially concerned with adjudicating the marriages of their subject peoples on the grounds of religious validity. For people of the 13th-century Pontic-Caspian steppe, therefore, the recognition of one's 'legitimate' marriage required official approval and was likely closely linked to membership in a politically recognized community. Golden (1992: 254), for example, mentions a tradition of taxing marriages among subject peoples in Volga Bulgaria and it is conceivable that the Mongols (together with their local Bulgur officials) continued the practice in some way. In any case, such practices point to the power of rulers in this period to impose the terms of (il)legitimate kinship on their subjects more generally. In the context of the large influx of denationalized, predominantly Kipchak-speaking Tatars to the upper Volga region following the Mongol conquest of the Bulgar kingdom—whether as subjects of the Mongols or fugitives from Mongol power—it becomes conceivable that these Tatar *qaçaqs* were perceived by Mari and Oghuric speakers as 'solitary' or 'unmarried' men, i.e., 'without clan or kin'—probably the best translation of 'mujarrad' encountered in the Arabic gloss of the first Mamluk Kipchak attestation of *qazaq*.⁴¹ Indeed, modern Chuvash *xusax* denotes not only an 'unmarried man' but also a 'widower' or 'widow' (вдовец, вдова), or just a 'man' (парень), suggesting that the meaning extended beyond simply that of a 'bachelor' to include strangers or foreigners living outside the kinship structures established through recognized relations of affinity. This also helps to explain the early use of both **QAČAQ** and **QAZAQ** in some contexts as a synonym for 'slave' since it was exactly those without kith and kin who were considered outside the 'mutuality of being' that defines kinship ties cross-culturally (Sahlins 2013). 'What makes a slave different from a serf, a peon, captive or inmate,' Graeber and Wengrow (2021: 187) point out, 'is their lack of social ties. In legal terms, at least, a slave has no family, no kin, no community; they can make no promises and forge no ongoing connections with other human beings.'

At the same time, it appears that both **QAČAQ** and **QAZAQ** were used not only as labels for people living outside of state-recognized kinship relations but were also adopted by these people themselves as designations for new, alternative forms of kinship and group affiliation in a politically fractious and rapidly shifting social order. In the Caucasus, as noted in section I, *qaçaq* was borrowed from Kipchak speakers into most regional languages as a synonym for *abrek*—itself derived from a Middle Iranian form of Old Persian *aparaq* (cf. New Persian *avare*), a word mean-

⁴¹ Houtsma (1894: 86), clearly influenced by the later development of **QAZAQ**, mistakenly translates the Arabic مجرد (*mujarrad*)—the gloss for Mamluk-Kipchak 'qāzāq'—as 'landstreicher' ('vagabond'), an error pointed out by Pritsak (2006) and others. More recent editions of the Mamluk Kipchak Glossary (Toparlı *et al.* 2000:115/ms.26b6) have translated the term as 'homeless' and 'without shelter' ('evsiz, barksız'). Given both that *qazaq* is preceded in the same glossary by 'ergen' and 'evli' and followed by 'başdaq'—respectively glossed as 'al-ʿazab min ar-rijāl' ('single'), 'al-mutāʾahhil' ('one who takes a wife'), and 'l-ʾārī min al-aḥl wa ǧairihim' ('free of family and the like')—and also that *qazaq* appears in another 14th-century Mamluk Kipchak text in the compound 'qazaq başlı' with the gloss 'āzib' ('single') (Atalay 1945: 189/ms.24b11), it seems certain that *qazaq* denoted some form of kinship status (or lack thereof). In modern Arabic, 'mujarrad' can mean 'bare', 'denuded', 'free' (Wehr 1976: 120). Formerly, it also seems to have been used in the meanings of 'unmarried', 'solitary', or 'alone' as seen by its uses in medieval Persian (Steingass 1992: 1177). My contention is that this points to a more general meaning of 'without kith or kin'. I am grateful for the significant feedback I received on this topic from one of my anonymous reviewers.



ing ‘homeless’ or ‘vagabond’ that passed into the Caucasian languages together with *qaçaq* under the influence of East Iranian and Turkic speaking steppe refugees, who had fled from Mongols into the mountains in the early Middle Ages (Bobrovnikov 2007: 245–246; Grant 2009: 83–84). In the Caucasian highlands, according to Soviet scholarship, the label *abrek* and *qaçaq* originally denoted someone ‘expelled from his family or clan’ (Bobrovnikov 2007: 245–253): The *qaçaq* could be a ‘homeless vagabond’ who survived by theft or banditry, but they could also be refugees from blood feuds or intradynastic political struggles who were granted hospitality as an expatriate among a foreign people. They were, as Grant (2009: 83) describes them, ‘detachable kin’ who existed ‘not only as figures intended for temporary or permanent exile because of blood feuds but as limit-figures whose entry into new communities was often seen as consolidating those who received them.’

A similar set of meanings is observed in the development of *qazaq* in 15th and 16th-century Russian sources, as outlined in section II, where the label was used to describe not only an amorphous class of ‘fugitives’, ‘runaways’, and ‘bandits’ living on the Pontic-Caspian steppe, but also exiled Tatar princes and their retainers who sought refuge with Muscovy in the course of the intradynastic struggles in the Kazan Khanate—a common occurrence resulting from local factional struggles within the Khanate and the shifting system of alliances between Russians, Crimean Tatars, and the disintegrating Golden Horde that defined its politics between the 15th and 16th centuries (Lee 2015; Davies 2016; Pritsak 1967). This system of political *qazaqliq*, as Lee (2015) describes, subsequently spread into Central Asia, where it came to denote precisely those who fled from their communities in an act of political rebellion.

The phenomenon of *qazaqliq* therefore represented an inversion of the traditional system of political kinship in Turkic societies in two respects. Firstly, it shifted emphasis from a system of patrilineal descent that normally structured the politics of these steppe societies to one of matrilineal consanguinity, since in many contexts rebellious princes could expect support from their maternal relatives against their paternal uncles, cousins, or half-brothers in the course of dynastic struggle—a feature of Turkic political life that Ecsedy (1972: 250) argues can be observed as early as the 6th and 7th centuries, where kinship structures responded to wider sociopolitical conditions, cycling between ‘patrilineal (as a rule; in prosperity)–matrilineal (temporarily, in need)–patrilineal (after having been strengthened and advantageous conditions)’ systems of kinship. Secondly, the term described a new form of social affinity defined not through marriage but membership in a *qazaq* warband—an important sociopolitical institution in post-Mongol Central Asia that was defined by free association and loyalty to a military leader over tribal or ethnic identities (Gommans 2017). Such warbands, and the new political and tribal structures they produced, as Golden (2009: 118) describes, ‘stemmed more from the *nökürized* world of the Chinggisid armies and personal forces than from the traditional bonds of ‘kinship’. In contemporary Turkic heroic epics, the personal bond between the chieftain and his *nökürs* was the bedrock of political loyalty.’ This potentially accounts for why *qazaq* was first borrowed into OAT in the meanings of ‘traveling companion’ or ‘friend’ (Kanar 2011: 435). In both examples, the label *qazaq* described people living temporarily outside an established system of clan and tribal hierarchies, or in a political sense, those who were fighting to restructure this system.

