

DISTINCTIVE DIACRITICS IN THE ATIL-TURKIC EPITAPHS AND IN THE YARKAND DOCUMENTS

Marcel ERDAL

Prof., Goethe University Frankfurt a. M., Department of Turcology, merdal4@gmail.com, ORCID: 0000-0002-1604-4193

Erdal, M. (2025). Distinctive diacritics in the Atil-Turkic epitaphs and in the Yarkand documents. In O. Çınar, F. Başbuğ & H. Aydemir (Eds.), *Contemporary studies in linguistics I: IMU Linguistics 15th anniversary commemorative volume* (pp. 109–116). Artsürem. <https://doi.org/10.7816/imuling-15-2025-01X005>

ABSTRACT

Scribes using the Arabic alphabet for writing Arabic, Persian or early varieties of Turkic used non-obligatory ‘distinctive’ diacritics beside the standard diacritics of the writing systems of their languages. After quoting Onur (2024) on the presence of such diacritics in Arabic, Persian, Khāqānī Turkic and Early Middle Turkic, the author documents their appearance in 13th-14th century epitaphs in Arabic and Atil Turkic, a language closely related to Chuvash, spoken in the Volga-Kama area of Russia, and in 11th-12th century legal documents in Uyghur script discovered in Yarkand, East Turkestan. The distinctive diacritics consist either of dots or of small Arabic letters placed either below or above the line of writing.

Keywords: The Arabic writing system, Atil Turkic, Early middle Turkic texts, Yarkand documents

ÖZ

Arap alfabesini Arapça, Farsça ya da Türkçenin erken dönem deęişkelerini yazmak için kullanan kâtipler, dillerin yazı sistemlerinde yer alan ölçünlü diyakritik işaretlerin yanı sıra, zorunlu olmayan “ayırt edici” diyakritik işaretler de kullanmışlardır. Yazar, Onur (2024)’te Arapça, Farsça, Hākānī Türkçesi ve Erken Orta Türkçede bu tür diyakritiklerin varlığına yapılan atıftan sonra, söz konusu işaretlerin 13.–14. yüzyıllara ait Arapça ve Rusya’nın İdil–Kama bölgesinde konuşulmuş ve Çuvaşça ile yakından ilişkili bir dil olan Atil Türkçesi mezar kitabelerinde ve Doęu Türkistan’da, Yarkent’te bulunmuş 11.–12. yüzyıla ait Uygur yazılı hukuk belgelerinde de görüldüğünü belgelemektedir. Ayırt edici diyakritik işaretler, ya noktalardan ya da yazı satırının altına ya da üstüne yerleştirilen küçük Arap harflerinden oluşmaktadır.

Anahtar Sözcükler: Arap yazı sistemi, Atil Türkçesi, Erken Orta Türkçe kaynaklar, Yarkent belgeleri

1. Introduction

The Arabic alphabet consists of 28 letters, representing consonants. Some of these letters differ from each other in that their simple (*rasm*) forms get a dot or two dots placed above of below them, or three dots placed above them: Thus, *b* differs from *t*, from *n* and from *θ* just by having a dot under it while *t* has two dots above it, *n* (in the beginning and middle of a word) one dot above it and *θ* has three dots above it. *γ*, *δ*, *ḍ*, *z* and *z̤* differ from ^c (*ayn*), *d*, *ṣ*, *r* and *ṭ* respectively just by having a single dot above them. *j* [dʒ] differs from *h* and from *x* just by having a single dot under it while *x* has one dot above it. *š* differs from *s* just by having three dots above it.¹ This consonant pointing, called *iʿjām*, is obligatory; however, some early Qurʾān manuscripts still only have the *rasm* forms of the letters without the *iʿjām* pointing.

The standard Arabic writing system also has a system of non-obligatory diacritic signs, called *taškil*; these include the vowel marks called *ḥarakāt*, the *sukūn* to show that no vowel is to be pronounced, the *šadda* above consonants which are to be pronounced doubly, the *tanwīn*, nunation, for the marking of indefinites and for adverb formation, the *madda* and the *waṣla* above the letter *alif* (the first to indicate a long /a:/, the second to mark positions where the *hamza* – the glottal stop – is deleted).

