Dmitry Bondarev, Darya Ogorodnikova

How to Spell Loanwords? Integration of Arabic Etymons in Bilingual Islamic Manuscripts of West Africa

Abstract: Different stages of the integration of Arabic loanwords into West African languages are studied on examples of spelling patterns identified in two distinct manuscript traditions of Old Kanembu and Soninke. Both languages are written in Arabic script (Ajami) providing interlinear translations of Arabic texts. These Ajami writings exhibit a high degree of Arabic loanwords, some spelled etymologically and some deviating from the Arabic graphemic source to the point of complete dissimilarity. The paper outlines a preliminary typology of the spelling of Arabic loanwords in interlinear Ajami writings and suggests that retentions and divergences from etymological spelling are probably motivated by established orthographic practices specific to teaching–learning circles, whereas the individual linguistic sensitivity of scribes does not seem to play a prominent role in the selection of spelling features.

1 Introduction

Arabic came to sub-Saharan Africa with the spread of Islam, starting from the ninth century CE.[1] Lexicons of regional languages have since been under the significant influence of Arabic. Many sub-Saharan societies with a long history of Islam or contact with Muslims had parts of their vocabulary changed and expanded by Arabic borrowings covering a wide semantic range, including religion, administration, warfare, science, trade, time and counting systems and mathematics. Some Arabic borrowings have been entirely integrated into the lexicon of the target languages, so that the Arabic etymons have changed considerably, for example, Soninke *sèyìdì* from Ar. *shahīd* 'martyr', Hausa *lādân* from Ar. *al-ʾaðān* 'muezzin' or Kanuri *ashâm* from Ar. *aṣ-ṣiyām* 'fast(ing)'. However, a large part of Islamic vocabulary remains transparent as to its source in Arabic.

---

1 On the chronology of propagation of Islam in sub-Saharan Africa, see Hiskett 1984; Levtzion and Pouwels 2000; Insoll 2003; Salvaing 2020.

The relative chronology, different mechanisms and types of borrowing from Arabic into sub-Saharan languages have been predominantly studied based on spoken data. How the process and result of borrowing were reflected in writing is largely unknown. This is partially due to the paucity of early written data. Epigraphic evidence, for example, of the interaction between Arabic and Songhay on funeral inscriptions dates to the thirteenth to fourteenth centuries. But the type of texts incised onto cliffs and tombstones is too short and too Arabic-oriented to be of significance for investigating the borrowing processes.[2] Writing on paper opens up a much more generous resource, which, however, is of much later time. The most 'ancient' manuscripts where Arabic is accompanied by local languages in Arabic script (Ajami) go back to the mid-seventeenth century, and even this date is exceptional – only Qur'an manuscripts from Borno have such a noble age. The rest of the sub-Saharan region of West Africa feature Arabic-Ajami manuscripts dating, at best, to the late eighteenth century.[3]

This kind of bilingual and multilingual Islamic manuscripts from West Africa have only recently started emerging from obscurity, and the novelty of the material explains its absence from the previous research on Arabic loans in sub-Saharan languages. This paper is just a first step to complement studies of lexical borrowing in languages of West Africa by manuscript data by taking a preliminary look into some select Ajami manuscripts from two unrelated regions of the Senegambia and Lake Chad. Our slightly more ambitious aim is to sketch some tendencies of loanword spelling observable in manuscripts and map distances from the original Arabic graphemic words to their spelling in the recipient language (what we call a 'gradient of etymological spelling'). We then relate the types of spelling thus obtained to the question of graphemic visual interaction between the donor and recipient language (spelling affected by the visual proximity to the original, and/or by the level of the scribe's literacy) and will ask whether the typology may have the potential for the study of the relative chronology of borrowing and degrees of incorporation of the loanwords into the recipient language.

The paper is structured as follows. We sketch out a brief history of adaptation of Arabic script to the languages of West Africa in the next section. Section 3 deals with previous studies of channels and patterns of Arabic loanwords in these languages. Section 4 starts with the specification of the kind of linguistic data represented by the interlinear Ajami material under study (Subsection 4.1). We then introduce our methods and principles for the study of loanwords in

---

2 De Moraes Farias 2003. On epigraphy in Essouk-Tadmakka, see Nixon 2017.

3 See a summary on types of Ajami manuscripts in Bondarev 2021, 708.

interlinear Ajami (Subsection 4.2). Section 5 presents an analysis of selected etymons in Old Kanembu (Subsection 5.1) and Soninke (Subsection 5.2) manuscripts. Section 6 provides a comparative summary of our findings, and Section 7 draws some conclusions.

2 Adaptation of Arabic script: Ajami

Inscriptions on cliffs and tombstones provide the earliest evidence of the interaction between Arabic and local languages, for example, Songhay. But the words in Songhay are very few, the texts in Arabic do not betray any linguistic influence of Songhay, and nothing can be said about the process of borrowing and adaptation of the Arabic lexicon at the time when the inscriptions were made. We should, thus, start with much later written practices witnessed by manuscripts on paper. The first such evidence comes with the Qur'an manuscripts of the Borno sultanate, what is now roughly north-east Nigeria and south-east Niger. We find an enormous amount of annotations in Old Kanembu, a language that was exclusively used for the translation of the Qur'an and other Arabic texts used in traditional Islamic education and scholarship, in the interlinear and marginal space of the manuscripts written from the mid-seventeenth century to the early nineteenth century.[4]

Soninke, the language spoken in north-eastern Senegal, south-eastern Mauritania and western Mali, is another sub-Saharan language that has an early written attestation comparable to Old Kanembu. At least one manuscript with interlinear annotations in Soninke to a theological treatise by Muḥammad b. Yusūf al-Sanūsī (d. 1486) is datable to the late eighteenth century.[5] Writing in Soninke must have been practiced much earlier however, because it was probably one of the earliest languages exposed to Islam in the western Sahel.[6]

---

4 Old Kanembu was a Western Saharan language closely related to Kanuri and Kanembu of the same sub-branch in the Saharan languages. Circumstantial evidence points to much earlier practices of writing in Old Kanembu, very possibly before the fifteenth century. On this topic and also on relationship between Old Kanembu, its modern-day descendant exegetical translational language Tarjumo and Kanuri/Kanembu dialects, see, *inter alia*, Bondarev 2013a; 2013b; 2014a; and 2014b.

5 See Ogorodnikova 2023.

6 Soninke was the language of the ruling elite of ancient Ghana, one of the first West African polities to adopt Islam; see Cuoq 1984.

There is no certain periodisation of Ajami writing in the other languages of West Africa, and various publications give different chronological pictures. Thus, Tamasheq, a Tuareg Berber language mostly spoken in the north of Mali, has been reported to exist in a sixteenth-century manuscript, although the proof of this claim is yet to be seen.[7] Songhay, after its prominent, if fragmented, appearance in medieval stone epigraphy, seems to reappear in manuscripts only in the late eighteenth century, but even this dating is speculative because the scribes are typically tacit about the time of their writing. The chronology does not improve in the case of Fulfulde (Fula), the language of the Fulbe (Fulani) who live across West Africa from Senegal to Chad and were the major propagators of Islam across the whole region in the nineteenth century. Possible earliest hard evidence has been found in manuscripts going back to the eighteenth century.[8] The Wolof language of Senegal, known for its significant body of contemporary literature in Ajami (Wolofal), was used in 1817 in a bilingual treaty between the Bar kingdom and France.[9] Incidentally, 1817 is also the date for the first extant example of Ajami in Hausa, the language spoken in the past in several urban kingdoms to later evolve into the most spoken language of West Africa, centred in northern Nigeria and southern Niger.[10]

As is the case of Old Kanembu and Soninke, the other languages must have existed in written form much earlier than the earliest manuscript evidence suggests. A French travel account of the mid-seventeenth century, for example, reports that Mandinka spoken in the lower Gambia region was written in Arabic characters.[11]

In spite of the chronologically limited range of the extant manuscript data, there is one clear tendency in the development of various Ajami orthographies, especially visible in Old Kanembu, Soninke and Hausa writings: the older the manuscripts, the more the conservative approach is manifest in the adaptation of Arabic script. Thus, the Arabic characters in the earliest Ajami writings were rarely modified to better represent the sounds of the target languages and no innovation is visible before the late nineteenth century. In Old Kanembu, for example, only nineteen of twenty-eight Arabic consonants were actively used

7 Gutelius 2000. On Ajami writing in the other Berber languages (beyond the scope of this survey), see van den Boogert 1997.

8 Salvaing and Hunwick 2003.

9 Ngom 2017.

10 'The first dated example of Hausa Ajami (in a multilingual text) was written in the Caribbean diaspora in 1817', Bondarev and Dobronravin 2019, 254.

11 Cultru 1913.

for the reconstructed twenty-three consonants.[12] The main selection principle was homorganic proximity, i.e. the phonetic similarity in the place and manner of articulation. The Arabic letter <d>, for example, was chosen for the sound [d]. One or more distinctive features are omitted in less straightforward cases. Thus, for a bilabial voiceless plosive [p] absent in Arabic, the choice was either the letter <b>, which encodes a phoneme with two features shared with the [p], i.e. 'bilabial' and 'plosive' (but lacking 'voiceless'), or the Arabic letter <f>, sharing (in its phonetic representation) the features 'bilabial' and 'voiceless' (but lacking 'plosive').[13]

Some Arabic letters that represent sounds absent in Soninke, are used for specific Soninke sounds in Soninke manuscripts. Thus, the letter *ʕayn* is conventionally used for the Soninke /ŋ/. In rare instances, the letter is used as a graphemic support for a vowel diacritic. The letter *ðāl*, which represents the interdental fricative [ð] in Arabic, is 'reinterpreted' for the sound [t͡ʃ] (henceforth [c] in simplified transcription), or sometimes for the sound [d͡ʒ] (henceforth [j]), the latter sound being also represented by the letter *jīm*. The same letter *ðāl* is used in some manuscripts to encode the voiced palatal approximant [j] (henceforth [y] in the conventional notation of the Soninke phonology).

The Arabic consonants that were only sparsely used in non-Arabic Ajami content range from nine in Old Kanembu Ajami (*ṣād, ḍād, ṭāʾ, ẓāʾ, ʕayn, qāf, xāʾ, ðāʾ* and *šīn*) to eight in Soninke (*ṣād, ḍād, ṭāʾ, ẓāʾ, zāy, xāʾ, θāʾ* and *šīn*).[14] We will call these letters *Arabic-specific*. As will become obvious in the following sections, the scribes of respective manuscript traditions had the letters in reserve for spelling Arabic loanwords.

The special status of the Arabic-specific letters implies a conscious attitude to spelling in multilingual writing which goes hand in hand with scribes' self-

---

12 Bondarev and Dobronravin 2019, 244–245.

13 The encoding of vowels in early Ajami texts exhibits the same conservative principle. The picture depicted here is intentionally simplified to emphasise the tendencies. However, there is more to the selection of Arabic graphemes than can be sketched out in this short overview. For more details on orthographies in various Ajami traditions, see Souag 2010 and Mumin and Versteegh 2014; and for specific languages, see Bondarev 2014b on Old Kanembu; Dramé 2021 and Ogorodnikova 2023 on Soninke; Newman 2000 and Bondarev and Dobronravin 2019 on Hausa; McLaughlin forthcoming on Atlantic languages, such as Fula and Wolof; Vydrine 1998 and 2014, and Tamari 2017 on Manding.

14 In Soninke manuscripts analysed by Dramé 2022, *xāʾ* is regularly used for the Soninke /x/. In the Soninke manuscripts discussed in this study, the sound phoneme /x/ is usually expressed by the letter *qāf*, but can also be denoted by *kāf*. A rarer set of graphemes for this sound is *ɣayn* and *xāʾ*.

perception, evident in how they labelled non-Arabic content with the term *ʕajami* and other metalinguistic means.

The scribes were not just conscious of differences between their language and Arabic – this is well expected in translational context. They were marking their Ajami writing by metalinguistic means, such as a short tag signalling that the writing is in Ajami, or describing their phrases as written 'in our language'.[15] The spelling choices made by the scribes in writing loanwords may have been prompted by the use of such conscious metalanguaging, as will be discussed later.[16]

3 Arabic influence on lexicons of West African languages and types of borrowing

The Arabic lexical load on vocabularies of West African languages is substantial and the languages spoken by communities with a long history of Islam have a particularly high number of Arabic loans. Lexical borrowing comes about by a variety of routes. The earliest attempt to map different channels and patterns of Arabic borrowings across the breadth of languages of West Africa is by Paul Wexler.[17] With the exception of Sergio Baldi, later lexicographical studies dedicated to Arabic loans have focused mostly on individual contagious or co-regional languages.[18]

Our study is, in many ways, consonant with Wexler's approach to disentangle different patterns of borrowing from Arabic, and we briefly outline some of his concepts here. In an attempt to 'distinguish between borrowed vocabulary defined by the speakers themselves as "Arabic" (regardless of the origin) and

---

15 Ogorodnikova 2017; Ogorodnikova 2023; Bondarev 2021.

16 On the concept of metalanguaging, see Maschler 1994.

17 Wexler (1980) examined data available at the time for sixteen languages: Tamasheq (Berber family); Kanuri, Teda (Saharan); Dangaléat, Hausa, Mandara (Chadic); Maba; Fula, Wolof (North-Central Atlantic); Temne (Mel, of Atlantic-Congo macro phylum); Songhay (Nilo-Saharan?); Dagbani, Moore (Gur, of Atlantic-Congo); Bambara, Busa, Manding (Mande) and Mbay (Central Sudanic).