In approaching the evidence in this way, this paper argues that the etymological connection between **QAÇAQ** and **QAZAQ** can be placed within a plausible historical narrative that accounts both for the phonetic transformations that produced the change *qazaq* < *qa(t)saq* < *qaçaq* among the emergent, multilingual society of Tatars living in the greater Volga region during the 13th



and 14th centuries, as well as for the different meanings historically ascribed to both **QAÇAQ** and **QAZAQ** over time and in different parts of the Turkic speaking world. And it further suggests that in considering both forms together we can develop a more robust account of their different social valences over time and glean something about the sociolinguistic realities of these societies. In final section below, the paper examines how the reflexes of **QAÇAQ** and **QAZAQ** were similarly ‘enregistered’ in the languages of Caucasus and Central Asia as contrasting forms with different semantic valences. In so doing, it makes an argument for the usefulness of register and enregisterment as analytical frameworks through which to understand larger processes of language change within historical linguistics.

IV: ON THE USEFULNESS OF THE ANALYTICAL CONCEPT OF REGISTER TO HISTORICAL LINGUISTICS

This paper has argued for the common etymology of modern reflexes of CT **QAÇAQ** and **QAZAQ**, arguing that both derived from a common source: a CT derived nominal form *qaç(g)aq*. The steps outlined for this evolution were as follows: 1) the loss of the *g* within the western branches of CT (*qaçaq* < *qaçgaq*), a shift already documented by the 13th century; 2) a shift in the affricate in the NW Turkic (*qa(t)saq* < *qaçaq*) and its borrowing into the dialects of Middle Oghuric in the Volga-Kama region during the 13th and 14th centuries as *qasaq*; and 3) the voicing of intervocalic consonants in Middle Oghuric (*qazaq* < *qasaq*) and its borrowing in this form back into Western CT Turkic (as evidenced in Mamluk Kipchak) and into other regional languages like East Slavic, Mari, and Udmurt. Doubtless, however, this is a very oversimplified picture of what was certainly a more complicated and uneven historical-sociolinguistic process, in which not only language internal sound changes and interlanguage borrowing, but also a complex world of multilingualism and linguistic variation—shaped by conscious and unconscious forms of linguistic differentiation—likely also played a meaningful role. The concepts of register and enregisterment are therefore useful in grounding this analysis, allowing us to speculate on how the social roles and positions of different speakers might have shaped the adoption of **QAÇAQ** and **QAZAQ** across different historical-sociolinguistic contexts.

The most enduring analytical debates in comparative historical linguistics over the past century and a half have largely centered around the problem of linguistic variation. In part these responded to 19th-century controversies around the relationship between a language’s phonetic regularity and its level of evolutionary development: in Germany, Steinthal developed an influential model of language evolution wherein the movement away from variability toward fixed representations within an ‘inner linguistic form’ served as a primary metric of a language’s level of evolutionary development; and followers of Steinthal in the United States like Powell and Brinton, similarly pointed to the presence of ‘alternating sounds’ and ‘synthetic sounds’ in the indigenous languages of the Americas as markers of their linguistic ‘primitivism’. As McElvenny (2019) describes, two important albeit divergent critiques of this evolutionary framework emerged at the end of the 19th century: the first, developed by Boas and his followers in the United States, stressed that regular phonological rules govern all speech communities and that the perception of ‘variation’ was largely an error on the part of European-American linguistics, whose own phonological consciousness shaped their (mistaken) perception of the sound rules governing indigenous languages.



In Germany, a second major critique of the connection between phonetic variation and linguistic evolution developed targeting the Neogrammarians and their insistence on the absolute regularity of sound changes and their autonomy from other aspects of language change.⁴² Gabelentz attacked the Neogrammarian comparative method⁴³ as overly rigid and unable to recognize the ‘radical mutability of linguistic form’ which, not unlike Steinthal, Gabelentz saw as being more pronounced in languages at a ‘lower level of culture’ (for Gabelentz those language communities without high level of formal education in literary standards) (McElvenny 2019: 49–52). In a more systematic and substantiated way, Schmidt, Schuchardt, and other early proponents of the competing *Wellentheorie* (‘wave model’), argued that the Neogrammarians adopted an overly rigid distinction between language-internal and autonomous developments in phonetic and phonology (i.e., sound laws) and morphosyntax (i.e., change by analogy, such as the substitution of one grammatical paradigm for another) on the one hand, and language-external borrowing (i.e., the lexicon) on the other. And they pointed out that this Neogrammarian distinction failed to account for other forms of linguistic kinship beyond the narrow parameters of direct descent, inasmuch that the Neogrammarians claimed that there exists a clear demarcation between language-internal innovations (the basis for genetic relationships) and the external influences of genetically unrelated languages (François 2015).⁴⁴

The Neogrammarian model has remained influential within historical linguistics to this day. And not without reason, because this model has proven powerfully successful, allowing for the production of generalized guidelines for predicting sound change, thereby providing a theoretical basis on which the historical reconstruction of non-attested linguistic forms can be plausibly generated. It is on the basis of these premises, for instance, that this paper builds major pillars of its own argument (e.g., its analysis of the timing *qaçaq* < *qaçgaq* in section I or its analysis of the shift in the articulation of CT affricate and intervocalic voicing in Middle Chuvash in section III). More recent scholarship in mainstream historical linguistics, therefore, aims to combine

⁴² The Neogrammarian concept of the ‘sound law’ (*Lautgesetz*) states that phonology is largely autonomous from either syntax or semantics and thus has no influence on a word’s denotational content. Consequently, a phonetic shift away from the norm of the speech community among a speaker or subset of speakers does not change the ‘meaning’ of the affected words. Over time, moreover, any sound shift is predicted to regularly occur in all analogous phonetic environments (i.e., it becomes universal across a given language). Any apparent exception to a ‘sound law’ (*Lautgesetz*) must therefore be explained with reference to a still further sound law through the which the exception itself is universalized across the language or dialect in question (e.g., Verner’s law as a universal account for the apparent exceptions in Grimm’s law). Linguistic variation, in this model, is therefore understood as a process of phonetic divergence within speakers of a common language that remains entirely disarticulated from traditional philological concerns around meaning (wherein ‘semantic drift’ is conceived as an independent sociolinguistic process).

⁴³ In the Neogrammarian model of historical descent and diachronic sound change, the relationship of one language to another is evaluated through descent from a common ancestor—a genetic relationship popularly visualized as a ‘family tree’ (*Stammbaum*) wherein languages gradually diverge (or ‘split’) over time. The Neogrammarian model thus puts great emphasis on the regularity of sound changes and their autonomy from other aspects of language change.

⁴⁴ Proponents of the diffusionist model of language change, in contrast, have argued that genetic relations of descent are complicated and cut through by influences from neighboring languages. And they point to regional affinities between otherwise ‘unrelated’ languages (as captured in terms such as the ‘Sprachbund’, ‘contact languages’, and ‘creoles’) that go beyond lexical borrowings to encompass both phonetic and morphosyntactic influences—such as, for example, work on the interaction and mutual influence between the phonological systems of Common Turkic, Oghuric and Finno-Ugric languages in the Volga-Kama region cited above (see Baitchura 1975; Róna-Tas 1982b; Johanson 2000; Agyagási 2019).



insights from both approaches, as exemplified in the ‘network analysis’ model which approaches the problem of genetic affinity between languages as a function of the statistically quantifiable percentage of features shared between them (Heggarty *et al.* 2010). Regardless of its precise views on the nature of linguistic kinship, however, much of mainstream historical linguistics—both proponents and critics of the Neogrammarian framework—continues to approach language as a bounded and regular code for which a stable set of rules can be enumerated and in which phonological variation is relegated to a problem of temporal or spatial displacement that remains disconnected from synchronic social and cultural forms. In this model, phonological variation within a language is relegated to a problem of ‘dialect’, and variation within a dialect is relegated to a problem of ‘idiolect’ (i.e., the language of an individual speaker)—an analytical approach that, not unlike that of the Neogrammarians, continually seeks to resolve exceptions to the autonomous laws governing phonological change by shifting to a second-order rule and/or a smaller scale of analysis.