This paper is about a different set of non-obligatory diacritics hardly mentioned in descriptions of the writing system: the distinctive signs, called ‘Differenten’ in German, which were in use till the 14th century. The job of the distinctive signs was to make sure that *rasm* letters differing by *iʿjām* pointing were read and understood correctly, whether the *iʿjām* dots were actually present or were lacking: The central idea would have been that the *iʿjām* markings could have been effaced on the stone or in the manuscript, leading to a wrong reading. The distinctive diacritics consist of single dots under the letters to be read as ^c, *d*, *ṣ*, *r* or *ṭ*, and mostly of three dots under the letter to be read as *s*; there are some other distinctive signs as well, to which we will come when discussing the Turkic Yarkand manuscripts.

2. The Arabic Evidence

The oldest instance mentioned by Grohmann (1971: 42) for such signs consists of a dot placed under the letter *ṭ*, which would be read as *z* if there was a dot above it, in l. 2 of an epitaph from the year 883. In an 1163/64 epitaph from Baṣra, *d* is marked with a dot under it; if the letter had a dot above it, it would be read as *ḍ*. In a Baṣra epitaph from the year 1159 mentioned by Grohmann, the *s* is distinguished from *š* with a single dot above it, but in widespread use elsewhere, *s* is made to differ from *š*

¹ I am referring to the standard system used for Classical Arabic. I will not be going into phonetic or graphemic detail here, as the facts are well-known. Nor is knowledge of the terms of the next paragraph, part of Arabic grammar, necessary for understanding the present paper.

with three dots under it, whereas š has three dots above it: Grohmann thus mentions a 1226 inscription in Syria in the word شمس (*šams*, ‘the sun’), and in the word بسم (*bismi* ‘in the name of ...’) in the *basmala* on the sarcophagus of a certain Asʿad ibn ʿAli baq in Ġazna (Afghanistan, late 11th century) where the *s* is spelled as پ, although the context would have made the reading clear enough. These examples also show the geographical spread of the phenomenon in Arabic. Onur (2024: 850), footn. 47 mentions an Arabic manuscript with this feature, from the year 1356.

The early 12th century Arabic legal manuscripts edited by Gronke (1986) also show distinctive diacritics. She discusses the phenomenon on p. 458 with the phrase *aḥad ḥudūdihi* ‘one of its borders’ with dots under its first and second *dāls* to show that they are not *ḍāls*. No dot appears under the third *dāl*; it very often happens in all our sources that – without explanation – some of the relevant letters are marked with distinctive diacritics while others are not. The term *waṣiyy* ‘guardian’ has a dot under its *šād* to mark it as *šād*. The names *Yūsuf* and *Ilyās* and the title *sultān* are, e.g., spelled with three dots under their *sīn*, as is the Turkic title *sūbaşı*: The documents were commissioned by the same group of Turkic persons who were responsible for the ones edited by Erdal (1984). Gronke also mentions distinctive diacritics in the 13th-14th century documents from Ardabil (Persian Azerbaijan) which she edited, and in Arabic papyri discussed in another work by Grohmann.

3. Evidence in Other Languages

The Arabic script was used also for writing Persian and Turkic varieties and, as documented by Onur (2024: 849-852), the distinctive diacritics were put to use in texts written in those languages as well. For Persian, Onur mentions a pharmacological tractate from 1055/56, whose distinctive signs are already discussed by Orsatti (2019: 53). Further, an Andarz-name (a text with advice and injunctions for proper behaviour) from 1090, a 12th/13th century manuscript from Bukhara, and a document from Bāmiyān (Afghanistan) from the year 1211; full bibliographical details for all of these are given in the publication.

4. Distinctive Diacritics in Turkic

Onur discovered distinctive diacritics also in four Turkic manuscripts, examples of which are presented in facsimiles. The topic was brought up by the author in connection with the dating of the Tahrān manuscript of the Khwarazm-Turkic *Qīṣaṣu ʿl-Anbiyā*, completed by Rabġūzī in 1311: With a number of orthographical, grammatical and lexical arguments, he convincingly shows this manuscript to be much closer to the original than the British Museum manuscript, probably written down in the 15th century but hitherto thought to be the oldest and best source. One of the arguments which Onur gives for his early dating of the Tahrān manuscript is based on the recourse to the diacritics discussed in the present

paper.² Another source with distinctive diacritics which Onur gives is the Ferghana copy of the Kutadgu Bilig, clearly the earliest of three manuscripts of this Khākānī text: This copy is taken to have been written down in the first half of the 14th century. A third manuscript with this orthographic feature is the Middle Turkic-Persian interlinear translation of the Qurʾan, kept in the Abu Rayhan Biruni Institute of Oriental Studies of the Academy of Sciences of Uzbekistan with the archive number 2008; the editor of this text dated it to the 13th century. Fourthly, Onur found such diacritics in the first few pages of the 1365 manuscript of the Nahju ʿl-Farādīs, another Khwarazm-Turkic text.