18 Baldi 1999 deals with the periodisation of Arabic loans in Chadic languages and Baldi 2008 provides large-scale surveys of Arabic loans across African languages. Kossmann 2005 is an exemplary study of Arabic loans in Hausa and Kanuri contrasted to loans from Berber. An excellent overview of literature and topics dealing with the Arabic impact on sub-Saharan languages is given in Souag 2016.

genuine Arabic loans no longer identified as such',[19] Wexler draws a distinction between 'Arabisms' and 'Arabic elements'. Thus, 'Arabisms' are (1) the words purposefully borrowed from Arabic, precisely because of their Arabic origin, and (2) the words recognised by the speakers to be of Arabic origin (even if wrongly). 'Arabic elements' are Arabic loans 'no longer recognised as such by speakers'[20] as a result of a 'chance diffusion of Arabic elements'.[21]

Islamisation was a gradual process, initially involving elite social groups and only later reaching out to wider populations, therefore, a useful distinction has to be made between the loans coming at an earlier stage of demographically narrower interaction with Islam and those introduced later into predominantly Muslim societies. However, the identification of such different stages of borrowing is complicated by 'the possibility that the original norms of Arabic in the target language may have undergone change' and, therefore, 'the age of the borrowing cannot easily be determined simply from the type of formal integration in the target languages'.[22] Another complicating factor is the possibility of different or multiple contacts of the target languages with the secondary intermediate channels of Arabic etymons. The eastern regions of the Hausa language, for example, were exposed to indirect contact with Arabic from the east through its contact with Kanuri, whose speakers were influenced by Islam from the eleventh century, and then later, around the fifteenth century (and simultaneously with the ongoing Kanuri contacts), the western regions of Hausa came into contact with Songhay. At the same time, Islamic education was constantly developing, encompassing a greater number of people, leading to new or renewed direct borrowing from written Arabic.

> The result of widespread reciprocal impact is that West and Central African languages frequently have doublet forms of a single Arabic etymon – one form borrowed directly from Arabic and one form introduced through a neighbouring lingua franca.[23]

A classification similar to Wexler's has been advanced by Nico van den Boogert in his study of the Berber literary tradition of the Sous.[24] Van den Boogert distinguishes between four types of borrowing in written Berber: quotations, classi-

---

19 Wexler 1980, 524–525.

20 Wexler 1980, 525.

21 Wexler 1980, 526.

22 Wexler 1980, 529.

23 Wexler 1980, 531.

24 Van den Boogert 1997.

cisms, colloquialisms and Berberised loans.[25] Quotations are loans from Classical Arabic that retain the standard Arabic orthography. Classicisms are borrowings of nouns with their morphological properties, such as plural and singular forms, the definite article and case affixes. Colloquialisms are borrowings from Moroccan colloquial Arabic. Finally, the Berberised loans consist of the oldest Arabic layer of borrowed verbs, and are distinguished from the other loans by their complete integration into the Sous Berber morphology.

Types of borrowing discussed by Wexler and van den Boogert are commensurate with the general framework of lexical borrowing. Gerrit J. Dimmendaal, drawing on the examples of African languages, outlines different types and mechanisms of borrowing.[26] In order to avoid terminological ambiguity, we first list some of the relevant concepts discussed by Dimmendaal, and then align the terminology we use in our study with that of Wexler, van den Boogert and Dimmendaal.

Our data (in Section 5) is introduced from the less to maximally integrated loans, therefore, our selection of terms mentioned by Dimmendaal is also given in a similar order. The first term is *unadapted borrowing*, which refers to the type of borrowing when items are not phonologically integrated into the target language. Unadapted borrowing occurs 'in particular when there is widespread bilingualism'.[27] The second type is *morphosyntactically integrated* borrowings without phonological integration of the borrowed item.[28] The third type is *loan blends* or *hybrids*, which refers to 'partial morpheme substitution in loanwords'.[29] The fourth type is *phonological adaptation*, that makes the loanword completely integrate into the target language by various mechanisms, from the insertion and deletion of sounds to phone substitution and phonological restructuring.[30]

In our analysis of Arabic loans, we try to ascertain to what extent the scribes were making intentional decisions in their spelling choices (from less integrated to more integrated). This sociolinguistic dimension is also discussed by Dimmendaal under the notion of an 'act of perception',[31] whereby the speakers are conscious of their multilingualism, prompting unadapted borrowings.

---

25 Van den Boogert 1997, 223–224.
26 Dimmendaal 2011, 179–188.
27 Dimmendaal 2011, 182.
28 Dimmendaal 2011, 182.
29 Dimmendaal 2011, 185.
30 Dimmendaal 2011, 182.
31 Dimmendaal 2011, 183–184.

We are now in a position to introduce the terminology we use in our study in comparison to the concepts by the three authors mentioned above. Note that not all concepts are fully compatible with each other.

Table 1: Terminology in comparison.

This study	Wexler	van den Boogert	Dimmendaal
insertion	Arabisms (1) the words borrowed from Arabic purposefully	quotations	unadapted borrowing
copy-spelling; etymological spelling	Arabisms (2) the words recognised by the speakers as of Arabic origin	classicisms	morphosyntactically integrated
partial etymological spelling			loan blends or hybrids
complete integration; adapted spelling	Arabic elements	Berberised	phonological adaptation

4 Types and mechanisms of borrowing discernible in manuscripts

All data on Arabic loans in African languages, except for van den Boogert's research into the Berber literary tradition,[32] have come so far from spoken African languages. The general assumption that written Classical Arabic was a significant (or even the primary) point of contact between West African languages and Arabic has never been corroborated by the study of Ajami writing.[33]

The type of written data presented in this study has the potential for testing various assumptions about borrowing processes related to written Arabic. Unlike independent Ajami texts that exist on their own without direct relationship with Arabic source texts, the interlinear annotations are in constant interplay

---

32 Van den Boogret 1997.
33 Wexler's important conclusion that 'written Arabic can be considered the most important contributor of Arabic elements to the West and Central African languages spoken by Muslims' (Wexler 1980, 556) results from his study of grammars and dictionaries that do not deal with literary variants of the languages.

with the Arabic in the translational source-target frame. At the same time, this kind of written texts has its difficulties and limitations. We, firstly, discuss the peculiarities of such written data and then introduce our method of dealing with the spelling of Arabic loans.

4.1 Peculiarities of Ajami texts in interlinear annotations

Linguistic data represented by the Ajami manuscripts under study are highly specialised. The Ajami texts consist of glosses that translate the Arabic texts, mostly of a religious genre which deal with various subjects of the classical Islamic curriculum. The manuscripts were produced in the domain of intermediate and advanced classical Islamic education.[34] Ajami glosses represent a written counterpart of oral exegetical practices based on the translation of small units of the source text, most typically parsed into short noun phrases and verb phrases. In spite of such a source-dependent relationship between the units of the parsed Arabic texts and the units of translation, the translational output is rarely influenced by the grammatical structures of Arabic, and the phrases of the target language are usually grammatically well-formed.[35] At the same time, the Ajami glosses represent specialised exegetical varieties of the respective languages which differ from the ordinary spoken languages and may, therefore, be incomprehensible to the speakers outside the learned circles.[36] One specific characteristic of these exegetical and translational registers is their technical vocabulary, drawn largely from Arabic.[37] In this context, loanwords are often preferred over the language-internal vocabulary to unambiguously convey the concepts of the main text. Such an increased load of the Arabic vocabulary in exegetical spoken and written practices makes it difficult to draw a line between the borrowings already incorporated into the target language and the specialist 'technical' terms used exclusively in the educational and learned domain.

A similar problem of differentiating borrowings from specialist vocabulary was identified by Maarten Kossmann in relation to the linguistic content of Berber manuscripts:

---

34 Bondarev 2017; Ogorodnikova 2021.
35 Tamari and Bondarev 2013, 9; Bondarev 2022.
36 Tamari and Bondarev 2013, 15–22; Bondarev 2022.
37 Tamari and Bondarev 2013, 15–22.

> In Islamic treatises and admonitions, the text genre entails the usage of much Arabic vocabulary. Some of this is no doubt genuine borrowing, while others are necessary insertions in order to explain concepts not nameable otherwise. There are also many terms that seem to be inserted from Arabic, even though there are Berber forms available.[38]

Van den Boogert also stressed the increased 'use of Arabic loans instead of an Existing [sic] Berber word' as a characteristic of Berber manuscript verse texts.[39]

Another complicating factor in the analysis of borrowings in the glosses is that literate scribes were conversant with the phonology and morphology of the source languages. Therefore, they probably adhered to accurate pronunciation[40] and orthographic rules, both of which may trigger the etymological spelling of Arabic borrowings. As a result, manuscript glosses are likely to capture and reflect scholarly attitudes towards borrowings rather than general tendencies or patterns of integration into ordinary language. Moreover, the scribes' awareness of the words' etymologies may result in 'original/authentic' spellings even when the words had been well-integrated and undergone phonological adaptations.

4.2 Method: Identification of distance from etymological spelling

Aware of peculiarities of the linguistic material represented by Ajami annotations, we set out the principles employed in our analysis of Arabic borrowings discernible in manuscripts. The basic procedure is to compare the spelling of a given Arabic loanword in Ajami with the original Arabic orthography of the word. The comparison is carried out following what we call a gradient of etymological spelling schematised in Table 2. The types of spelling are organised in descending order from the spelling identical to Arabic to the spelling unrelated to the Arabic original orthography.

Working with the etymological scale of graphemic representation, we pay close attention to the visual proximity of a given Ajami loanword to the original Arabic word. This is important to ascertain the extent to which the scribes are influenced in their spelling choices by the immediate graphemic prompts of the source orthography, rather than by their mental map based on the knowledge of Arabic orthography acquired during their previous cycles of learning.

---

38 Kossmann 2013, 47.
39 Van den Boogert 1997, 52.
40 Tamari and Bondarev 2013, 16.

Table 2: Types of etymological spelling of Arabic loans in Ajami: a scalar.

	Type of spelling
1	Identical: all graphemic segments retained
2	Partial
	2a All consonants retained
	2b Some consonants retained
	2c Vowels retained
	2d Some vowels changed
3	Dissimilar spelling: none of the original spelling is retained
4	Sensitivity to Arabic morphology: Arabic definite article
	4a retained
	4b absent

We also try to distinguish various characteristics of borrowing, partly drawing on Wexler's work discussed in Section 3, in our analysis of selected examples presented in Section 5.[41] The characteristics include relative chronology, channels of borrowing, the impact of written Arabic, integration into the target language, scribes' recognition of Arabic elements in their language, and semantic fields of loanwords. These characteristics are presented in Table 3.

Table 3: Characteristics of borrowing.

	Earlier vs later layers of borrowed words (including multiple cycles)
1	Direct vs intermediary source (including multiple sources)
2	Impact of scholarly and written Arabic vs non-written, colloquial Arabic
3	Non-integrated vs fully integrated etymons
4	Recognised vs unrecognised Arabic elements
5	Semantic fields related to Islamic concepts vs unrelated/non-bound to Islamic concepts

A systematic investigation of all these characteristics is beyond the scope of the present study. However, we will take them into consideration where possible in order to relate a preliminary typology of borrowing in Ajami texts to more general questions of processes of borrowing.

---

41 Wexler 1980.

5 Case studies

We now turn to the analysis of loanwords in Old Kanembu (henceforth, OKb) and Soninke Ajami manuscripts. Differences in the type of manuscripts and linguistic practices of translation represented by each Ajami tradition dictate a slightly different organisation of data. Old Kanembu glosses of the Qur'anic manuscripts retain much more archaic features not present in modern Kanuri. The glosses in Soninke are translations of non-Qur'anic texts and the similarity between Soninke glosses and modern spoken Soninke is much greater than that between OKb and Kanuri.