Scholars working in the fields of sociolinguistics and linguistic anthropology, in contrast, have argued that variation is encountered at all scales of sociolinguistic life and that it constitutes a meaningful and integrated piece of a language’s structure. Even at level of the individual speaker, people can and do speak in and through different social ‘voices’ (Bakhtin 1982), and therefore, phonemic variation, rather than disconnected from morphosyntax and semantics, is itself often conscious and socially meaningful. In his discussion of the inherent tension between the ‘push’ of sound laws and the ‘pull’ of enregisterment, Silverstein (2016: 54–55) makes exactly this point, taking issue with the Neogrammarian model of diachronic change precisely for how it effaces the sociolinguistic complexities of actual language use for a simplified model of mechanical change, as well as for how it relies on a reductive model of ‘standard language’ that denies the omnipresent fact of variation in favor of an ideology of a homogenous national Volk, or what he terms ‘the normative order of *langue* at the heart of earlier historical linguistics.’ He does not deny the predictive power of Neogrammarian-inspired sound laws in accounting for aspects of diachronic change. Rather, Silverstein calls into question the Neogrammarian division, largely preserved in modern historical linguistics, between phonological change as a mechanical process that is unconscious and best understood through the terms of natural science, and the study of morphosyntax and lexicon as sites of meaning-making, i.e., as conscious mental process coding affective and cognitive values. In opposition to this model, Silverstein explains how the concept of register redirects the study of language change from individual phonemic segments (sounds) to phonological structures (signs and their meanings) as they become indexically linked to the identities or affective dispositions of speakers within a dynamic sociohistorical context through a process he and others have termed ‘enregisterment’.

This analytical innovation differentiates this approach from earlier approaches to variation within sociolinguistics, as Eckert (2012) outlines: earlier approaches within sociolinguistics identified important correlations between linguistic variables and macro-sociological categories such as ethnicity and class (first wave of variation studies developed by Labov *et al.*) and subsequently sought to better understand how this variation was locally perceived and understood by speakers in everyday interactions (second wave of variation studies developed by Trudgill, Cheshire, *et al.*). However, these approaches tended to view ‘vernacular’ language as reflecting or marking sociological identities. In contrast, a ‘third wave’ of recent scholarship on linguistic variation in both sociolinguistics and linguistic anthropology—as exemplified by Silverstein’s approach—under-



stands registers as tools of meaning-making through which language users agentively construct their own and others' identities in the process of interaction.

Before considering how the concept of register is useful in an analysis of the etymological connection between **QAČAQ** and **QAZAQ**, let us consider a few historical examples of enregisterment from within Turkic linguistics to better illustrate the problem. Consider the etymology of the ethnonym 'Chuvash'. According to Agyagási (2019: 4–10), 'Chuvash' is likely derived from the Old Turkic⁴⁵ *yuyaç* meaning 'from the opposite bank [of the river]'. In the Middle Oghuric dialects, according to Agyagási, *yuyaç* underwent the change *yuyaç* > *çuvaş* among the speakers of the Ĵ-dialect (the variant spoken along the upper Volga and by most of the Bulgar elites) and *yuyaç* > *śuvaś* in the Ś-dialect (spoken around the Kama and a predecessor of modern Chuvash). Among speakers of the former, *çuvaş* became a new Bulgar-Tatar designation for those Ś-dialect speakers living along the Kama and the left bank of the Volga. Whereas for speakers of the Ś-dialect, *śuvaś* came to refer, from the converse perspective, to those communities living along the upper Volga and the river's right bank—i.e., speakers of both the Ĵ-dialect and Kipchak-Turkic speakers—thus eventually passing into Mari as *suas* in the meaning of 'Tatar'. Only later, following the migrations of Ś-dialect speakers to the right bank of the Volga around the turn of the 16th century did the ethnonym 'Chuvash' become fixed in its modern meaning. Agyagási analysis is compelling, but even more can be drawn from it using the concept of register, because this is likely not simply a story of sound change and borrowing: it appears just as likely that a variation in world initial *y* (as *ç* or *ś*) in the Middle Oghuric was itself a 'register shibboleth' that allowed speakers of both dialects to distinguish between members of other groups, and that overtime became productive of a shift in these ethnic designation's 'denotational' meanings in addition to their 'indexical' ones. 'Registers', as Silverstein (2016: 59–60) notes:

...have all the properties of languages as structures immanent in denotational discourse; since registers are, however, indexically particular to context, whether in a positive or negative stipulation, the set-theoretic union of the elements of registers in a community, socio-linguistically viewed, thus constitutes the inclusive envelope of the community's 'language' and its 'oscillations'.

In other words, peoples' language is not only meaningful because words have referents (i.e., denotational meaning) but because people's language—their accent, choice of words, style of address, manner of speaking—also reveals meaningful information about a speaker, such as their social background, their attitudes and stance, *et al.* (i.e., indexical meaning). Language variation, in this perspective, is inseparable from the larger question of language as structures of meaning. Among Middle Oghuric speakers living in the Volga-Kama region between 13–16th centuries, any interpretation of the meaning of an utterance like 'from the opposite bank' likewise responded to differences in enregistered variants of this Oghuric word (i.e., different dialectal pronunciations) that were in turn linked to contrasting speaker types (Ĵ-dialect speakers living on the Volga's right bank or Ś-dialect speakers on the river's left bank)—an indexical association through which the

⁴⁵ Here I am using the system of linguistic classification employed by Agyagási (2019: 4), although she labels the term both 'Old' and Proto-Turkic. There are competing accounts of the relationship between Oghuric and Common Turkic, as well the relationship of both to 'Old Turkic' (see Tekin 1991). However it is not my objective to adjudicate those here.



precise denotational content of this utterance was in fact apprehended. It was only following the gradual disappearance of the Ĵ-dialect—and consequently the loss of this enregistered contrast—coupled with the migration of Š-dialect speakers to the right bank of the Volga around the turn of the 16th century that allowed for ‘Chuvash’ to emerge as a stable designation for a particular ethnolinguistic group.

The problem, from the perspective of historical linguistics, is both that evidence for enregistered differences are not always captured in the written record as such, and that it is very difficult to distinguish, both analytically and conceptually, between variation as the product of processes of enregisterment on the one hand, and the mutual exchange of contrasting cognate forms between distinct language communities on the other, especially when there is a subsequent divergence in semantic valence between related forms over time. This is not surprising from the ‘third wave’ perspective on language variation, inasmuch that they critique the conceptual division of language into discrete, regularized codes and emphasize how real-world linguistic interaction always incorporates and reflects variation. This phenomenon is already familiar to many Turcologists. Consider the problem with categorizing linguistic variation as highlighted by Aqtay and Jankowski (2015: 8–9) in the introduction to their Karaim Turkish-English dictionary:

It is not an easy question to answer which sources should be included in a Crimean Karaim dictionary. Many old forms are similar to Western Karaim, while new ones are Turkish, more exactly Crimean Turkish. These Turkish words encountered in the texts and dictionaries are not always spelled in standardised forms... They all are known not only from Crimean Turkish and Crimean Tatar, but also from Anatolian Turkish dialects. A specially difficult question is how to spell Arabo-Persian loanwords, since unpointed Hebrew script is not a good system to reflect vowels. Moreover, it is certain that these loanwords were not pronounced in a uniform way, since they were style and register dependent... The primary reason of phonetic alternation in pronunciation was a tendency of adapting loanwords to the phonological rules of Karaim and Turkish. As a result, most loanwords existed in at least two forms, one formal and non-adapted, another informal and adapted. Since in some cases adaptation was gradual, some words were pronounced in more than two forms. This tendency is well-known in many Turkic languages including Turkish...