5. The Atil-Turkic Texts

We will here discuss two other Turkic corpuses showing distinctive diacritics; one is the corpus of Atil-Turkic epitaphs dealt with in Erdal 1993, the other corpus the Muslim manuscripts discovered in Yarkand in East Turkestan and published in Erdal (1984). The Atil-Turkic epitaphs, mostly discovered in present-day Tatarstan, were hitherto called ‘Volga Bulgarian’, but the name proposed here is more appropriate: The Kama-Volga river system is in Greek, Arabic, Persian and Chaghatay sources called *Atil*.³ The name Bulgarian/Bolgarian is confusing for readers, as modern Bulgarian is a Slavic language spoken in Bulgaria. Atil Turkic is an important branch of the Turkic languages, which is abundantly represented in Hungarian loans and, prehistorically, had an important connection to Mongolic. In their editions of the Atil-Turkic epitaphs, Róna-Tas & Fodor (1973), Khakimzyanov (1978), Khakimzyanov (1987) and Tekin (1988) mark distinctive diacritics both in their transliterations and their transcriptions, and Róna-Tas and Fodor mention them in their notes as well.⁴ This was clearly a general practice, in view of the fact that a number of the marks must, with time, have become unclear on the stone and could not be seen by the scholars.

The epitaphs are bilingual: Beside the Atil-Turkic texts formulated by the relatives of the deceased or by local professionals employed by

² Edited in Çelebi Çam (2025: 203-663). Note that the Tahrān ms. also has numerous errors.

³ Concerning Greek sources, see the entry Ἀτίλ in Moravcsik 1983. Aḥmad Ibn Faḡlān, who participated in the Abbasid expedition to the Atil Turks in 921-922, calls the large river flowing through the territories of the Bulgarian and Khazar Turks *Atil*, with *madda* above the onset *alif* in all instances, in his travelogue edited by Togan (1939). According to Togan, the name is spelled like this also by al-Bīrūnī in the 11th/12th century and by al-Qazwīnī in the 13th/14th century. In his *Shejere-yi Terākime* written in 1659/60, Abulgāzi Bahadır Khan also mentions the *Atil suyi*, the Atil river: According to the note in Ölmez (2020): 266, the name appears with (non-obligatory) *madda* over the *alif* in 5 among the 7 instances in the T manuscript.

⁴ In each of these publications, only parts of the corpus were edited; Tekin (1988), e.g., only has 90 among the 139 epitaphs discussed in Erdal (1993). I have not tried to decipher the facsimiles published in these and in earlier editions (referred to in the mentioned editions), as many of them are not clear and some appear to have been redrawn; some of the distinctive diacritics may have been missed by the editors, but there is no need for the present paper to be exhaustive. There is no reason to think that the editors added distinctive diacritics which are not there, as these marks were not topics of scientific discussions (e.g. whether the language is closer to Chuvash or to Tatar).

them, there are purely Arabic sentences traditionally found in Muslim epitaphs, showing a limited variation. The language used by the community which left us their epitaphs also had Arabic and Persian loans, partly subject to Atil-Turkic sound laws. The distinctive diacritics found in our epitaphs appear to concern only graphemes, whether the words in which we find them were part of the international Arabic epitaph corpus not necessarily meant to be understood, or of the Atil-Turkic corpus (including its loans).

In the Arabic sentences we find three dots under *sīn* to show that it is not *šīn* e.g. with the noun *nās* ‘people’ or the verb *sayamūtu* ‘they will die’; single dots under *dāl* to show that it is not *ḍāl* e.g. with the verb *dāhiluhu* ‘they will enter it’, under *rā* to show that it is not *zāy* e.g. with the adjective *kabīr* ‘great’, under *‘ayn* to show that it is not *ḡayn* e.g. with *al-‘aliyy* ‘the superior’.