5.1 Old Kanembu

As already mentioned, OKb is a written language exclusively used for the translation of Arabic texts. In this paper, we deal with the earliest written attestations of OKb preserved in Qur'anic manuscripts produced in the Borno sultanate from around the mid-seventeenth century to the early nineteenth century. Old Kanembu or, more precisely, its later written variants outlived the epoch of written Qur'anic translation practiced in Borno until the beginning of the nineteenth century.[42] The non-Qur'anic texts have continued to be translated into written OKb. The written translational language has functioned as a downsized counterpart of the oral exegetical language known locally as Tarjumo, also applied in exclusively translational contexts.[43] Although it is possible to trace a continuum of incremental linguistic changes from the earliest witnesses of OKb to the modern-day Tarjumo, the term 'Old Kanembu' refers to the language of the Qur'anic manuscripts before the early nineteenth century. Both written OKb and spoken Tarjumo are closely related to the Kanuri language spoken in northeast Nigeria and south-east Niger (around and off the western shores of Lake Chad) and to the Kanembu language spoken in the areas north, north-east and east of the lake.[44] It will, therefore, be necessary to compare some of the examples of Arabic loans in OKb with their equivalents in one of the major Kanuri dialects, Yerwa Kanuri. A wide-scale comparison across available data of the other Kanuri and Kanembu dialects is outside the scope of this paper.

---

42 Bondarev 2024 and forthcoming.

43 Bondarev and Tijani 2014.

44 Kanuri and Kanembu are usually considered a continuum of dialects.

Kanuri is one of the West African languages with a large number of Arabic loans, and various aspects of a borrowed lexicon in Yerwa Kanuri have been studied in previous literature.[45] John P. Hutchison described the Kanuri lexicon as chronologically layered (a notion applicable to almost any human language, no doubt) 'in the sense that it is possible to recognize words in the language that are older as distinct from words that have not been in the language as long'.[46] As our primary focus is written attestations of Arabic borrowings, we will assess the extent to which their spelling may help to unravel such chronological layers.

The discussion of Arabic loans in OKb is organised along the scale from etymologically identical (copy-spelling) to completely divergent spelling. The scale also takes into consideration the visual relationship between the target and source etymons. The spelling scale is, thus, subdivided into the following components.

Table 4: Spelling scale of loanwords.

	Type of spelling and visual proximity to the source word	Subtype
1	Copy-spelling (of entire lexeme) above the corresponding source word	Copy-spelling, no source word present
2	Partial etymological spelling above the corresponding source word	
3	Partial etymological spelling, not written above the source word	
4	No traces of etymology, but written above the corresponding source word	
5	No traces of etymology, no source word present	

These types of spelling are exemplified below by select etymons. Each etymon is introduced as a vocabulary entry followed by explanations. Each entry is provided with a reference to a manuscript and the Qur'anic text (e.g. 1YM/91:9 reads 'from the manuscripts 1YM, chapter 91, verse 9').

---

45 Greenberg 1960; Baldi 2002; Baldi 2003; Baldi 2007; and Baldi 2020.
46 Hutchison 1981, 10.

5.1.1 Etymological spelling written above corresponding Arabic words

Most of the etymons in this category consist of the terms related to theology, religious practice and conduct, nouns designating specific entities of geographic and botanical nature unknown in the host culture, and proper names (the latter is, however, not dealt with here).

{1} <aflaḥa> 'to prosper'.[47] The verb is a direct borrowing from Arabic ʔaflaḥa 'to prosper'. In the manuscript, the word is spelled without the initial *hamza* according to the *Warsh* reading of the Qurʾan and the word is, thus, an exact orthographic copy of the original. At the same time, it functions as a typical OKb verb with its own inflectional morphology, which is characterised in the example given by the suffix <-jī> denoting the 3rd person subject morpheme and the perfective category: <aflaḥa-jī> 'he has gained reward'.

{2} <ṭūr> 'Mount Sinai',[48] written above Ar. *ṭūr* 'mount (Sinai)'. Since Mount Sinai is the place of revelation of the Torah, the passage about Sinai is commented upon by another OKb term borrowed directly from the orthography of Arabic: <tawrē> 'Torah', the pronunciation specific to the *Warsh* reading. The same OKb spelling is found written next to the original Arabic term, as in 1YM/3:50.

{3} <aṣiyām> 'fasting',[49] above Ar. *al-ṣiyām*. This is a special case of phonetic spelling of the assimilated definite article before the coronal consonants ('sun letters'). The Arabic word *ṣiyām* 'fasting' when used with the definite article *al-* is written as <al-ṣiyām> and pronounced as [aṣ-ṣiyām]. The OKb loan <aṣiyām> is written above the Arabic <al-ṣiyām>, but, nevertheless, reflects a phonetic form. At the same time, the OKb written word does not indicate the phonetic gemination [ṣṣ] in its spelling, although the Arabic spelling clearly has the gemination marked by the diacritic *shadda*.[50]

{4} <kun-ẓūlm> 'wrongdoing, injustice',[51] above Ar. *ẓulm*. This loanword entered the language at the time when there was a productive nominal derivation through the prefix *k-* (and its variant *kVn-*, where V = vowel). Such prefixed forms are among the oldest integrated loans, and, nevertheless, the Arabic source of some of these is obvious to the scribes who copy the orthography of

---

47 1YM/91:9. Note that in transliteration of OKb the letter *ɣayn* is represented as <g>.
**48** 1YM/2:93.
49 1YM/2:183, 187.
50 That OKb <aṣiyām> is borrowed from the definite article form (rather than being a phonetic adjustment by insertion of an initial prosthetic vowel [a-]) is confirmed by the OKb loan that starts with the consonant <ṣ> <ṣūm> 'to fast' borrowed from the Arabic verb *ṣūm* 'to fast'.
51 1YM/2:61, 114; 4:160.

the corresponding Arabic etymon. A similar case of scribal awareness is attested in the loanword <kaṣūm> 'fasting', which is, however, not found in visual proximity to the source word (and, therefore, discussed with category 3 below), unlike its more frequent and supposedly later loan <aṣiyām> 'fasting' discussed above.
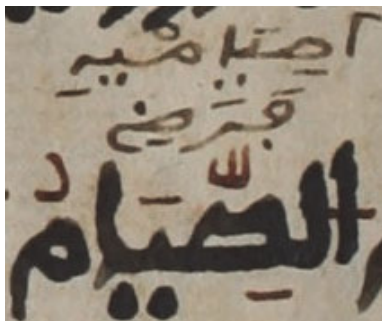
Fig. 1: The OKb loan <aṣiyām> is written above the Arabic <al-ṣiyām>.

Many words in this category start with the letter ʕayn <ʕ>, which is rarely used in non-Arabic etymons.

{5} <ʕaṣ> 'to disobey',[52] above Ar. ʕaṣā. The OKb verb form <ʕaṣ-inmi> (disobey-AUX.2s.PRF) 'you disobeyed' is written above Arabic ʕaṣay-ta (disobey-2s.m.PRF) 'you disobeyed'. The verb has been fully integrated into the Kanuri lexicon, having the phonological shape *ashi*.

{6} <ʕaðāb> 'punishment',[53] above Ar. *al-ʕaðābu*. Although the OKb form is a copy of the Arabic orthography, it does not retain the case ending -*u* necessarily present in the written Qur'an, but not always pronounced in recitation.

(1) 1YM/2:86	tandīkan	ʕaðāb-yi	itskāybō
	3p.on	punishment-SJ	lighten.3s.IMPF.NEG
	'the punishment will not be lightened for them'		

---

52 1YM/10:91.
53 1YM/2:86.

However, the etymon is spelled as <ʕaðābu>, i.e. with the final vowel -*u*, in other places of the same manuscript. This probably shows a process of phonetic adjustment to a more typical OKb syllable structure that avoids plosive consonants at the end of the syllable. This three-syllable form has been retained in Modern Kanuri: *azáwu*.

{7} <ʕinab> 'grape',[54] above Ar. *ʕinab-an* (-*an* is the suffix of the accusative case). The OKb word is borrowed from a singular Arabic form without case endings and is invariable in OKb. Thus, the same singular form is used to translate the plural *ʔaʕnāb-an* 'grapes-ACC' in 1YM/78:32. The OKb item has a final vowel after the consonant: <ʕinab-e>, probably due to the same phonetic adjustment described above.

{8} <ʕadas> 'lentil',[55] above Ar. *ʕadas*. In the other manuscripts, the same Arabic agricultural term is translated by vernacular generic terms for beans: 3ImI <gālū>, 1YM <gālgāla>.

5.1.2 Partial retention above corresponding Arabic words

These OKb words are probably well integrated, and, at the same time, the scribes are conscious about their origin, which shows in the selection of specifically Arabic graphemes for the spelling of the OKb words.

{9} <ʕājab> 'wonder'.[56] This etymon is a good example of a fully integrated word that does not always copy the source vowels, but invariably retains the etymological spelling of the consonants. The visual proximity to the source etymon does not, therefore, have a defining effect on the orthography of the vowels, as explained below.

The OKb loan <ʕājab> is based on the Arabic noun *ʕajab* 'wonder' and is used for not only the translation of this noun but also the Arabic adjective *ʕajīb* 'wonderful, amazing'. In 1YM/72:1, the Arabic phrase *innā samiʕnā qurʔānan ʕajaban* 'Indeed, we have heard an <u>amazing</u> Qur'an' is translated into OKb as <alqurʔān ʕajabū=ka krniyē> (the Qur'an amazing=DO we have heard). Although the consonants of both *qurʔān* 'the Qur'an' and *ʕajab* 'wonder' are retained in the OKb spelling, the OKb phrase is morphologically independent from the source text. Thus, the absence of the article *al-* in the source word 'Qur'an' is ignored and the word is given in its borrowed form with the article: *al-qurʔān*. The Arabic noun (in

54 1YM/80:28.
55 2ShK/2:61.
56 1YM/11:72, 50:21, 72:1.

attributive function) ʕajab-an is translated as an adjective in OKb, and has the final vowel <u> to form a permissible phonotactic structure in OKb (i.e. avoiding syllables with final plosive consonants).

The Arabic adjective ʕajīb(un) 'wonderful' (derived from the noun ʕajab 'wonder') in 1YM/11:72 is again translated by the OKb word <ʕajab(u)>, which (1) ignores the second vowel <ī> in the source word and (2) functions as a noun (in the OKb associative construction 'with') rather than an adjective.

(2) 1YM/11:72   attī   agō   ʕājab-wa

                   this   thing   wonder-with

         'this is an amazing thing' (lit. 'this is a thing with wonder')

Fig. 2: Arabic adjective ʕajīb(un) 'wonderful' is translated by the OKb word <ʕājab>.

The scribe's awareness of the Arabic source of this otherwise fully integrated lexeme becomes more obvious when compared to the genuine OKb word <agō> written next to it. If <agō> starts with an ʔalif and the vowel sign [a] fatḥa above it, the initial grapheme of the loan word is a copy of the source letter ʕayn that corresponds to the guttural voiced pharyngeal fricative sound [ʕ] in Arabic but is absent in Saharan languages including OKb. However, the rest of the OKb word is spelled deviating from the original. The most significant difference is in the vowel of the first syllable. It is a short vowel [a] in Arabic, whereas it is written as <a> + <ʔalif> in OKb, the combination representing a long [ā] in Arabic, but probably used here to mark a high tone á (cf. Kanuri cognate *ájap* 'wonder', *ájabba* 'amazing'). Similarly, the vowel of the second syllable in Arabic is spelled as <i> + <ya'>, standing for a long [ī], but it is <a> in OKb.

Other OKb words written with the letter ʕayn, such as <ʕabat> 'to worship' and <ʕalam> 'to teach', are in the same category of spelling.

{10} <ʕabat> 'to worship, serve', from Ar. ʕabada. The spelling of the final consonant in the OKb lexeme denotes a voiceless stop [t] rather than the voiced [d] of the Arabic source. The non-voiced feature shows that the verb had long been inte-

grated into the language. The OKb phonotactics observable in the manuscripts does not require a syllable-final voiced /d/ to become voiceless in any environment. Consider the inflectional forms of the verb <ʕabat> 'worship' (3) and <wud> 'pour' (4) for comparison, both having the auxiliary verb *n* suffixed to the final root consonant. In the case of <wud>, there is no change from [d] to [t].

(3) 1YM/3:51   nadīyi   tīka   ʕabat-n-ūgū

                     2p.SJ   3s.DO   worship-AUX-2p.IMV

                     'worship Him!'

(4) 1YM/80:25   andiye   wud-n-ē

                     1p.SJ   pour-AUX-1p.PRF

                     'we poured (water)'

At the same time, there are other Arabic verbs with the last consonant /d/ that changed in OKb to the root-final /t/, as in the OKb verb <wat> 'promise' from Arabic *wāʕada* 'promise'. The change of the final /d/ of the source word to the final /t/ in the target should then be due to some earlier sound change rules no longer productive at the time of the OKb of the Qur'anic manuscripts.[57]

     Conversely, the Arabic verbs with the final voiced plosive consonant /b/ retained the voiced feature of the consonant when borrowed into OKb: <jarab> from Ar. *jarraba* 'to test', <jab> from Ar. *jāwaba* 'to answer'.