The dilemma described by Aqtay and Jankowski, as they themselves acknowledge, is not unique to Karaim, but reflects a wider phenomenon that was likely especially pronounced among the multilingual societies of the Volga-Kama region and the wider Turkic world. In the introduction to his edited facsimile of Mirza Mehdi Khan’s *Sanglax* (1759), similarly, Clauson (1960: 4–5, 14–15) pushes back against mistaken impressions the *Sanglax* was primarily a Chagatai dictionary, noting that the original author had claimed it was a guide to the Turki Language that included entries and grammatical paradigms from both its two major literary standards (Chagatai and Rumī), along with entries from a number of other dialects and colloquial variants, including some entries which Clauson suspects are ‘imaginary words with imaginary meanings’ and which he glosses as ‘dubious’. Like in the case of **QAZAQ**, however, occasionally new words appear whose origins are unexplained by the expectations of sound laws but whose reality as a linguistic fact cannot be denied, even by prickliest of philologists. In these cases, especially, register provides a useful analytical tool through which to reconsider such innovations.



Let us return to the question of the relationship between **QAČAQ** and **QAZAQ** and the processes through which they became enregistered as contrasting forms. As argued in the previous section, **QAZAQ** first emerged as a contrasting, Oghuric-form of **QAČAQ** during a period of widespread upheaval that took on a more generalized meaning of a person without recognized social ties (i.e., ‘unmarried’ or ‘unaffiliated’)—a meaning captured in the Arabic glosses of Mamluk Kipchak *qazaq* as well as Chuvash *xusax*, one of two contrasting Chuvash variants of ‘Cossack’ (the other being *kasak*). Subsequently, **QAZAQ** entered into seemingly every language spoken on the Pontic-Caspian steppe, but the valences in which it was deployed seems to have depended on the sociolinguistic and political contexts in which it used. Among modern Bashkir and Chuvash speakers, as argued above, a contrast between the *kazak* and *kasak*—the reflex of **QAČAQ** in both languages (a borrowing in the case of Chuvash)—likely barely registers, except perhaps in the written language as an artifice of Russian literary standards. Among speakers of Kazan Tatar, moreover, the differences in the semantic valences of *qačaq* and *qazaq* are seemingly highly contextual—i.e., *qazaq* refers to, inter alia, a specific historically or socially located subset of *qačaq*. When one considers the Russian and Polish-Lithuanian uses of the word, on the other hand, where no contrastive form of **QAČAQ** exists, the semantic values in which *qazaq* was used (‘runaway’, ‘political refugee’, ‘rebel’, ‘bandit’) seems to even more closely mirror Turkic meanings of **QAČAQ** (see Lee 2015: 43–45). This is unsurprising, since Russian *qazaq* is in effect an Eastern Slavic borrowing of Turkic **QAČAQ**, and unlike among Turkic speakers, has no contrasting variant through which to focus its range of meanings.

As one moves farther from the vicinity of the Volga-Kama region, the semantic differences between the two reflexes become greater, so that their etymological relationship is increasingly obscured and they gradually transform from contrasting enregistered forms to unrecognized cognates with distinct denotational meanings. Thus among the Kipchak speakers of the Crimea, as discussed above, a *qačaq* continued to designate a ‘runaway’ or ‘refugee’, while a *qazaq* designated either ethnic ‘Russians’/‘Ukrainians’ or ‘slaves’ (Aqtay and Jankowski 2015; Królikowska-Jedlińska 2019). And among Kipchak speakers of the Caucasus, similarly, the *qačaq* or *abrek* referred to a distinct class of ‘foreigners’ and ‘political refugees’ that ranked above the *qazaq*, *lag*, *lay*, or *kuli*—labels applied to the lowest class of serfs or slaves in Caucasian society (Bobrovnikov 2007: 246; Takvul 2000). In both cases, a hierarchical pattern of enregisterment is observed wherein the ‘native’ Kipchak variant *qačaq* serves as a more neutral designation for sociopolitical status whereas *qazaq* designates low-status social outsiders. This is not to suggest, importantly, that all Turkic speakers understood the contrasts in these terms. Since any linguistic register is always ideological and therefore perspectival (Gal and Irvine 2019), it is very likely that the semantic value hierarchies that shaped the use of this term were enregistered differently across social space and were subject to contestation and renegotiation in practice. Unsurprisingly, we encounter a *different but related set of contrasts* in early modern Chagatai. And among the Cossacks of Ukraine, the label *kozak* did not designate a slave at all, but an ethno-political identity encompassing both Turkic and East Slavic-speakers that was organized around the values of communal autonomy and individual freedom.

Let us conclude our discussion by looking at how **QAČAQ** and **QAZAQ** were enregistered in early modern Chagatai, while further outlining how an analogous process of enregisterment produced a third contrasting Northern Kipchak variant of **QAČAQ**: *qaş(ş)aq*. As discussed above, Buḥārī (1892: 214) ascribes *qačaq* and other nominalized forms of the root *qač* with the meaning



of ‘fugitive’ (‘firārī’) or ‘wanderer’ (‘girīzende’).⁴⁶ For *qazaq*, on the other hand, Buḥārī (1892: 219) gives the primary definition of ‘without home or family’ (‘bī-ḥānumān’)—a meaning also encountered in the Kipchak languages (see Toparlı *et al.* 2003). It is only as a secondary meaning that he glosses *qazaq* as an ethnonym. At the same time, *qazaq* is listed with numerous derived forms whose glosses suggest a set of related, albeit more explicitly negative connotations, in comparison to Buḥārī’s glosses of *qaçaq*.⁴⁷ The glosses for *qaçaq* and *qazaq* show a set of related meanings in Buḥārī’s dictionary—recall Atlasi’s citation of them in his own argument for these words’ common etymology discussed above. But they are not synonymous: whereas the glosses that he ascribes to *qaçaq* (the ‘native’ Chagatai form) tend to be more neutral, those ascribed to *qazaq* appear to convey, at least in its derivational forms and from the sociolinguistic viewpoint of Buḥārī’s dictionary, a more negative set of meanings (not unlike in the Kipchak Turkic languages of the Crimea and Caucasus discussed above).