The Turkic text obviously includes many proper names, e.g. *Ismā‘īl* with dots under the *sīn* and the *‘ayn*, *Sāra* with dots under the *sīn* and the *rā* or *Xar(t)mās* with a dot under the *rā*. There are Arabic loans, e.g. the dative / accusative forms *āxirete* ‘to the thereafter’ with a dot under the *rā* and *mesjidsemne* ‘mosques’ with a dot under *dāl* and three dots under *sīn*. The names of the months are Arabic as well, e.g. *šafar* with a dot under the *šād* or *rabi‘ul āxir*⁵ with a dot under the *‘ayn* and a second one under the second *rā*. The numbers are well represented: Note, e.g., the spellings of *toxur* ‘nine’ with dots under the *ṭā* and *rā*, or *sekir* ‘eight’ with three dots under the *sīn* and one under the *rā*. Even better represented are the ordinal numbers, e.g. *altūši* ‘sixth’ with a dot under the *ṭā* and *bīrim* with a dot under the *rā*. *hīr* ‘daughter’ is often spelled with a dot under the *rā* and *erle-sa‘at* ‘early hour’ has dots under the *rā*, *sīn* and *‘ayn*. Finally there are infinite verb forms, among them *batuwi* ‘that (s)he went’, *wafāt(i) baltuwi* ‘that (s)he died’, both with dots under their *ṭās*, and *tanruwi* ‘that he did’; the modal form *baltur* ‘let it be!’ has distinctive dots under its *ṭā* and its *rā*.

Differentiae are noted in 49 of the Atil-Turkic epitaphs, dated between 1291 and 1358.

6. The Yarkand Documents⁶

The Turkic Yarkand documents, linked to the Arabic ones of Gronke (1986) both by content and by persons involved, were edited in Erdal (1984); Salur (2024) is an actualized Turkish translation of this paper, supervised by the present author. The graphemic system of these texts is Uyghur: no letters are taken over from the Arabic script, as Şinasi Tekin, a previous editor, thought. Uyghur /n/, /ŋ/ and /r/ look alike, but the first often has one dot above it (as in Uyghur script and not just in Arabic),

⁵ Spelled like this.

⁶ Clark (2010: 93-104) already discusses the features of the script described here, connecting them with Kāšġarī’s statements about the writing systems.

the third two dots; I will not be mentioning these normal features of the Uyghur writing system. However, various Arabic signs are used diacritically to distinguish cases in which the same Uyghur letter, or two or more Uyghur letters looking identical in certain positions, represent more than one phoneme or allophone. /z/ is often distinguished from final /n/ by two dots placed under it, while in the word *Težik* in I,32 there are three dots under an R (with the regular two dots above it): This must be reminiscent of the Persian spelling of the specifically Persian sound ž with three dots over R.

In the Arabic tradition, three dots under the letter S show that it is not to be read as /š/. This is opposite to Uyghur practice, where /š/ is occasionally marked against /s/ with two dots underneath.⁷ Our texts also often mark the /s/ with dots, and not the /š/: *qanmasa* ‘if he is not satisfied’ in I,17, the proper name *Arslan* in II,b3, the proper name *Narsi* in III,1 and *Ishaq jal(l)ab-qa* ‘to Mr. *Ishaq*’ as well as *kesek* ‘piece’ of II,3c3 have three dots each under their S's. In IV,2, the title to be read as *sübaşı* has a single dot under its fifth letter. *igsiz* ‘healthy’ in IV,3 has three dots under its S but there is a single dot under the S of the proper name *Sökmen* in IV,4. The S's of *kesek*, of *sat(t)ım* ‘I sold’ in IV,5 and IV,6 and of a verb derived from a Persian loan, *usparladım* in IV,8, have double dots underneath. *Ishaq* in IV,5 has a single light dot under its S. *Sulman* in V,a1 and *silihdar* ‘man-at-arms’ in V,a5 both have three dots under their S. The title *teksin* in V,b7 is written with two or three dots under S. In V,b8 *sarıg* ‘blond’, two dots are visible under the S. *inisi* ‘his younger brother’ in V,b9 has three dots under the S, while the same word in IV,3 has two joined dots above its S. The last instance being an exception, all /s/-dots – mostly but not always three – are *under* the letter.

7. Diacritic Letters

A single Uyghur letter, Q, stands for [q], [ç], /^ç/, and /h/. [q] can be marked by a superscribed single dot, in *Ishaq* in II,3c3, in *tanuqlar* ‘witnesses’ in IV,1 and in *qılsar* ‘if he does’ in IV,10. All other distinctive markings come from Arabic letters: *ğayn* or ‘ayn for [ç] and /^ç/, a small *hā* for /h/ and /x/ and small *hā* for /h/. In I,15, an Arabic *ğayn* is visible under the Q of *oylumız* ‘our son’, in I,23 small *ğayns* under the Q of *siçğan* ‘mouse’ and under the ‘ayn of *rabi^çal ahir* (spelled like this, and with Q as fifth letter). *hallāj* ‘cotton carder’ in I,30 has a *hā*-like sign under its first letter; in V,b5, the diacritic *hā* is written *over* its first letter. The instances of *haddi* ‘its border’ in II,c1, c5 and c7 have small *hās* under their first letter, but *haddi* in IV,6 has something like an Arabic *hā* above it. In *Ishaq* of II,3c3 a small Arabic *hā* appears under the third letter. The Persian loan *baha* ‘price’ in II,c6 and twice in IV,7 has the Arabic letter *hā* above it.