     {11} <ʕalam> 'to teach' from Ar. *ʕallama*. The OKb form is invariable in its syllabic and vocalic structure (i.e. CV-CVC), irrespective of the Arabic form that has the geminated /ll/ and changes vowels according to inflection. In 1YM/2:102, for example, this OKb verb <ʕalam> is written above Arabic *yuʕallimāni* 'two (angels) were teaching':

(5) 1YM/2:102   malayka   dīro   trgsā   yamka   ʕalam-kisa-ḥalan

                  angels   two   were sent   to people   teach-ing-ADV

                  'two angels were sent to teach people'

---

57 The final voiced 'emphatic' /ḍ/ in some Arabic nouns also changed to [t] in the process of borrowing, but not systematically: thus, the OKb loan from Arabic *al-ʔarḍ* 'earth' is sometimes written as <lart>, sometimes as <larḍ> (see below item {30}).

{12} The last example in this category of loans with a recognisably Arabic-specific letter is a lexeme that starts with the latter *šīn*, rarely used outside Arabic borrowings. This is the verb <šawar> 'to consult' in 1YM/3:159, that translates the Arabic *shāwir-hum* (consult.2s.m.IMV-them) 'consult them!', as in <ni-ye tandīka shawar-nē> (you.SJ they.DO consult-2s.IMV) 'consult them!'. The same lexeme is used as a noun in 1YM/2:233, where it translates the Arabic phrase *wa tashāwurin* '(by mutual consent) and consultation' as follows: <šawari-bi gābukan> (consultation-GEN after.from) 'after consultation'.

5.1.3 Borrowing from Arabic, not written above the source word

This group of loanwords retains some etymological graphemes but is not written in proximity to the corresponding source word. The Arabic corresponding terms exist in the Qur'an elsewhere, and the same OKb lexeme is sometimes also used to translate its etymological source.

{13} <kaṣūm> 'fasting', from Arabic *ṣūm* 'to fast'. This is another k-prefixed borrowing comparable to <kun-ẓulm> 'wrongdoing' in {4} that, in spite its ancient integration into the language, was recognised by the scribes as an Arabic loan. But, unlike <kun-ẓūlm>, the etymological spelling is not prompted by the corresponding Arabic word (e.g. *ṣūm* 'to fast'): no such Arabic word is present in the verse (Q2:51) where OKb <ka-ṣūm> 'fasting' occurs. However, the event described in the verse refers to Moses's fasting on Mount Sinai and, thus, the OKb phrase explains this as follows:

(6) 1YM/2:51

būnēkami	fīdegbi	kaṣūm-nn	watkīnīyē
night.from	forty.GEN	fasting-ADV	we.appointed

'we appointed forty nights of fasting (for Moses)'

{14} <ḥaṣm>/<ḥaṣum> 'to argue' from Ar. *xaṣm* 'opponent' or *xāṣama* 'to argue'. In 1YM/4:109, the verb <ḥaṣm> translates a different Arabic lexeme *jadala* 'to argue', rather than the verb *xāṣama* 'to argue'.

(7) 1YM/4:109

ʔālagin	amū	dū	ḥaṣm-nū
God.with	persons	who	argue-3p.FUT

'who will argue (on their behalf) with God?'

In 1YM/22:19, a participle form of <ḥaṣum> is used to translate its etymological source, also in nominalised form. Thus, the Arabic *xaṣmāni* 'two adversaries' corresponds to the OKb <ḥaṣum-kita-ndi> (argue-ing-two) 'two arguing ones'. The OKb word is fully integrated in both cases, but retains the etymological connection through the use of the letter *ṣād*. Notably, the first consonant is represented by the letter *ḥāʾ* (it denotes a fricative voiceless pharyngeal sound in Arabic) rather than by the original *xāʾ* (used for the same set of articulatory features, except for the place of articulation, which is velar/uvular rather than pharyngeal). This is interesting in light of the existence of the OKb loans that copy the etymological *x*, e.g. <xalq> 'to create' from Ar. *xalq*. In modern Kanuri, the lexeme corresponding to the OKb <ḥaṣm> does not have the initial consonant: *ásəm* 'argument'.

{15} <safr> 'to travel' from Ar. *safar* 'voyage'.[58] The OKb loan verb is used to translate a different Arabic verb *ḍaraba* with the synonymous meaning 'to travel': <nadīye safr-nīyūya> (above Ar. *wa ʔiðā ḍarab-tum* 'and when you travel').
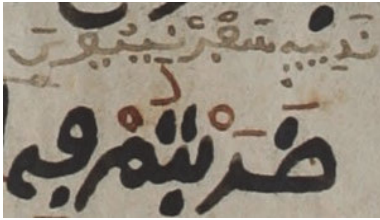

Fig. 3: The OKb loan verb <safr> is used to translate a different Arabic verb *ḍaraba*.

The same loan may function as a noun 'journey' in OKb and is used to interpret the corresponding Arabic noun. Thus, the Ar. *asfarinā* 'our journeys' in 1YM/34:19 is translated into OKb as <safr-dē> 'journey-our'. Note that the Arabic plural is not expressed in OKb.

The Arabic term *safar* 'travel' is sometimes translated by a genuine OKb word <blāgrū> (Kanuri *bəláwúró*), rather than by the loan <safr> (as in 1YM/2:184).

It is difficult to assess whether the scribes meant to copy the original orthography with items such as <safr>. This is because all consonantal letters of the original word are equally used in the genuine OKb lexemes. Therefore, such words overlap with those in the following two categories, but, at the same time, represent items in their own right because the Arabic words to which they corre-

---

58 1YM/4:101.

spond do not have specific consonants that can be used as a diagnostic of the scribal orthographic awareness.

5.1.4 Written above, but no etymological spelling

The orthography of the lexemes in this category do not have any signs of original etymology, even though they are used as translations of the original Arabic etymons and are actually written above or in close vicinity to the source words.

{16} <ka-sūgu> 'market', from Arabic -*suq* 'market', written above Arabic *ʔaswāq* 'markets'.[59] This k-prefixed word has been discussed extensively in the literature on lexical contact in West Africa as an unambiguous example of borrowing pathways from Arabic into a language of wider communication (lingua franca), such as Kanuri, and the k-prefixed word's subsequent spread into many other languages.[60] The fact that the spelling of OKb <ka-sūgu> is not sensitive to the source word *suq*, let alone its plural form *ʔaswāq*, seems to betray the scribes' oblivion of its Arabic connection.

{17} <alū> 'writing board' from Ar. *al-lawḥ*. The OKb item has almost no traces of the original word, except for the element of the definite article manifest in the initial vowel <a>. The consonant /l/ could probably have been pronounced as a geminated [ll], which would make it closer to the source word, and this is the case in Kanuri: *alló*. The example below (8) and Fig. 4 demonstrate an interplay between this well-integrated word and an orthographic copy/insertion, although the latter has the long vowel misplaced (to the third syllable instead of the second).

(8) 1YM/85:22

Q.	*lawḥ-in*	*maḥfūẓ-un*	
OKb	alū	maḥfuẓū=gin	kaθikō
	tablet	preserved=in	be.3s.AFP
	'in the preserved tablet'		

---

59 1YM/25:7, 20.

60 Greenberg 1960; Kossmann 2005; Baldi 2020.

Fig. 4: OKb <alū> 'writing board' and *maḥfūẓ* 'preserved' written above the corresponding Arabic phrase.

The simultaneous use of the non-etymological and etymological spelling of two etymons in the same short phrase suggests that the word <alū> was not perceived by the scribe as a loan from Arabic.

{18} <jagaru> 'blast',[61] from Arabic *zajra* 'shout, cry, blast' (variants: <jgaru> (2ShK), <jugaru> (Arabe 402)). The lexeme is found in the phrase <jagaru tilō> (blast one) 'one blast' written above Ar. *zajratun wāḥidatun* (blast first) 'the first blast' (of the Day of Judgement). The first consonant of the OKb <jagaru> represented by <j> probably stands for the voiced non-palatal affricate [dz],[62] which would be a close approximation to the Arabic [z] *zajra*. The second consonant <g> points to a source other than Classical Arabic, to a dialect where the Classical Arabic /j/ corresponds to /g/. But this alone does not explain why a colloquial Arabic source would be considered for a very specific word used in the context of the Day of Judgement.[63]

{19} <ʔala> from Ar. *allāh* 'God'. The etymon has the most transparent connection to the original Arabic word, and it is remarkable that the OKb employs this very specific spelling convention for the most significant religious concept God. However, this non-etymological spelling only occurs in the first part of 1YM, about 65% of the entire 470-folio manuscript (see item {30} for more details). The remaining 35%, and indeed many other manuscripts, exhibit the etymological spelling <allāh>.[64]

---

61 1YM/79:13.

62 Bondarev 2014b.

63 Lameen Souag confirms the doubts about a colloquial source (Souag's comments on an earlier draft of the paper).

64 The manuscripts that have the etymological spelling <allāh> or, much more often, non-vocalised spelling, are T.Kano, 3ImI, 4MM, Kad.Ar.33, 5.Konduga.

5.1.5 Borrowing from Arabic, but not associated with the source language

The Arabic words in this group of borrowings are either not used in the Qur'an or used only once or twice.

{20} <algāma> 'wheat', from Ar. *al-qamḥ* (Kanuri *algáma*, *ləgáma*, *laáma*). The Arabic word *qamh* 'wheat' does not occur in the Qur'an, and the OKb term is written above the Qur'anic Arabic *fūm* (Q.2:61), which may mean either 'garlic' or 'wheat'.[65] The OKb scribes interpreted the Arabic word *fūm* as 'wheat' following certain exegetical sources.[66] This is also corroborated by the very choice of the OKb word, or rather, by the absence of the OKb term for 'garlic', which would resemble the modern Kanuri *kəngálmú* 'garlic' (unattested in OKb).

Due to its significance for the reconstruction of the scribes' perception of the integrated loanwords, the whole sentence translating various agricultural terms in this Qur'anic verse is illustrated below in (9). The two Arabic loans <algāma> 'wheat' and <albāsar> 'onion' are integrated into the list of the vernacular botanical terms, and it is obvious that both terms are not perceived as loans. Thus, <algāma> 'wheat' does not have its original equivalent in the verse and elsewhere in the Qur'an, whereas <albāsar> 'onion' is written above the corresponding Arabic term in the spelling, entirely disconnected from the original word (*baṣal*).

(9) 1YM/2:61

Q.	*min baqli-hā*	*wa qabaθāʔi-hā*	*wa fūmi-hā*	*wa ʕadasi-hā*	*wa baṣali-hā*
OKb	**kālō klī**-jī kāmi	**fālī**-jī kāmi	**algāma**-jī kāmi	**gālgāla**-jī kāmi	**albāsar**-ji=kami
	leaf green-its from	watermelon-its from	wheat-its from	beans-its from	onion-its from

'from its (earth's) herbs, its cucumbers, its wheat, its beans, its onions'

---

65 Badawi and Haleem 2008, 721.
66 Ibn Kathīr (d. 1373) referred to this ambiguity in his *tafsīr*: 'Others said that *fūm* is wheat, the kind used for bread. Al-Bukhari said, "Some of them said that *fūm* includes all grains or seeds that are eaten"' (https://quran.com/2:61/tafsirs/en-tafisr-ibn-kathir; accessed on 12 January 2024). Similar explanations in Arabic are found in various OKb manuscripts.

{21} <kn-jikl> 'hardship', from Ar. *shikāl* (pl. *shukul*) 'fetter(s) for shackling the feet of riding animal' or *shakl* 'state of confusion'.[67] The k-prefixed word does not exist in modern Kanuri, and the cognate Kanuri term *shíwol* 'distress, trouble' must have entered the language at a later stage and possibly from a source different to <knjikl>.[68]

{22} <fahama> 'hear', from Ar. *fahima* 'understand'. Some obvious loans that retain Arabic orthography seem so well integrated into the language that their spelling peculiarities should be considered part of the core graphemic inventory of OKb. Such is the word <fahama>, with the letter *hāʾ* normally used in borrowings and as a word-spacing device.[69] The lexeme is used as a verb 'to hear' in the OKb manuscripts. It is unlikely that the OKb term was borrowed from the Qur'an since there is only one occurrence of the Arabic *fahima* in the Qur'an, where it occurs in the causative derivation *fahhama* 'to make understand, explain' (Q.21:79). What the Okb term <fahama> consistently translates is the Arabic *samiʕa* 'to hear' and, thus, the loan word has its own path of semantic derivation from 'understand' to 'hear', making it completely independent from the source language (however present the latter may be in such a translational relationship). The only occurrence of the Arabic *fahima* (a causative *fahhama*) is translated into OKb by another OKb verb *fam*, <fam/fan> also with the meaning 'to hear' (<itfamgiye> 'we made him understand').[70]

5.1.6 Definite article retained

Many Arabic words came into OKb and Kanuri-Kanembu with the Arabic definite article *al-*. The article in such loans may be written as either <al-> or only

---

67 1YM/90:4. As pointed out by Lameen Souag in his comments on an earlier draft of the paper, the Arabic *shakl* as a borrowing source raises the problem of the difference in vowel, /a/ in Arabic vs /i/ in OKb. It is also possible that the OKb lexeme is internal Saharan rather than a loan from Arabic. Thus, in Tubu, the verb *kakal* 'belästigen' has been attested by Lukas (1957, 185), with the basic root *kal*. This would not, however, explain the affricate <j> in the OKb <kn-j-kl>, and the matter, therefore, is left for future research.