However, a further contrast with *qaçāq* is seen in Buḥārī’s gloss of the word *qaş(ş)aq*. Like *qazaq*, *qaş(ş)aq* is ascribed the meaning of ‘without home or family’, as well as a range of meanings from ‘powerless’ (‘bī-bīza’at’ (sic)) to ‘destitute’ (‘ser-ü-pā-birehne’).⁴⁸ Like *qaçaq*, moreover, *qaş(ş)aq* can be confidently connected to a form like modern *қазақ* (*qaşaq*), a Kazakh reflex of **QAÇAQ** meaning ‘fugitive’, as the word *qaş(ş)aq* is very likely derived from the root *қау-* (*qaş-*): the expected variant of *qaç-* in the Northern Kipchak languages, sometimes represented in official orthography (e.g., modern literary Kazakh) and sometimes only present in vernacular pronunciation (e.g., some dialects of Kazan Tatar).⁴⁹ How should we understand this? The most plausible explanation, I argue, is that in the centuries-old Chagatai literary tradition as compiled and described in Buḥārī’s dictionary there appears a hierarchy of cognate derivational variants, both ‘native’ Chagatai forms, and ‘borrowings’ from neighboring Tatar and Kipchak dialects that developed from a common etymon, CT *qaç(g)aq*, structured by at least a three-way enregistered contrast between 1) *qaçaq* (and various other forms likely consciously associated by Chagatai writers with the root *qaç-*); 2) *qazaq* (and its variants); and 3) *qaş(ş)aq*.

⁴⁶ The full entries are: Qāçār (قاچار): serī’l-ḥareket, qaçan ve firar eden, qaçāq; Qaçmāşlıq (قاچماشلىق): qaçmaqliq, firar, giriz, ‘azimet; Qāçāğ (قاچاغ): firari; qaçqin, girizende, ğulām, bende, firari, abd girizpa (sic?); and Qaçğaq (قاچغاق): qaçiq, hezimet, bozğun ve firari.

⁴⁷ For example, Buḥārī glosses ‘qāzāqci’ with various terms meaning ‘thief’ or ‘bandit’ (‘ḥırsis (sic); ḥarami; quttā’-i tarik; qaltabān; oğru; rāhzan’) and he glosses a related verbal form, ‘qāzāqmaq’, with the meanings ‘to engage in highway robbery and banditry’, ‘to wander’, or ‘to counterfeit’ (‘ḥırsızlıq, ḥaramılıq, quttā’-i tarikliq etmek; gezgincilige çıqmaq; pul kesmek’). Finally, Buḥārī gives an abstract nominalized form, ‘qāzāqlıq’, in the meaning of ‘the state of soldiers during travel in times of conflict’ (‘ceng eyyāmi seferde asker ve leşkerin ḥālī’) (Buḥārī 1892: 219).

⁴⁸ The complete gloss is ‘poor, beggar, without home or family, without wealth, without money, destitute (*lit.* with naked head and feet)’ (‘faqir; müflis; gedā; bī-ḥānumān; bī-bīza’at (sic); qaşāq, pārasız, ser-ü-pā-birehne’) (Buḥārī 1892: 220).

⁴⁹ Astute readers might ask about the apparent gemination of fricative *ş*—a feature that is preserved in the modern Uzbek word as well (see below). However, Buḥārī also lists the word without gemination in the same entry (e.g., as ‘qaşāq’), suggesting the two forms are related and that their pronunciation likely varied between speakers. That said, we can speculate that presence of gemination is potentially explained by an emphasis on the fricative in the spoken language—a phenomenon that makes sense if we consider the word derives its meaning from a contrast with the affricate in the standard literary form of the word (i.e., a contrast between *ş* and *ç*).



Table 1: Buḥārī’s Hierarchy of Enregistered Forms

Proposed Variants of QAČAQ	Linguistic Classification	Semantic Values (according to Buḥārī)
<i>qačaq</i> : ‘qāčāğ’, ‘qačğaq’, ‘qačāq’	‘native’ Chagatai forms	i. ‘runaway’, ‘fugitive’
<i>qazaq</i> : ‘qazāq’, ‘qāzāq’, ‘qazāq’; (derived) ‘qāzāqci’	Oghuric-influenced Tatar forms	i. ‘without home or family’; ii. ‘a great people living on the Great Desht-i Qipchaq north of Bukhara and Khwarazm’; iii: (derived) ‘thief’, ‘bandit’
<i>qaş(ş)aq</i> : ‘qāşşāq’, ‘qaşāq’	Northern Kipchak forms	i. ‘without home or family’; ii. ‘poor’, ‘destitute’,

In Buḥārī’s glosses, the ‘native’ Chagatai form, *qačaq*, retains a set of more or less neutral meanings (‘fugitive’, ‘flee’, ‘break off relations’, ‘fight against’, *et al.*). In contrast, *qazaq* shows greater diversity in meanings: in its simplest form, it is used both in the sense of ‘without home or family’ and in a second sense akin to its modern ethnonymic meaning. In its derivational forms, *qazaq* also becomes used to refer to various forms of banditry, or in a different context, to the state of soldiers on a military campaign—no doubt a class of persons greatly feared and mistrusted in broader society. Finally, the third form, *qaş(ş)aq*, is similarly given the sense ‘without home or family’ but also the further sense of ‘poor’ and ‘destitute’, as if to refer to people or peoples who, once reduced to a state of permanent exile or homelessness, no longer posed the same threat as those persons or communities described by either of the other two forms. In socio-semantic terms, *qaş(ş)aq* is the lowest of the three forms, at least within the contrasting hierarchy of meanings offered by Buḥārī. Buḥārī’s glosses point to an ideological correspondence between social categories and sociolinguistic forms that positions—in terms of the contrasting socio-semantic senses in which these forms are used—the ‘native’ or ‘standard’ Chagatai *qačaq* above the other two forms, and they are therefore potential evidence for a ‘language ideology’ (Irvine *et al.* 2009) that privileges certain varieties of speech as more or less valuable and prestigious. Nor is the word a unique product of Buḥārī’s imagination: it continues to be attested in dictionaries of modern Uzbek in the words ‘qashshoq’ (‘poor’, ‘beggardly’) and ‘qashshoqlik’ (‘poverty’) (Butayev and Irisqulov 2008: 708).

CONCLUSION

This paper has outlined an argument for a common etymology of QAČAQ and QAZAQ and advanced a related argument for the usefulness of register as an analytical category in historical linguistics. While no one piece of evidence presented in making this argument can be considered conclusive by itself, I argue that the totality of existing evidence positions this proposal as the most plausible etymological explanation available. For the moment complete certainty remains impossible given both the large gaps in our knowledge and the possibility that new evidence can still emerge.



In any case, this article has outlined a case for the analytical usefulness of register in examining diachronic language changes whose contribution to historical linguistics, I hope, can extend beyond the confines of its etymological argument. It is likewise hoped that this article can contribute to further comparative sociological and anthropological discussions of the **QAÇAQ** and **QAZAQ** across time and in different societies, as well as of their relationship to analogous sociopolitical phenomena as they have appeared in different historical periods.⁵⁰ While there is a growing body of impressive work of the historical development of *qazaqlıq* across Eurasia (see Lee 2015), there is still relatively little literature on *qaçaqlıq* as a wider sociopolitical phenomenon in the Turkic world and its uptake as a category among other language communities.⁵¹ Going forward, such comparative work might serve to focus an alternative analytical lens on how people have understood and remade communal identities in periods of political upheaval and rapid social change.

ACKNOWLEDGEMENT

I am very grateful to the article's two anonymous reviewers for their careful reading and insightful suggestions, as well as to Helga Anetshofer, whose early feedback was invaluable in bringing the article into its final form. I also owe a deep debt of gratitude to the late Carol Rosen, who first introduced me to the study of historical linguistics many years ago.

REFERENCES

Primary Sources

- CARPINE, Giovanni da Pian 1903. *Ystoria Mongalorum*. In: Willem van Ruysbroeck (ed.) *The texts and versions of John de Plano Carpini and William de Rubruquis, as printed for the first time by Hakluyt in 1598, together with some shorter pieces*. No. 13: London: Hakluyt Society.
- ÇELEBİ, Evliyâ 1996. *Seyahatnâmesi*, vol. 1–10. Eds. Robert DANKOFF *et al.* Istanbul: YKY.