⁷ It is also contrary to what we find in p. 6 of the *Dīvān Luğāti* ‘t-Turk: In a list of the “Turkic” alphabet which Maḥmūd offers with Arabic correspondences, the Uyghur letter corresponding to Arabic *šīn* has two dots under it, while the one corresponding to *šīn* has nothing.

taqaġu ‘chicken’ in II,b1 has an ‘*ayn*’ below its second Q. The names to be read as ^c*Uθmān* in V,b2 and ^c*Alī* in V,b4 have little ‘*ayns*’ under their first letters.

The early practice for writing Arabic texts is not limited to dots either: The *hamza* also, after all, comes from a small ‘*ayn*’. Grohman (1971: 43-46) mentions other signs, mostly simplified letters, which were placed under or over the *rasm* letters to ease their identification. Gronke (1986: 458), e.g., mentions a letter *fā* added above the *fā* of the text.

8. Conclusion

What is unique in the Turkic Yarkand documents is that the distinctive diacritics appear in texts which are not only in a different language from Arabic, but also are imposed on a different writing system – albeit used in situations where the Arabic language had a prominent position. The paper gives evidence for a widespread graphic practice in early Muslim epigraphic and manuscript texts for markings beyond the standard system, originally intended to prevent misreadings, but often also used when misreadings were unlikely.

References

- Clark, L. (2010). The Turkic script and the *Kutadgu Bilig*. In H. Boeschoten & J. Rentzsch (eds.) *Turcology in Mainz*. Wiesbaden.
- Çelebi Çam, S. (2025). *Kısaşü'l-Enbiyā (Rabġūzī) Tahran Nūshası (transkripsiyonlu metin). Tahran ve Londra nüshaları arası eş anlamlılık*. Ankara.
- Erdal, M. (1984). The Turkish Yarkand documents. *Bulletin of the School of Oriental and African Studies*, 47, 260–301.
- Erdal, M. (1993). *Die Sprache der wolgalbolgarischen Inschriften*. Harrassowitz. Wiesbaden.
- Grohmann, A. (1971). *Arabische Paläographie. 2. Teil: Das Schriftwesen. Die Lapidarschrift*. Böhlau Nachf.
- Gronke, M. (1986). The Arabic Yarkand documents. *Bulletin of the School of Oriental and African Studies*, 49(3), 454–507.
- Khakimzjanov, F. S. (1978). *Jazyk epitaġii volġskix Bulgar*. Moscow.
- Khakimzjanov, F. S. (1987). *Epigrafičeskie pamjatniki volġskoj Bulgarii i ix jazyk*. Moscow.
- Moravcsik, G. (1983). *Byzantinoturcica II: Sprachreste der Turkvölker in den byzantinischen Quellen*. Brill.
- Onur, S. (2024). Which is the oldest manuscript of Rabġhūzī’s *Qisas al-Anbiyā*? *Türkiyat Mecmuası*, 34(2), 839–875.
- Orsatti, P. (2019). Persian language in Arabic script: The formation of the orthographic standard and the different graphic traditions of Iran in the first centuries of the Islamic era. In D. Bondarev, A. Gori, & L. Souag (eds.), *Creating standards: Interactions with Arabic script in 12 manuscript cultures* (pp. 39–72). De Gruyter.
- Ölmez, Z. (ed.). (2020). *Ebulgazi Bahadır Han. Şecere-yi Terākime*. TDK.

- Róna-Tas, A., & Fodor, S. (1973). *Epigraphica Bulgarica. A volgai bolgár-török feliratok* (Studia Uralo-Altaica, 1). Szeged.
- Salur, G. (2024). Türkçe Yarkent belgeleri. *Hacettepe Üniversitesi Türkiyat Araştırmaları Dergisi*, 41, 261–327.
- Tekin, T. (1988). *Volga Bulgar Kitabeleri ve Volga Bulgarcası*. TDK.
- Togan, A. Z. V. (1939). *Ibn Faḍlān's Reisebericht* (Abhandlungen für die Kunde des Morgenlandes XXIV, 3). Leipzig.