68 Apart from the prefix *k-*, the major difference here is in the vowel quality in the second syllable of *shíwol*, rather than in the quality of the second consonant, because the historical sound change k > w in the intervocalic position is common in Kanuri.

69 Bondarev and Dobronravin 2019, 249.

70 It is unclear whether the OKb <fam/fan> (Kanuri *fan*) is a different borrowing of the same Arabic *fahima* or an internal Saharan root with such cognate forms as **b**as in Teda-Daza and Bakoore Beria *kɛ**b**ɛnô* given regular correspondences between Kanuri-Kanembu f- and Teda-Daza and Beria b- (Chonai 1999, 21, 183, 217–218).

retaining the second segment – the consonant /l/, or as reflecting a form assimilated to the following coronal consonants/graphemic 'sun letters' (see etymon {3} <aṣiyām> 'fasting'). We only consider the two first types of the retained article, <al-> and <l->, in this subsection.

The borrowings with the definite Arabic article are listed here in a way similar to the previous examples, from the closest etymological spelling to the items spelled independently of Arabic source words. Given that the Arabic definite article is an easily identifiable grammatical item with a very high occurrence frequency, it is expected that the scribes who constantly deal with the Arabic and OKb content would naturally analyse the OKb words containing the initial element *al-* as Arabic loans. Some of the article-based OKb loans have other distinctive Arabic graphemic elements and, thus, plausibly represent the conscious Arabic-related repertoire of scribal choices. However, the orthography of some OKb words with the retained definite article is unrelated to the original Arabic words, making such loans more difficult to classify from the point of view of scribes' etymological awareness. This discrepancy between the clear presence of the Arabic definite article and non-transparent spelling of the rest of the word also complicates categorisation of the loanwords that retain the Arabic article in a clipped form, with the initial vowel lost, such as <laqabur> 'grave' from Arabic *al-qabr*. If the word 'grave' is spelled with the Arabic letter <q>, irrespective of the changed (phonetic and graphemic) shape of the article *al-*, why does a more obvious article-based loan <al-bāsar> 'onion' from Arabic *al-baṣal* not reflect the original *ṣad* <ṣ> and is spelled instead with the letter *sīn* <s>? A plausible answer to this question could only be obtained through the analysis of a much larger set of data than that selected for this study.

5.1.6.1 Words with the Arabic article al- retained in full

{23} <alaxira> 'afterlife; the hereafter',[71] from Ar. *al-ʔaxira* 'the afterlife' and written above the Arabic word used in genitive case (*al-ʔaxira(ti)* 'of afterlife'). This genitive grammatical relationship is translated into OKb with the respective genitive marker *-be* (<alaxira-be> 'of afterlife'). The only deviation from the original spelling is the disregard of the *tāʔ marbūṭah* ending *-ti* in the Arabic *al-ʔaxira-ti*, but this is typical in almost all borrowings in OKb or Kanuri-Kanembu.

{24} <alqurʔān> 'the Qur'an', from Arabic *al-qurʔān*. This OKb word denoting the sacred scripture always occurs with the definite article integrated into the lexeme, irrespective of the grammatical status of the original word. Thus,

---

71 1YM/79:25.

the OKb <al-qurʔān> in 1YM/72:1 is written above the Arabic word without the article (*qurʔān*). The form with the whole definite article is, however, less frequent in this manuscript than the one with the clipped initial vowel (see <luqurān> and its variants in the next list). However, the form with the whole article is preferred in the manuscript T.Kano: <alqurān> (Q.46:2). At the same time, the T.Kano variant lacks the original *hamza* <ʔ> in its spelling.

{25} <albāṣal> 'onion'. Ultimately from Arabic *al-baṣal*, via an intermediary language. The same etymon in Kanuri has the article integrated into the clipped form: *luwásar*.

{26} <algāma> (repeated {19}) 'wheat'. Ultimately from Ar. *al-qamḥ*, but borrowed via an intermediary source. The etymon in Kanuri occurs in three (idiolect/dialect-specific) forms: *algáma, ləgáma, laáma*.[72]

5.1.6.2 Words with the definite article without the initial vowel a, i.e. <l->

{27} <luqurān>/<luqrān> 'the Qur'an'.[73] Unlike the form with the unchanged definite article *al-* (etymon {24}), this is fully integrated into the language and the only sign of the original etymology is the letter <q> rather than the Okb standard <k> used for the voiceless velar stop /k/. It is this clipped form that exists in modern Kanuri, with the velar being elided between vowels: *luwurân*.

{28} <laqabur>/<laqabr> 'grave',[74] from Ar. *al-qabr* (pl. *al-qubūr*). The form with the clipped article is not known in Kanuri, where the cognate (borrowed)

---

72 See Wexler 1980, 536, on the problem of borrowing paths of both etymons (onion and wheat). Baldi (2008, 244–245) provides an extensive list of the etymon 'onion' across forty different languages, the majority of which have either the whole or clipped article in the first syllable. The etymon 'wheat' with the integrated Arabic article is attested in twenty languages, see Baldi 2008, 415–416. The form with the voiced [g] after the article *al-* is only reported in Kanuri and Munjuk (a Bio-Mandara [Central Chadic] language spoken in Cameroon and Chad). Lameen Souag's suggestion that the voiced [g] (g < q) 'would readily reflect Libyan or Sudanese Arabic' may well explain the voiced feature of the velar in this OKb/Kanuri borrowing (Souag's comments on an earlier draft of the paper).

73 1YM/17:105 and 1YM/2:176 respectively.

74 <laqabur> in 1YM/9:84, 82:4; <laqabr> in 1YM/35:22, 100:9. In the case of <laqabur>, the insertion of a vowel into the consonant cluster (here, -*br*- in the source word al-qabr) to form a permissible syllable structure is a typical mechanism in Kanuri. Exemplarily, K. *laadar* (*lagadar*) 'capacity' < Ar. *al-qadr* 'extent'; K. *lámar* 'event, situation' < Ar. *al-ʔamr* 'matter, affair'. The quality of the inserted vowel /u/ in <laqabur> is conditioned by the preceding bilabial /b/. Lameen Souag suggested an alternative source from a Maghribi colloquial form *l-əqbəṛ* (Souag's comments on an earlier draft of the paper). However, since the common pattern of vowel insertion plausibly justifies Classical Arabic as a direct source of borrowing, a distant link to Maghribi sources seems unnecessary.

word is *káwar*. However, the OKb variant has been preserved in Tarjumo: *lawura*.

{29} <lqalam>, variant <lqālm> 'pen',[75] from Ar. *al-qalam*. All three known instances of the word in 1YM exhibit a peculiar combination of the copy-spelling and phonetic encoding. The letter <q> is a reference to the source word, but the integration of the article in the clipped form rather than the whole <al-> suggests an adapted pronunciation [ləkalam] or, possibly, [ləkálam] if the long vowel in the variant <lqālm> encodes a high tone. This is plausible, since the Kanuri cognate also has a high tone on the second syllable: *alkálam*. However, unlike OKb, the Kanuri form (and its numerous variants *alkərəm, alláram, ar'álam* and *arkərəm*) has the whole article integrated into the word.

{30} <larḍ>/<lart> 'earth', [76] from Ar. *al-ʔarḍ* 'the earth'. The word *al-ʔarḍ* occurs 461 times in the Qur'an, and only with the article, which is always preceded by a vowel. This results in the pronunciation [l-arḍ]. The prominence of this phonetic form in the Qur'an suggests that the OKb (and Kanuri/Kanembu) borrowed the term directly from the Qur'an. Note, however, two distinct types of spelling: one with the copy of the original letter <ḍ> for the emphatic voiced alveolar stop [ḍ] and the other with the letter <t> used for the voiceless alveolar [t]. The former spelling convention <larḍ> is consistent in the first part of the manuscript 1YM, whereas the latter <lart> corresponds to the second part. The watershed between the two spelling types runs in Q.34. In Q.34:1, the etymon is written as <larḍ> and in the next verse, Q.34:2 (and in different hand!), the same word is written as <lart>. It is hard to ascertain the significance of this orthographic difference. On the one hand, being an obvious deviation from the Arabic original, the variant <lart> seems to suggest an adapted pronunciation independent from the source word, something like [lartə]. This is plausible given that the modern Kanuri word is *lárdə* 'earth' and that historically, the voiceless alveolar stop /t/ became voiced between sonorant/vowel and vowel (e.g. *atə > adə* 'this'). On the other hand, the same second part of the manuscript consistently copies the orthography of the etymon 'God' <allah> (الله). This co-occurrence is, however, reversed in the first part of the manuscript. There, the etymon 'earth' is copy-spelled as <larḍ>, but 'God' is written as <ʔāla> (ءال), probably reflecting pronunciation (cf. Kanuri *Ála*, with the first vowel having a high tone). The two

---

75 <lqalam> in 1YM/68:1, 96:4; <lqālm> in 1YM/31:24.

76 <larḍ> is found in the first part of 1YM (e.g. Q.2:30, 36, 116, 117, 164, 3:50, 4:97, 18:14, 27:25), whereas <lart> occurs in the second part of 1YM (e.g. 50:7, 44, 51:20, 57:4, 5, 10, 79:30, 81:2 84:3, 86:12, 88:20, 89:21, 91:6). Other spelling variants are rare, e.g. <lārd> in 1YM/9:2 and <lard> in 1YM/78:6. In 2ShK/91:14, the word is also written as <lart>.

parts of the manuscript 1YM are different in many other ways as well,[77] and it is plausible that they were produced by different students taught within different schools of Qur'anic interpretation, each of which had specific preferences for the choice of phonetic- and etymology-based spelling.

{31} <k-l-īslām> 'Islam'. Unlike the other k-prefixed loan nouns, the source for this word was the Arabic lexeme with the definite article *al-* (*al-ʔislām*). The orthography of the OKb <klīslām> is consistent and invariable, irrespective of the form and proximity of the source word. Thus, in 1YM/3:19, it is written above *al-ʔislām*, in 1YM/2:208, above *al-silm* 'submission to God, peace' and in 1YM/2:257, it is used to specify the Arabic word *nūr* 'light' in the sense of the 'light of Islam': <klīslām-bi nūr>. The OKb k-prefixed form based on the source word with the article *(a)l-* did not find its way into Kanuri, where the concept of Islam is expressed by a prefixed form *kər-məsələm* 'Islam', which derives from the Arabic word without the article *muslim* 'Muslim'.

5.2   Soninke

Soninke was the language of the Ghana empire, which was among the first sub-Saharan polities to adopt Islam.[78] Despite the long history of contact with Islam and, hence, the Arabic language, there was relatively scarce research on Arabic loans in Soninke. Although examples of Arabic borrowings into Soninke have been documented in lexicographical works[79] and quoted in comparative surveys,[80] the only work (we are currently aware of) addressing this topic in detail is that of Seydina-Ousmane Diagana.[81]

The present section examines how Arabic borrowings are represented in Soninke Ajami writings.[82] The materials discussed here appear in manuscripts mainly from the early nineteenth century. These manuscripts were produced by speakers of Mandinka,[83] who resorted to a related language, Soninke, as a means of exegesis.[84] Theological treatises and Islamic law manuals are among

---

77 Bondarev forthcoming.

78 Cuoq 1984.

79 Dantioko 2003; Baldi 2008; Ousmane Moussa Diagana 2013; Bathily 2017.

80 Souag 2016.

81 Seydina-Ousmane Diagana 1992.

82 On spelling of Arabic borrowings in Soninke Ajami, see also Dramé 2021.

83 The scribes used their native Mandinka language to write colophons and some glosses, which they marked with the phrase *fī kalāminā* 'in our language'. See Ogorodnikova 2017 and 2023.

84 See Ousmane Moussa Diagana 1995, 19; Tamari and Bondarev 2013, 18; Tamari 2016, 45.

the texts annotated by the scribes. Because of the texts' specific nature, the scribes extensively incorporated words of Arabic origin.

The examples of borrowings in this section are divided into three groups, depending on their spellings: (1) etymological, (2) partially etymological and (3) non-etymological. The spelling of borrowed words does not seem to depend on their proximity to the source word.

The spellings of borrowings are compared to their assumed source, on the one hand, and to the modern-day Soninke form, on the other. It is frequent across the manuscripts under examination that the same word is spelled differently by either the same or different scribes. As a result, the spellings range from less to more distant from the assumed etymon. Although spelling variants and their frequency are considered when categorising examples into a particular group, the statistical analysis of these variations and occurrences of each item will be carried out in the future.

5.2.1 Etymological spelling

This group includes borrowings that retain the orthographies of the source in the Ajami glosses, even though many of these words in modern-day language have undergone phonological changes. It comprises concepts and words central to Islam, such as God and the holy scripture, words from the religious exegesis and practices.