⁵⁰ As encountered, for example, in the figure of the *gazi* in medieval Islamic steppe societies (Gommans 2017), or in the figure of the *levend* in 17th and 18th-century Ottoman society—a connection already made by contemporaneous Ottoman officials (Veinstein 1999).

⁵¹ This kind of comparative approach might contribute, for instance, in efforts to connect the medieval and early modern phenomenon of political *qazaqlıq* documented across the Eurasian steppe to more recent forms of anti-state rebellion, such as the Albanian Kachak Movement that challenged the Kingdoms of Serbia and then Yugoslavia in the first decades of the 20th century, or as those exemplified by the political resistance of the *abrek/qaçaq* figures who fought Russian occupation in the 19th and 20th-century Caucasus. It might likewise allow us to draw parallels between the social status of the medieval *qazaq* and the status of refugees, undocumented migrants, or stateless persons living in contemporary Turkic states under the status of a *kaçak/qaçaq*; or to consider the relationship between older forms of *qazaqlıq* as an institution with overlapping political and socioeconomic factors and, for instance, the modern institution of 'kaçakçılık' ('smuggling') as a mode of economic survival and political resistance in Turkey and the wider post-Ottoman world.



Dictionaries

- АХМЕРОВ, К. З. *et al.* [Ахмеров К.З. и др.] 1958. *Башкирско-русский словарь*. Москва: Государственное издательство иностранных и национальных словарей.
- AQTAŲ, Gulayhan and Henryk JANKOWSKI 2015. *A Crimean Karaim-English Dictionary*. Poznań: Katedra Studiów Azjatyckich UAM.
- БАММАТОВ, З. З. [Бамматов З. З.] (ed.) 1960. *Русско-кумыкский словарь*. *Русча-къумукъча сёзлюк*. Москва: Государственное издательство иностранных и национальных словарей.
- BEZDOSIKOVA, L. M., N. K. ZABOEVA, R. I. KOSNYREVA [Безносикова, Л. М., Н. К. Забоева, and Р. И. Коснырева. 2003. *Русско-коми словарь*. Сыктывкар: Коми книжное издательство.
- БЛАГОВА, Г. Ф. [Благова, Г.Ф.] *et al.* 1997. *Этимологический словарь тюркских языков. Общетюркские и межтюркские лексические основы на буквы “К”, “Қ”*. Москва: Российская Академия Наук.
- BUDAGOV, L. Z. [Будагов Л. З.] 1869. *Сравнительный словарь турецко-татарских наречий*. Москва: Рипол Классик.
- BUĞARİ, Şeyh Süleyman 1892. *Lügat-i Çağatayi Ve Türki-yi Osmani* [A Dictionary of Chagatai and Ottoman Turkish]. İstanbul: Mihran Matbaası.
- ЧУМАКАЕВ, А. Е. *et al.* [Чумакаев, А. Э. и др.] (ed.) 2018. *Алтайско-русский словарь*. Горно-Алтайск.
- COLLINDER, Björn 1955. *Finno-Ugric Vocabulary. An etymological dictionary of the Uralic languages*. Stockholm: Almqvist & Wiskell.
- Derleme Sözlüğü* [Dictionary of Turkish Dialects] Vol. 8. 1993. Ankara: Türk Dil Kurumu. [1st ed. in 1975]
- ЕГОРОВ, В. Г. [Егоров, В. Г.] 1964. *Этимологический словарь чувашского языка*. Чебоксары: Чувашское книжное издательство.
- FAİK, Ömer. 1896. *Deutsch-türkisches Wörterbuch*. Druckerei: Ossmanieh.
- ФЕДОТОВ, М. Р. [Федотов, М.Р.] 1996. *Этимологического словаря чувашского языка*. (В 2-х тт.). Чебоксары: Чувашский государственный институт гуманитарных наук.
- GURIEV, T. A. [Гуриев, Т. А.] 1993. *Осетинско-русский словарь*. Владикавказ: Изд-во Северо-Осетинского института гуманитарных исследований.
- HINDOĞLU, Artin 1838. *Hazine-i lugat-ı ou dictionnaire abrégé turc-français*. Beck.
- HOUTSMA, Martijn Theodoor (ed.) 1894. *Ein türkisch-arabisches Glossar*. Leiden: E.J. Brill.
- KALIEV, F. *et al.* 2005. *Қазақ тілінің аймақтық сөздігі*. Алматы: Арыс баспасы.
- KANAR, Mehmet 2011. *Eski anadolu Türkçesi sözlüğü* [Dictionary of Old Anatolian Turkish]. İstanbul: Say Yayınları.
- KELEKIAN, Diran 1911. *Kamus-ı Fransevi* [French Dictionary]. İstanbul: Mihran.
- ХАМЗАЕВ, М. Я. [Хамзаев, М.Я.] 1962. *Словарь туркменского языка*. Ашхабад: Изд-тво АН ТССР.
- Mari-English Dictionary*. Mari Web Project: Department of Finno-Ugric Studies at the University of Vienna. (Available online at: <https://mari-language.univie.ac.at/dict.php>, last access: 9 July 2022)
- MENINSKI, Franciscus 1680. *Thesaurus linguarum orientalium, Turcicae, Arabicae, Persicae*. Viennae Austriae.
- MOLINO, Giovanni 1641. *Dittionario della lingua italiana, turchesca*. Rome.
- NIŞANYAN, Sevan *Nisanyan Sözlük—çağdaş Türkçenin etimolojisi* [Nisanyan Dictionary—Modern Turkish Etymology]. (Available online at: www.nisanyansozluk.com, last access: 3 July 2022).
- PAVET DE COURTEILLE, P. M. 1870. *Dictionnaire Turk-Oriental*. Paris.
- ПЫРЕРКА, А. Р. and N. M. TERESHENKO [Пырерка, А. П. и Н. М. Терещенко] 1948. *Русско-ненецкий словарь: около 15 000 слов*. Гос. изд-во иностранных и национальных словарей.



- RADLOV, Vasily 1893. *Versuch eines Wörterbuches der Türk-Dialecte*. Vol. 1. St. Petersburg: Commissionnaires de l'Académie impériale des sciences.
- REDHOUSE, J. W 1856. *Redhouse's Turkish Dictionary: In Two Parts, English and Turkish, and Turkish and English*. London: Barnard Quaritch.
- RUSTEMOV, O. D. [Рустемов, Олег Диливерович] 2020. *Язык кадиаскерских книг Крымского ханства и ханских ярлыков: особенности становления крымского городского наддиалектного койне XVII-XVIII вв.* Казань.
- Sami, Şemseddin 1899. *Kamus-ı Türki* [Turkish Dictionary]. Istanbul.
- SHAW, Robert 1878. *A Sketch of the Turki Language: As Spoken in Eastern Turkistan*. Calcutta: Baptist Mission Press.
- SKVORTSOV, M. I. [Скворцов, М. И.] 1985. *Чувашско-русский словарь*. Москва: Русский язык.
- STEINGASS, Francis Joseph 1992. *A Comprehensive Persian-English Dictionary*. [Revised, Enlarged, and Entirely Reconstructed] New Delhi: Asian Educational Services.
- Tarama Sözlüğü* [Historical Dictionary of the Turkish Language] 1974. Ankara: Türk Dil Kurumu.
- TIETZE, Andreas 2016. *Tarihi ve Etimolojik Türkiye Türkçesi Lugati* [Historical and Etymological Dictionary of the Turkish of Turkey], vol. IV. Ankara: TÜBA TBA.
- TOPARLI, Recep, Hanifi VURAL and Recep KARAATLI 2003. *Kıpçak Türkçesi Sözlüğü* [Dictionary of Kipchak Turkic]. Vol. 835. Türk Dil Kurumu.
- [TTAS] *Tатар теленең аңлатмалы сүзлеге* (Vol III K) 2017. Казан: тәһси.
- ÜNLÜ, Suat 2014. *Çağatay Türkçesi Sözlüğü* [Dictionary of Chagatai Turkic]. Istanbul: Eğitim Yayınevi.
- Vefik, Ahmed [Paşa] 1889. *Lehce-i Osmânî* [The Ottoman Language]. Istanbul: Mahmud Bey Matbaası.
- WEHR, Hans 1976. *A Dictionary of Modern Written Arabic*. [3rd ed. Ed. J. Milton COWAN] Wiesbaden: Otto Harrassowitz Verlag.