{32} <alqurʔān> *al-qurʔān* 'the Qur'an'. The term for the holy scripture of Islam is one example of complete etymological spelling throughout the manuscript corpus, with consonantal base and vocalisation rendered exactly the same as in the source. The lexeme is attested in two forms in modern-day spoken language – with and without the definite article – *àlixùràané* and *qùráanà/qùràané*.

{33} <allāh> *Állà* 'God'. The word 'God' in Soninke Ajami glosses is often written in proper Arabic spelling. However, spellings reflecting the phonetically adjusted form *Állà* are also frequent: <alā> or <ala> (the medial consonant gemination being underrepresented). The latter graphical representation with *ʔalif-lām* is usually attached to the following segment, similar to that in <alataʕalā>[85] *Állà tāʕalā* 'God the Exalted' or <alamakiri>[86] *Állà mà gìrí* 'God is eternal' (lit. 'God did not come to an end'). Some spellings differ from the source entirely,

---

85 EAP 1042/9/2 p. 22.
86 BULAC MS.ARA.219bis fol. 2ʳ.

replacing the initial *ʔalif* with *ʕayn*, followed by *lām*, i.e. <ʕala>[87], or *lām-ʔalif maqsurah*, i.e. <ʕalā>[88] (corresponding to the spelling of the Arabic preposition *ʕalā* 'on, upon').

5.2.2   Partial etymological spelling

Several etymologically spelled borrowings exemplified below reproduce the consonantal base (including long vowels) but may differ in vocalic notation.

{34} <ṣāḥibi> *saahibe* 'companion' from Ar. *ṣāḥib* (pl. *aṣḥāb*, *ṣaḥb*) 'associate, companion'. The Soninke plural suffix *-nu* is often attached to the core form, spelled etymologically, i.e. <fāri ṣāḥibinu> *fàaré saahibe-nu* 'prophet's companions'. Word formation with indigenous suffixes suggests borrowing rather than insertion or copy-spelling. This word is not listed in dictionaries, which suggests that its usage is specific to exegetical contexts.

{35} <ṭālibi> *táalíbè* 'student' from Ar. *ṭālib* (pl. *ṭullāb*) 'student'. The main phonological difference between the source and the borrowing is the emphatic /ṭ/ transformed into an alveolar voiceless plosive /t/. Most Ajami spellings retain the initial consonant of the Arabic original, but sporadic spellings with *tāʔ* for the adapted form also occur. Similar to the previous example, the word produces derivatives with Soninke suffixes (e.g. <ṭālibinu> *táalíbè-nú* 'students'), indicating its integration into the Soninke lexicon. The term refers to learners at intermediate and advanced stages of classical Islamic education and has several synonyms, such as *qàrànlénmè* 'pupil, student' (a compound formed with an Arabic borrowing; see below).

The next group of examples contain the sounds /ħ/ (fricative voiceless pharyngeal) transliterated as <ḥ> and the glottal fricative /h/. While the former is outside of the Soninke phonemic inventory, the latter exists in eastern dialects, regularly corresponding to *f* in western dialects.[89] The /ħ/ in the Arabic borrowings is usually adjusted to /h/ (without being changed to the dialectal *f*) in the spoken domain.[90] In the written domain, however, the tendency is to retain the graphemes of the original: <ḥaramu> *hàráamè* 'forbidden' from Ar. *harām* 'forbidden, unlawful', <ḥaqi> *háqqè* 'right' from Ar. *ḥaqq* 'right', also

---

87 BL Or. 6473 fol. 206ᵛ; UbL Or. 14.052(5) fol. 1ʳ.

88 BL Or. 6473 fol. 221ʳ; PGL ORI 11/2 fol. 12ʳ⁻ᵛ.

89 Creissels 2016, 13. Ousmane Moussa Diagana (1995, 19) notes that due to intellectual migration, the language of religious exegesis is characterised by systematic usage of *f* (and not *h*).

90 See Seydina-Ousmane Diagana 1992, 207; Ousmane Moussa Diagana 2013, 77–78.

with attached Soninke suffixes:[91] <ḥijāna> *híijáanà* 'pilgrim' from Ar. *ḥajj/ḥijjah* 'pilgrimage' + *-áanà* (NMAG). At the same time, the glottal fricative /h/ is represented by the letter *hā*: <jāhilāku> *jàahìláaxù* 'ignorance' from Ar. *jāhil* 'ignorant' + *-áaxù* (ABSTR), <ẓāhiriⁿkīti> *jàahìrìnkíitè* 'transparent judgement'[92] from Ar. *ẓāhir* 'visible, clear' + Soninke *kíitè* 'judgement', and <halaki> *hàlákì/fàlákì* 'ruin, destroy' from Ar. *halaka* 'to perish, to be destructed'. However, these graphemes are sometimes used interchangeably. Yet, there is one word that appears consistently across the corpus in the source spelling with *hāʾ* even though in the phonetically adjusted form it is dropped:

{36} <fahamu> *fàámù* 'understand' from Ar. *fahima/fahm* 'understand/understanding'.[93] While the modern Soninke form does not have the medial /h/,[94] the corresponding Arabic grapheme *hāʾ* is systematically retained in writing. At the same time, the verb behaves as fully integrated, taking the Soninke grammatical items, such as aspectual suffixes, for example, the imperfective gerundive suffix *-nV* in the example below:[95]

(10) Gloss to [*kull mā*] *yafhamu al-jāhil* [*minhi*] '[everything that] the ignorant understands [from it (the Qurʾan)]': <jāhilinun kwʾ fahamunu>

jàahìli-nú-n	*gà*	*wá*	*fàámù-nú*
ignorant-PL-D	SBD	INACP	understand-GER

'the ignorant (ones) understand'

The following examples demonstrate approaches to spelling Arabic words with voiced alveolar /z/ (encoded with the grapheme *zāy*) and voiced interdental ð (encoded with *ðāl*), both of which correspond to /j/ in Soninke borrowings.

{37} <lazimi> *làajímí* 'become obligatory' from Ar. *lazima* 'to be necessary'. This is another word of the exegetical domain in which the original consonant (*zāy*) of the source word is retained in the Ajami spellings. Although in a few cases, it is

---

91 Morphological categories are marked based on Creissels 2016.

92 See this example in Dramé 2021, 195, and also MAAO AF 14722(87) fol. 154ᵛ.

93 Seydina-Ousmane Diagana (1992, 128) quotes *fahama* as the basis of the Soninke borrowing. For the cognate Bamana verb *fàamu*, Zappa (2011, 237) suggests either *fahm* (*maṣdar*) 'understanding' or *ifham* (2sg m imperative) 'understand!'.

94 Creissels (2016, 25) argues that the vowel sequence *à-á* of the word *hàámù/fàámù* resulted from the dropping of the *h* in the intervocalic position.

95 MAAO AF 14722(87) fol. 190ᵛ.

also spelled with *jīm*, thus, betraying its oral counterpart pronounced as /j/. Despite the etymological spelling, <lazima> behaves as a genuine Soninke verb:

(11) Gloss to *bi-mā yalzamu* 'with what is necessary, required': <ʾa ka wu sirī lazimini>

a	*gà*	*wó*	*sèré*	*làajímí-ní*
3S	SBD	INACP	person	become.obligatory-GER
				'that [what] is obligatory to a person'

It also produces various derivatives, such as *làajímí-yé* 'obligation' ('become obligatory' + NMLZ) and *làajímí-nd-áanà* ('become obligatory' + ANTP + NMAG) 'necessary, requiring'.[96] The core-element <lazimi> tends to be spelled etymologically in all these forms.

   The retention of etymological spelling in the verb *làajímí* contrasts with the spellings of *jíidì* 'to increase' from Ar. *zāda* or *zid*.[97] The latter is much more frequently spelled with *jīm* or *ðāl*, rather than the original *zāy*.

   {38} <ðunubiⁿ> *jùnúbù* 'sin' from Ar. (s. *ðanb*) *ðunūb* 'sins'. The target spellings retain the initial consonant *ðāl* of the source, even though it is pronounced as an affricate /j/ (for which the grapheme *jīm* is a better match). The vocalization pattern of the Soninke item suggests that it derives from the Arabic plural form *ðunūb*. Contrary to the attested modern form (for singular and plural) *jùnúb-ù*,[98] the spellings in the glosses mark the last vowel with *kasra* <ðunubiⁿ>,[99] giving /junub-i/ or /junub-e/. Thus, the scribes possibly reinterpreted the borrowed *junub-u* as the Soninke plural and, using the *-e* ending, derived its singular form.[100]

---

96 The gloss *làajímíndáanà* corresponds to the main text's (active participle) *lāzim* 'necessary, requisite'.

97 Zappa (2011, 237) suggests the imperative *zid* of the verb *zāda* as the etymon of the Bamana *jíidi* 'increase'.

98 The Soninke plural is formed by the suffixes *-o*, *-u* and *-nu*. The singular word forms ending with *-e* regularly form the plural by changing the final vowel to *-u*, while singular forms ending with *-u* form the plural by attaching the suffix *-nu*. The Soninke words with the singular marker *-u*, including Arabic borrowings, have the lowest rate of occurrence. See Ousmane Moussa Diagana 1995, 54–58. Thus, *jùnúb-ù* seems to have an exceptional and irregular number morphology.

99 AAN1 p. 100; BL Or. 4897 fol. 168ʳᵛ; JRL MS 780[825] fol. 37ʳ.

100 However, the glosses are sometimes ambiguous as to whether the form <ðunubiⁿ> is to be interpreted as singular or plural. Additionally, I could not find out whether there are forms such as /junubu/ or /jubuni-nu/ for the plural.

{39} <ḥarafu> *harafe* 'letter' from Ar. *ḥarf* 'letter, consonant'. An opposite case is the word *harafe* 'letter', in the modern language attested with the final -*e* but spelled in manuscripts with the final *ḍamma* for /u/, i.e. <ḥaraf-u>. This corresponds to the Arabic source form in the nominative case *ḥarf-un*. From the examples below in the glosses, it follows that <ḥarafu> is used in singular (12), and forms the plural by attaching the suffix -*nu* (13):[101]

(12) Gloss to [*mā xaṭṭa bi yadihi*] *ḥarfun* '[he (the Prophet) did not write with his hand] a letter': <ḥarafu bāni ya>

haraf-u	*báané*	*yá*	*ní*
letter-SG	one	FOC	COPEQ
	'a single letter'		

(13) Gloss to *bi-al-ḥurūf* 'with the letters': <tiḥurufūnu>

tí	*harafu-nu*
with	letter-PL
	'with letters'

It is plausible that the item *haraf-u* was at some point reinterpreted as a plural, producing the modern-day singular *haraf-e*.

{40} <xibārindiyi> *qìbáarìndíyè* 'information' from Ar. *xabar* pl. *ʔaxbār* 'news'. The Soninke *qìbáarì-ndí-yé* (*qìbáarí-ndì* 'to inform' + -*yé* NMLZ) 'information' seems to be another borrowing which derived from the broken Arabic plural noun. The Soninke *qìbáarì/qìbáarè*, with the long vowel /ā/ in the second syllable, corresponds to the Arabic plural *ʔaxbār* with the initial vowel omitted (i.e. *xbār*). Interestingly, in one instance, the scribe retained the entire Arabic plural form (i.e. retaining the initial vowel and adjusting vocalic notation) in the Soninke borrowing <axbāriⁿ di yin dū> /axbaarindiyen du/ *qìbáarìndíyèn dí* 'in the news, information' when translating the respective Arabic phrase *fī al-ʔaxbār*.[102] Given the close proximity of the source word, copy-spelling of the form /axbaarindiyen/ cannot be ruled out. However, it is then unclear as to why the scribe omitted the definite article *al-* and added the Soninke suffixes.

---

101 Examples 12–13 are from MAAO AF 14.722(87) fol. 208ʳ and fol. 199ʳ.
102 MAAO AF 14722(87) fol. 105ʳ.

In this and other instances, the scribes retain the etymological grapheme *xāʾ* in the spelling of *qìbáarè/qìbáarìndíyè* (without the initial vowel). Although the Arabic grapheme *xāʾ* is a good match for the Soninke uvular /q/ or its allophone [χ], it is not a common convention to use it in writing indigenous lexemes in the Ajami material under examination, and the grapheme *qāf* is typically used. Another example of retaining the source orthography <xayraⁿ>[103] with the grapheme *xāʾ* is *qéerì* 'joy, fortune' from Ar. *xayr* 'prosperity'.

The graphemic representations of borrowings in the following group are somewhat 'hybrid', since they reflect phonetically adjusted forms while retaining traces of the source orthographies, especially conspicuous with Arabic-specific graphemes.

{41} <qara>/<karā> *qàrá* 'study, read' from Arabic *qaraʔa* 'read, recite'. Some of the scribal spellings replicate the orthography of the etymon, retaining the final *ʔalif* (although without *hamza*). The final *ʔalif* may appear even when the first consonant is not spelled etymologically, also in composites: <karāⁿmūdi>[104] *qàrànmóodì* 'teacher'. The usage of the long vowel in the second syllable of <qarā>/<karā> could be explained by the scribal intention to mark the high tone. However, the example of the compound word <karāⁿmūdi> shows that a long vowel is marked even when it bears a low tone, and it is probably a reference to the source orthography.

{42} <ṣali> *sállì/-è* 'pray/prayer' from Ar. *al-salāh/ṣalla* 'prayer/pray'. This word could possibly count among the oldest borrowings due to the antiquity of Islam among the Soninke. The scribal spellings often conform to the original orthography by using the emphatic *ṣād*, even in compounds, such as <ṣali yi>[105] *sállí-jí* 'ablutions (lit. prayer-water)' or <ṣalifana> *sállìfàná* 'afternoon prayer'.[106] However, as indicated earlier, the etymological and adapted spellings coexist, and one finds spellings with *ṣād* or *sīn* even on the pages of the same manuscript.[107]

---

103 BULAC MS.ARA.219bis fol. 9ᵛ; TCD MS 3499 fol. 47ᵛ.

104 BL Or. 4897 fol. 154ᵛ.

105 UbL Or. 14.052(8) fol. 21ʳ⁻ᵛ. The grapheme *yāʾ* in this spelling apparently represents a palatal glide /y/ rather than rather than the affricate /j/ expected in the Soninke word *jî* 'water'. This may evidence a specific Soninke dialect or an influence of a scribal native language.

106 EAP 1042/9/2 p. 3. The Soninke gloss *sállìfàná* corresponds to the Arabic *ẓuhr* 'midday prayer' in the main text. As Lameen Souag explains, the Soninke term *sállìfàná* 'lit. prayer-first' is in fact a calque of Arabic or Berber term *ṣalat al-ʕūlā* 'the first (among the obligatory) prayer(s)', with it another appellation for *ẓuhr*. See Souag 2015, 361–362.

107 BnF Arabe 5675 fol. 71ᵛ, fol. 74ʳ, fol. 75ʳ; TCD MS 2179 fol. 20ᵛ, fol. 21ᵛ.

{43} <laṣili>/<lasili> *lásìlì* 'origin' from Ar. *al-ʔaṣl* 'origin'. The word retains two etymological features: the definite article (even though with the clipped initial vowel) and the emphatic <ṣ>, that fairly often appears in the spellings of the otherwise adapted CV-CV-CV form. Spellings with *sīn*, indicating a phonetic adjustment to /s/, are also found in the glosses.

{44} <laʕada> *láadà* 'custom, tradition' from Ar. *al-ʕādah* 'custom'. Similar to the previous example, the word was borrowed with the (clipped) definite article. The manuscripts display remarkably consistent spelling of this item with *lām-ʕayn-dāl* (لعد) vocalized with *fatḥa*s.

{45} <sariʕa>/<šariʕā> *sàríyà* 'law' from Arabic *šarʕ* 'Revelation, divine law' or *šarʕīah* 'canonical Islamic law'. Interestingly, the phonetical form of the Soninke borrowing *sàríyà* corresponds more to the Arabic *šarʕīah*, whereas its graphical representation reproduces the Arabic *šarʕ* almost identically. The variants with *sīn* at the beginning of the word likely signal the phonetically adapted form, while those with *šīn* evoke the original etymon. As for the retention of the final *ʕayn*, it does not seem to have any phonological motivation since it usually represents the velar nasal /ŋ/ or serves as a support for a vowel, which would then give /sariŋa/ or /saria/. Thus, the retention of the *ʕayn* seems to be a purely visual reminder of the source, and, in this case, is to be interpreted as /y/ for /sariya/.

{46} <niʕima> *néemà* 'peace, grace' from Ar. *niʕmah* 'benefit, blessing'. The etymological spellings with the medial <ʕ> appears alongside an adjusted orthography of the word with *yāʔ* and *kasra* for /ī/: <nīma>. Therefore, the *ʕayn* is conceivably a way of encoding the long vowel, while, at the same time, referring to the word's source orthography.

Similar to examples 42–45, the feminine suffix *-ah* encoded by *tāʔ marbūṭah* is omitted in the spelling, and the last syllable is vocalised with *fatḥa* (for /a/).[108] By contrast, the word *umatunu* 'peoples'[109] from Ar. *ummah* (pronounced [*ummatun*]) 'community', is a borrowing derived from the literary Arabic form (with /t/ at the end) followed by the Soninke plural suffix *-nu*. The phonetically integrated *tāʔ marbūṭah* ending resulted in spelling the last syllable with *tāʔ* <t>: <ʔumatunu> or <ʔumaⁿtūnu>.[110]

---

108 The word *sóora* 'sura' from Ar. *sūrah* may occasionally be spelled with *tāʔ marbūṭah* <ẗ>, i.e. <sūraẗ> (e.g. UbL Or. 14.052(8) fol. 72ᵛ). However, spelling <sūra> without *tāʔ marbūṭah* also occur (e.g. BL Or. 4897 fol. 155ʳ).

109 The word *unmatoonuu* (pl.) 'prophet disciples' is listed in Mody Bathily's dictionary as a borrowing from Arabic *ummah*, see Bathily 2017, 362. A cognate word *mànton* 'people, nation, group' is attested in Bamana.

110 BL Or. 6473 fol. 214ʳ; TCD MS 2179 fol. 13ʳ.

5.2.3   No etymological spelling

Several words of Arabic origin display spellings quite dissimilar to the source orthography. Some of these words possibly came via an intermediary language, and the Arabic etymology of some is even disputable.

{47} <sʔumu>/<sumu> *súumè/súumù* 'fast' from Ar. *ṣaum* 'abstention, fasting'.[111] The adapted spelling of the word appears twice in the proximity of the source word.

In the first case (14), it explains *taṣūmu* 'she is to fast'.

(14) <ā nā sʔumu kafunu ki furu yā>

à	*nà*	*súumì*	*kàfínì*	*ké*	*fòró*	*yá*
3S	SUBJ	fast	with	DEM	blood	FOC

    'she should fast with this bleeding'

In the second instance (15), the gloss is linked to the Arabic phrase *wa lā yaqṭaʕu-hā* 'and he does not interrupt it', in which the pronoun *-hā* relates to the *ṣiyām ramaḍān* 'the fast of Ramdan' written two lines above.

(15) <ʼa nti ramaḍāna sumuyi kutunu>

à	*ntá*	*Ramaḍān*	*súumìyé*	*kútú-nú*
3S	INACP.NEG	Ramadan	fasting	interrupt

    'he does not interrupt the Ramadan fasting'

Interestingly, the name of the month in the phrase <ramaḍāna sumuyi> *Ramaḍān súumìyé* 'Ramadan fasting' is spelled etymologically, while the word *súumìyé* appears as well integrated, both in spelling (no etymological consonant) and derivation (with the Soninke nominalising suffix *-yé*).

{48} <bataqiⁿ>/<batāqi>/<batāki> *bàtáaxè* 'letter, message' from Arabic *biṭāqah* 'slip of paper, card'.[112] The Soninke lexeme does not appear in the proximity of the source word in any instances and is used to explain different Arabic

---

111 Seyidina-Ousmane Diagana (1992, 201) suggests the 2 pl. imperative form *ṣūmū* 'you all fast!' as a basis for the Soninke borrowing.
112 Seydina-Ousmane Diagana 1992, 197.

words, such as *ṣaḥāʾif* 'leaves, pages' or *jarīdah* 'list, register'. The spellings of the Ajami forms – with slight variations: <bataqiⁿ>, <batāqi> or <batāki>[113] – reflect the phonetically adjusted form. Thus, the initial *bāʾ* is vocalised with *fatḥa* (for /a/) and the second consonant is written as *ṭāʾ* as opposed to the emphatic *ṭāʾ* of the original, followed by the long or short /a/. The consonant of the last syllable is ambiguously represented by the letters *qāf* or *kāf*, each of which can encode both /x/ and /k/. The *tāʾ marbūṭah* is omitted (cf. examples above).

{49} <hakili> *háqîlè* 'mind, intelligence' from Ar. *ʕaql* 'sense reason'. The spellings of this word range from etymologically transparent to obscurer forms, i.e. from <ʕaqili> or <ʕakili> to <ḥaqili>/<ḥaⁿkili> or <haⁿkili>, with the latter spellings predominant in the corpus.[114] The variation between *kāf* and *qāf* may be purely graphical since both graphemes are used for uvular /q/, which occurs both in the source and Soninke borrowing. On the other hand, spellings with *kāf* <ḥakili> may, in fact, reflect Mandinka pronunciation with the velar /k/. Regarding the initial letters *ḥāʾ* or *hāʾ*, they stand for the Soninke /h/, which, in its turn, is a result of the phonological adaptation of the initial /ʕ/ of the source word.[115]

{50} <ḥariziki>/<ḥarziqi> 'favour, fortune' from Arabic *al-rizq* 'livelihood, blessing of God'. Although <ḥariziki>[116] corresponds to Arabic *yurzaqūna* 'they are bestowed/blessed' in one instance, its spelling only remotely resembles the original. On the other hand, the second spelling <ḥarziqi>,[117] without the source word visible nearby, reproduces the etymon *al-rizq* (pronounced [*ar-rizq*]) almost identically. While the *zāy* in the third syllable for /j/ hints at the word's etymology, the first *ḥā* represents /h/, apparently inserted to avoid an initial vowel-only syllable.[118] Unlike the word *háqîlè* examined above, the forms

---

113   TCD MS 3499 fol. 46ʳ.

114   MAAO AF 14722(87) fol. 5ᵛ; PGL ORI 11/2 fol. 13ʳ.

115   The change of initial /ʕ/ to /h/ is not very common. Seydina-Ousmane Diagana (1992, 210), for instance, cites only one further example in addition to the one discussed here. Interestingly, forms for the etymon *ʕaqil* with the initial /h/ are attested in a few other sub-Saharan languages, see Baldi 2008, 347–348.

116   TCD MS 2179 fol. 16ᵛ.

117   MAAO AF 14722(87) fol. 137ʳ.

118   It is noteworthy that Creissels (2016, 24) cites /w/ and /ŋ/ as being more typically added to 'regularise the proper syllable structure CV'. He also notes that many Arabic borrowing are retained in Soninke with the initial /a/ unpreceded by a consonant. However, /h/ appears at the beginning of the compound *hádámárénmé* 'human being', which is a calque from the Arabic *banī Adama* 'sons of Adam'. Interestingly, this word is mainly written etymologically with the initial *ʔalif*, and only exceptionally with *ḥāʾ*, in the glosses.

<ḥariziki>/<ḥarziqi> seem to have survived only in manuscripts' margins since this item is attested as *àríjaxè*/*wàrjàxé* in modern-day Soninke.[119]

{51} <ɣalabi>[120] x*àlíbè* 'writing tool'. Next to the word <ɣalabi>, which is only found in one instance, the scribe added the Arabic *al-qalam* 'pen'. However, the vowels of the modern form suggest a derivation from Tamasheq *a-ɣanib* '(reed) pen' rather than from the Arabic *qalam* '(reed) pen'. The manuscript form <ɣalabi> may, thus, represent a hybrid between *a-ɣanib* and *qalam*, motivated by the scribe's knowledge of Arabic.[121]

{52} <ʕākiⁿ>/<aⁿkiⁿ> from Ar. *ʔax* 'brother'. The word *áaxí*/*wáaxì* 'brother' in Soninke is a borrowing from Arabic, but not always recognised as such in scribal spellings. In two occurrences,[122] for instance, both linked to the Arabic text *yā ʔaxī* 'oh, my brother', the glosses are as follows: <ʕā kiⁿ ū> *áaxìn wó* and <yā aⁿ kiⁿwū> *ya áaxìn wó*. Neither spelling resembles the etymon. Interestingly, in two other instances,[123] the scribes chose the native Soninke kinship term *máarénmè* 'brother' to translate the same Arabic phrase. Finally, in one instance, the scribe wrote the word <axi> *áaxî* in the same spelling as Arabic.[124]

Another group of examples written in deviating orthographies are probably borrowings from Arabic that underwent successive phonetic changes in transmission from one linguistic community to another. Their spellings vary from etymological to 'native' in the glosses:

{53} <šuɣula> ~ <suqula>/<sūla>[125] *súxùlà* 'preoccupy, trouble' from Ar. *šaɣala*/*šuɣl* 'occupy, concern'.

<naqasi> ~ <nakasi>/<nāsi>[126] *nàqásì* 'decrease' from Ar. *naqaṣa* 'decrease'.
<waqati> ~ <wakati>/<wati>[127] *wáxátì* 'time' from Ar. *waqt* 'time, moment'.
<lawḥi> ~ <walaḥa>/<wulā>[128] *wáláhà*/*wáláxà* 'wooden board' from Ar. *(al-)lawḥ* 'board, tablet'.