Secondary Sources

- AGYAGÁSI, Klára 2019. *Chuvash Historical Phonetics. An areal linguistic study. With an Appendix on the Role of Proto-Mari in the History of Chuvash Vocalism*. [Turcologica 117.] Wiesbaden: Harrassowitz Verlag.
- AGYAGÁSI, Klára 2002. 'Some Middle Bulgarian loan words in the Volga Kipchak languages.' *Acta Orientalia Hung.* 55/1: 25–28.
- ATALAY, Besim (ed.) 1945. *Ettuhfet-üz-zekiyye fil-lûgat-it-Türkiyye* [The Gift of Understanding in the Turkish Language]. Istanbul: TDK.
- ATLASI, Hadi 1911. *Seber Tarihi* [History of Siberia]. Kazan.
- BAITCHURA, Uzbek 1975. 'The Sound Structure of the Turkic Languages in Connection with that of the Finno-Urgic Ones. Part II: The Turkic Consonatism.' *Central Asiatic Journal* 19/4: 241–263.
- BAKHTIN, Mikhail 1982. *The Dialogic Imagination: Four Essays*. Austin: University of Texas Press.
- BERINDEI, Mihnea 1972. 'Le problème des 'Cosaques' dans la seconde moitié du XVIe siècle. A propos de la révolte de Ioan Vodă, voievode de Moldavie.' *Cahiers du monde russe et soviétique*: 338–367.
- BOBROVNIKOV, Vladimir 2007. 'Bandits and the State: Designing a 'Traditional' Culture of Violence in the Russian Caucasus.' In: Jane BURBANK, Mark VON HAGEN, and Anatolij Viktorovič REMNĖV (eds.) *Russian Empire: Space, People, Power, 1700-1930*. Bloomington: Indiana University Press, 239–267.
- BOESCHOTEN, Hendrik et al. 1995. *The Stories of the Prophets*. Vol. 2. Leiden and Boston: Brill.
- BUTAYEV, Shavkat, and Abbos IRISQULOV 2009. *English-Uzbek, Uzbek-English Dictionary: 70.000 Words and Expressions: 70.000 So'z Va Ibora*. O'zbekiston Respublikasi Fanlar Akademiyasi.



- CLAUSON, Gerard 1960. *Sanglax, A Persian Guide to the Turkish Language*, by Mahammed Mahdi Xan. London: Luzac.
- CLAUSON, Gerard 1972. *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford: Clarendon Press.
- DANKOFF, Robert, and James KELLY (eds.) 1985. *Diwân Lugât at-Turk by Mahmûd al-Kashgari*. Part III. Cambridge: Harvard.
- DOERFER, Gerhard 1963. *Türkische Und Mongolische Elemente Im Neupersischen: Unter Besonderer Berücksichtigung Älterer Neupersischer Geschichtsquellen, Vor Allem Der Mongolen- Und Timuridenzeit*. Vol. 3. Wiesbaden: F. Steiner.
- DRIMBA, Vladimir 2000. *Codex Comanicus: édition diplomatique avec fac-similés*. Bucharest: Editura Enciclopedica.
- ECKERT, Penelope 2012. 'Three waves of variation study: The emergence of meaning in the study of socio-linguistic variation.' *Annual review of Anthropology* 41: 87–100.
- ECSEDY, Hilda 1972. 'Tribe and tribal society in the 6th century Turk empire.' *Acta Orientalia Hung.* 25: 245–262.
- ERDAL, Marcel 1991. *Old Turkic Word Formation: A Functional Approach to the Lexicon*. Vol. 1. Wiesbaden: Otto Harrassowitz Verlag.
- ERMERS, Robert 2017. *Arabic Grammars of Turkic: The Arabic Linguistic Model Applied to Foreign Languages & Translation of 'Abû hayyân Al-'Andalusî's Kitâb al-'Idrâk li-Lisân al-'Atrâk*. Leiden and Boston: Brill.
- FRANÇOIS, Alexandre 2015. 'Trees, waves and linkages: models of language diversification.' In: Claire BOWERN and Bethwyn EVANS (eds.) *The Routledge Handbook of Historical Linguistics*. London and New York: Routledge, 161–189.
- GABAIN, Annemarie von 1960. 'Kasakentum, eine soziologisch-philologische Studie.' *Acta Orientalia Hung.* 11/1: 161–167.
- GAL, Susan, and Judith T. IRVINE 2019. *Signs of Difference: Language and Ideology in Social Life*. Cambridge: Cambridge University Press.
- GOLDEN, Peter B. 1980. *Khazar Studies: An Historico-Philological Inquiry into the Origins of the Khazars*, Vol. I. Budapest: Akadémiai Kiadó.
- GOLDEN, Peter B. 1992. *An introduction to the history of the Turkic peoples: Ethnogenesis and state-formation in Medieval and Early Modern Eurasia and the Middle East*. Wiesbaden: Otto Harrassowitz.
- GOLDEN, Peter B. 2007. 'Irano-Turcica: the Khazar Sacral Kingship Revisited.' *Acta Orientalia Hung.* 60/2: 161–194.
- GOLDEN, Peter B. 2009. 'Migrations, Ethnogenesis.' In: Denis SINOR (ed.) *The Cambridge History of Inner Asia: The Chinggisid Age*. Cambridge: Cambridge University Press, 109–119.
- GOLDEN, Peter B. 2011. 'Georgio-Turcica: Some marginal notes on pre-Ottoman/ Safavid Oğuz and non-Oğuz Turkic elements in Georgian.' In: Peter B. GOLDEN: *Studies on the Peoples and Cultures of the Eurasian Steppes*. [Ed. Cătălin HRIBAN] București – Brăila: Editura Academiei Române, 247–264.
- GOMMANS, Jos 2018. 'The Warband in the making of Eurasian empires.' In: Maaiké VAN BERKELAND and Jeroen DUINDAM (eds.) *Prince, Pen, and Sword: Eurasian Perspectives*. Leiden and Boston: Brill, 297–383.
- GOULD, Rebecca Ruth 2016. 'The Abrek in Soviet Chechen Literature.' In: R. R. GOULD *Writers and Rebels: The Literature of Insurgency in the Caucasus*. New Haven and London: Yale University Press, 33–91.
- GRANT, Bruce 2011. *The Captive and the Gift: Cultural Histories of Sovereignty in Russia and the Caucasus*. Ithaca: Cornell University Press.



- GRAEBER, David and David WENGROW 2021. *The Dawn of Everything: A New History of Humanity*. New York: Macmillan Publishers.
- GULYÁS, László and Gábor CSÜLLÖG 2015. 'History of Kosovo from the First Balkan War to the End of World War II (1912–1945).' *West Bohemian Historical Review* 15: 219–237.
- GYÖRFI, Dávid 2014. 'Khwarezmian: Mapping the Kipchak component of Pre-Chagatai Turkic.' *Acta Orientalia Hung.* 67/ 4: 383–406.
- HAKIMZJANOV, Farid Sabirzjanovič 1986. 'New Volga Bulgarian Inscriptions.' *Acta Orientalia Hung.* 40/1: 173–177.
- HEGGARTY, Paul, Warren MAGUIRE and April MCMAHON 2010. 'Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories.' *Philosophical Transactions of the Royal Society B: Biological Sciences* 365/1559: 3829–3843.
- HILDINGER, Erik (trans.) 1996. *The Story of the Mongols Whom We Call the Tartars by Friar Giovanni di Plano Carpini*. Boston: Branden Publishing Company.
- IRVINE, Judith T., Susan GAL, and Paul V. KROSKRITY 2009. 'Language ideology and linguistic differentiation.' In: Alessandro DURANTI (ed.) *Linguistic Anthropology: A Reader*. Malden, MA: Blackwell: 402–434.
- JOHANSON, Lars 1998. 'The History of Turkic.' In: Éva Ágnes CSATÓ and Lars JOHANSON (eds.) *The Turkic Languages*. London: Routledge, 81–125.
- JOHANSON, Lars 2000. 'Linguistic convergence in the Volga area.' In: D. G. GILBERS, J. NERBONNE and J. SCHAEKEN (eds.) *Studies in Slavic and General Linguistics*. [Languages in Contact, Vol. 28.] Leiden and Boston: Brill, 165–178.
- KEENAN, Edward L. 1979. 'Kazan' — "The Bend". *Harvard Ukrainian Studies* 3: 484–496.
- KHALIKOV, A.H. and J. G. MUHAMETSHIN 1977. 'Unpublished Volga Bulgarian Inscriptions.' *Acta Orientalia Hung.* 31/1: 107–125.
- KRÓLIKOWSKA-JEDLIŃSKA, Natalia 2018. *Law and Division of Power in the Crimean Khanate (1532–1774): With Special Reference to the Reign of Murad Giray (1678–1683)*. Leiden and Boston: Brill.
- KUUN, Géza (ed.) 1880. *Codex cumanicus: Bibliothecae ad templum divi Marci Venetiarum primum ex integro edidit prolegomenis notis et compluribus glossariis instruxit come Géza Kuun*. Budapest: Academia Scientiarum Hungarica.
- LEE, Joo-Yup 2015. *Qazaqlıq, or Ambitious Brigandage, and the Formation of the Qazaqs: State and identity in post-Mongol central Eurasia*. Leiden and Boston: Brill.
- LEEZENBERG, Michiel 2016. 'The Vernacular Revolution: Reclaiming Early Modern Grammatical Traditions in the Ottoman Empire.' *History of Humanities* 1/2: 251–275.
- LIGETI, Louis 1981. 'Prolegomena to the Codex Cumanicus.' *Acta Orientalia Hung.* 35/1: 1–54.
- MCELVENNY, James 2019. 'Alternating sounds and the formal franchise in phonology.' In: James MCELVENNY (ed.) *Form and Formalism in Linguistics*. [History and Philosophy of the Language Sciences 1.] Berlin: Language Science Press, 35–58.
- MENGES, K. G. [Карл Генрич Менгес] 1979. *Восточные элементы в 'Слове о полку Игореве'*. Ленинград: Наука.
- NÉMETH, Gyula 1930. *A honfoglaló magyarság kialakulása*. Budapest: Könyvnyomda.
- OLCOTT, Martha Brill 1987. *The Kazakhs*. Stanford: Hoover Institution Press.
- PELLIOT, Paul 1949. *Notes sur l'Histoire de la Horde d'or Suivies de Quelques Noms Tures d'Hommes et de Peuples Finissant en 'Ar'*. Vol. 2. Paris: Librairie d'Amérique et d'Orient.
- POPPE, Nicholas 1964. *Bashkir Manual: Descriptive Grammar and Texts with a Bashkir-English Glossary*. Bloomington: Indiana University.



- PRITSACK, Omeljan 2006. 'The Turkic Etymology of the Word "Qazaq" 'Cossack'.' *Harvard Ukrainian Studies* 28/1: 237–243.
- RÄSÄNEN, Martti 1949. *Materialien zur Lautgeschichte der türkischen Sprachen*. [Studia Orientalia 15.] Helsinki: Societas Orientalis Fennica.
- RÓNA-TAS, András 1982a. 'A kazár népnévről.' *Nyelvtudományi Közlemények* 84: 349–379.
- RÓNA-TAS, András 1982b. 'On the history of the Turkic and Finno-Ugrian affricates.' *Acta Orientalia Hung.* 36/1: 429–447.
- RÓNA-TAS, András 1983. 'Újabb adatok a kazár népnév történetéhez.' *Nyelvtudományi Közlemények* 85: 126–133.
- SAHLINS, Marshall 2003. *What kinship is—and is not*. Chicago: University of Chicago Press.
- SCHAMILOGLU, Uli 1990. 'The formation of a Tatar historical consciousness: Şihabüddin Märçani and the image of the Golden Horde.' *Central Asian Survey* 9/2: 39–49.
- SILVERSTEIN, Michael 2016. 'The "push" of Lautgesetze, the "pull" of enregisterment.' In: Nikolas COUPLAND (ed.) *Sociolinguistics: Theoretical Debates*. Cambridge: Cambridge University Press, 37–67.
- SINOR, Denis 1999. 'The Mongols in the West.' *Journal of Asian History* 33/1: 1–44.
- TEKİN, Talat 1991. 'A New classification of the Turkic languages.' *Türk dilleri araştırmaları* 1: 5–18.
- TOPARLI, Recep, M. Sadi ÇÖĞENLİ, and H. Nevzat YANIK 2000. *Kitâb-ı mecmû-ı tercümân-ı Türki ve Acemî ve Mugalî* [The Compiled Translator's Book of Turkish, Persian, and Mongolian]. Vol. 763. Ankara: Türk Dil Kurumu.
- UKUDEEVA-FREEMAN, Jamilya Tolenovna 2003. *Collective action problem: Mobilization of national-democratic movements in Azerbaijan and Kyrgyzstan*. (Ph.D. Dissertation. University of California, Riverside)
- URBAN, Mateusz 2017. 'European 'cassock', Tkc. *қазақ* and Per. *kazağand*. Relative chronologies and possible links.' In: Michał NÉMETH, Barbara PODOLAK, and Mateusz URBAN (eds.) *Essays on the History of Languages and Linguistics*. Krakow: Archeobooks, 753–791.
- VEINSTEIN, Gilles 1999. 'Early Ottoman Appellations for the Cossacks.' *Harvard Ukrainian Studies* 23/3: 33–44.
- ZIMONYI, István 1990. *The Origins of the Volga Bulgars*. [Studia Uralo-altaica 32.] Szeged: University of Szeged.
- ZIMONYI, István 1992. 'The Volga Bulgars Between Wind and Water (1220–1236).' *Acta Orientalia Hung.* 46/2: 347–355.