---

119 However, Maninka has the form *hàrijɛɛ*. See Vydrine 1999, 249.

120 TCD MS 2179 fol. 11ʳ.

121 We are grateful to Lameen Souag for his interpretation of the manuscript form as a hybrid and for elucidating the borrowing path of the Soninke word. He further added that 'Despite appearances, this [*xàlíbè*] is at best very distantly cognate with Arabic [*qalam*]: *a-ɣanib* is a Tuareg-specific variant of pan-Berber *a-ɣanim* "reed", a loanword from Phoenican *qān-īm' (Lameen Souag's comments on an earlier draft of this paper).

122 EAP 1042/9/2 p. 24 and BULAC MS.ARA.219bis fol. 5ᵛ.

123 BULAC MS.ARA.219bis fol. 10ʳ; UPenn Lewis O35 fol. 34ʳ.

124 UPenn Lewis O35 fol. 86ᵛ.

125 AAN1 p. 179; PGL ORI 11/3 fol. 102ᵛ; BnF Arabe 5675 fol. 77ᵛ and fol. 80ʳ; TCD MS 2179 fol. 9ʳ.

126 TCD MS 2179 fol. 16ʳ; UbL Or. 14.052(8) fol. 58ʳ and fol. 59ʳ; BL Or. 6473 fol. 210ᵛ and fol. 212ʳ; TCD MS 3500 fol. 21ᵛ.

127 UbL Or. 14.052(8) fol. 63ʳ; BmT 2234 p. 104; ZOC1 di 5302.

This spelling variation reflects phonetic changes in (1) consonants: γ, $\d{h} \rightarrow x$ or k/g; (2) syllabic structures: metathesis l-w-$\d{h} \rightarrow w$-l-x; C $\rightarrow$ CV; and (3) dropping of the intervocalic uvular/velar. While the adaptation (1) and (2) probably happened primarily due to borrowing into Soninke, (3) is a clear secondary adaptation of Soninke into Mandinka, where the velar between two vowels is rare.

The next group of examples is concerned with different sources, channels or chronological layers of borrowing.

{54} *làgàré* 'last, youngest'. This word is regularly linked in the glosses to the Arabic *al-ʔāxīr* 'last, extreme', as well as other derivatives from the same root, such as *muʾaxxar* 'rear part, end'. Although it is possible that the Arabic word is the source for Soninke,[129] the etymological connection is not evident from spellings: <lākariⁿ> and <laʾri>,[130] <lāri>[131], or <laqari>.[132] The graphemes *kāf* and *qāf* may stand either for /g/, /k/ or /x/, as in /lagare/, /lakare/ or /laxare/, whereas the spellings with ʕ*ayn* and ʔ*alif* most likely indicate the dropping of the intervocalic velar/uvular, as in /laare/. Thus, both the graphic and phonetic realisations of this borrowing are relatively distant from its source. Interestingly, the word *láaxàrá* 'afterlife, hereafter', which is a phonetic adaptation of the Arabic *al-ʔāxirah*, in writing is typically spelled etymologically (i.e. retaining the definite article and the consonant *xāʾ*). Therefore, the different channel and/or time of borrowing might be assumed for these two words deriving from the same root.

{55} <ḥāli>/<ḥāri>[133] *hárì* 'even, until, so that'. This Soninke function word corresponds systematically to its potential source, the Arabic preposition/particle *ḥattā* 'until, even, so that'. The initial consonant *ḥāʾ* suggests a borrowing.[134] The consonant of the second syllable is represented either by *lām* or *rāʾ*, and never by *tāʾ* of the assumed etymon. The transformation from /t/ in Arabic to /l/ or /r/ in Soninke is unusual and not attested in other examples. Hence, the grapheme *ḥāʾ* seems inspired by its Arabic counterpart rather than

---

128 TCD MS 3499 fol. 167ᵛ and MAAO AF 14722(87) fol. 196ᵛ and fol.197ᵛ, respectively.

129 Vydrine (1999, 271) suggests the Arabic *al-ʔāxīr* 'the last' as the source for the cognate Maninka word *làgare* 'youngest child'.

130 TCD MS 2179 fol. 31ʳ and fol. 32ʳ.

131 BL Or. 6473 fol. 222ᵛ.

132 BnF Arabe 5675 fol. 69ᵛ; UbL Or. 14.522(8) fol. 18ʳ and fol. 26ʳ.

133 PGL ORI 11/1 fol. 15ʳ; TCD MS 3499 fol. 164ᵛ.

134 The initial consonant in the word *hárì* is not explained by the dialectal variation *f~h*, but rather that this word is a borrowing; see Ousmane Moussa Diagana 2013, 77–78.

dictated by an etymological connection. Instead, Tamasheq can be suggested as a source for the Soninke particle *hárì*.[135]

{56} *mìsídè* 'mosque'. The Arabic word *al-masjid* 'mosque' in the glosses is explained in two ways: <mišijidi>[136] and <misdi>[137]. While the first spelling clearly mimics the Arabic, the second, with the omitted <j>, reflects the Soninke modern-day *mìsídè*. However, the omission of /j/ is unlikely to be a result of phonetic change.[138] The form <misdi> and modern-day Soninke *mìsídè* 'mosque' could instead be derived from Ḥassānīya Arabic *msīd* 'mosque'.[139]

{57} <šafāri> *sáfàaré* 'medication' used in translating the Arabic *dawāʾ* or *ṭibb* 'medicine'. In a few instances, the scribes encoded the first consonant with *šīn*, a grapheme not frequently used in indigenous words, which led to previous assumptions that *sáfàaré* derived from Arabic *šafā* 'cure'.[140] However, this word was most certainly borrowed from another source. The word *safari* 'cure, treat' exists in Songhay, where it was borrowed from Tamasheq.[141] In this case, the Soninke Ajami spellings of *sáfàaré* with *šīn* is either another (untypical) way of encoding /s/, an awareness of the scribes that the word is ultimately a borrowing, or (as suggested by Lameen Souag) a mistaken assumption by the scribes that it derives from *šifāʾ*.

5.2.4 Definite article retained

Some borrowed Arabic nouns retain the definite article, integrated into the lexeme without any grammatical function. This section presents examples of such words, looking at the relationship between graphemic representation and phonetic realisation in Arabic and Soninke items.

---

135 Vydrine (1999, 252) suggests Tamasheq as a source for the Maninka *hálì* 'even'. Whether the Tamasheq *har* has derived from the Arabic *ḥattā* is inconclusive. Heath (2006, 215), for instance, indicates a potential relation of the Tamasheq *hàr* 'until' to the Arabic *ḥattā*. Kossmann (2005, 137–138), on the other hand, considers such a derivation unlikely.

136 Ubl Or. 14.522(8) fol. 60ᵛ.

137 BnF Arabe 5675 fol. 78ʳ.

138 See Seydina-Ousmane Diagana 1992, 206.

139 See Wexler 1980, 540–541. Lameen Souag also points out that 'msīd is a pan-Maghrebi form attested at least from Tunisia and Morocco, sometimes meaning "Quranic school"' (Souag's comments on an earlier draft of the paper).

140 Ogorodnikova 2023, 179. Note that Ousmane Moussa Diagana (2013, 180) considers the word *sáfàaré* to be an Arabic borrowing.

141 See Heath 1998, 212; Baldi 2005; Souag 2016.

The article is kept intact in some words, and it is represented in writing the same way as in Arabic by *ʔalif-lām* with *sukūn*, even though some modern forms feature a vowel added after *l*: *aljannà* <aljana>[142] from Ar. *al-jannah* 'Paradise', *alhaalà* <alḥāla>[143] from Ar. *al-ḥāl* 'condition, state', and *álíxíyáamà* <alqi-yāma>[144] from Ar. *al-qiyāmah* 'resurrection, final judgment'.

The definite article is typically retained in words borrowed from the words with the initial *ʔ* and *ʕ*. The borrowed forms in modern Soninke begin with the consonant *l-*: *liimànàaxù* from Ar. *al-ʔīmān* 'faith'; *lásílì* from Ar. *al-ʔaṣl* 'origin'; *láaxàrá* from Ar. *al-ʔāxirah* 'the hereafter'; and *láadà* from Ar. *al-ʕādah* 'custom'. The initial *ʔalif* of the article may still be written in the glosses, but such spelling is infrequent, and the words are mostly written with the initial *lām*, representing phonologically adapted forms.[145] Peculiarly, on one occasion, the /l/ of the definite article in the word *l-Àráabù* 'Arabs' <ḍaʕarābu>[146] was encoded with the Arabic *ḍād* <ḍ> (which is sometimes used for /l/).

The phonetic assimilation is ignored in a few words beginning with the 'sun letters', adhering to the conventional writing of the article with *lām*: <al-dālīliⁿ>[147] from Ar. *al-dalīl* 'sign, proof', <alðāti>[148] from Ar. *al-ðāt* 'essence', <[a]lṭālibinu>[149] from Ar. *al-ṭālib* 'student', <alšayx>[150] from Ar. *al-šayx* 'shaykh, elder' or <al-tawḥīdi>[151] from Ar. *al-tawḥīd* 'asserting oneness (of God)'. Nevertheless, the forms of these same words without the article occur much more frequently in the glosses: for example, *dàlíilè* 'proof', *jàatí* 'divine essence' or *táalíbènù* 'students'.

There are two exceptions to the previous habit – both beginning with *nūn* (which is the 'sun letter') – where the article is represented in its 'assimilated' rather than 'graphical' form. Thus, the most common spelling for the borrowing *ànnábìnínmè* 'prophet' (from Ar. *al-nabīy* pronounced as [an-nabī]) in the Ajami glosses is with the initial *ʔalif-nūn* <anabi yūmi> (the reduplication of the first consonant is underrepresented),[152] or sometimes with *ʔalif-nūn-nūn* <anna-

---

142 BULAC MS.ARA.219bis fol. 10ᵛ; PGL ORI 11/3 fol. 102ʳ.

143 BnF Arabe 5657 fol. 53ʳ; PGL ORI 11/1 and ORI 11/3 fol. 19ʳ.

144 BnF Arabe 5675 fol. 67ʳ; EAP 1042/9/6 p. 52.

145 TCD MS 3499 fol. 62ᵛ.

146 BULAC MS.ARA.219bis di 1233.

147 BL Or. 6473 fol. 205ᵛ; EAP 1042/9/3 p. 13; MAAO AF 14722(87) fol. 52ᵛ.

148 EAP 1042/9/3 p. 20.

149 EAP 1042/9/8 p. 18; PGL ORI 11/3 fol. 10ʳ.

150 PGL ORI 11/3 fol. 4ʳ.

151 BL Or. 6473 fol. 222ᵛ; PGL ORI 11/3 fol. 20ʳ and fol. 25ʳ.

152 DNN2 di 0001; MAAO AF 14722(87) fol. 2ʳ, fol. 190ᵛ, fol. 198ʳ; TCD MS 3499 fol. 45ʳ.

biyunmi>.[153] However, even the first *ʔalif* of the article can sometimes be 'replaced' with an *ʕayn* as a support for the initial vowel, with spellings such as <ʕanabiy>[154] *ànnábì* 'prophet' or <ʕaniya>[155] *ànníyà* 'intention, desire' (<Ar *al-niyyah* pronounced as [*an-niyyatu*]).

6 Towards a typology of spelling loanwords in annotated manuscripts

The OKb and Soninke data presented in Section 5 have both distinct and similar characteristics of loanwords in terms of their spelling and relationship with the original Arabic orthography. The main difference is in the parameter of the visual proximity between the Ajami word and the corresponding Arabic word. The proximity of the Arabic original in OKb seems to play a more significant role than in Soninke. It is possible that a conservative environment imposed on OKb by the Qur'an text created tighter connections between Ajami and Arabic. Such connections are best seen in OKb presented in Subsections 5.1.2 and 5.1.3. At the same time, visual proximity is a good diagnostic for the integration of a loanword when it does not exhibit etymological orthography while written close to the source word, as discussed in Subsections 5.1.4 (OKb) and 5.2.3 (Soninke).

Both OKb and Soninke data are similar in how they show some predictable tendencies along the scale from etymological to non-etymological spelling. Thus, the closest (identical) orthography is displayed by the etymons in the semantic fields of Islamic terminology – a tendency noticed in previous literature.[156] At the same time, both OKb and Soninke Ajami show that some prominent religious concepts and terms related to literacy (OKb items 17–19; Soninke items 47, 48, 50, 51) are graphemically very distant from the source words.

The orthography divergent from the original is often applied to well-integrated words, usually with a longer history of phonological adaptation. But, at the same time, the undoubtedly oldest loans that had undergone significant morphological adjustments (e.g. OKb items 4, 13) are written with clear references to the original spelling.

---

153 DNN2 di 0028 and 0029.

154 MAAO AF 14722(87) fol. 190ᵛ, fol. 208ᵛ.

155 BnF Arabe 5675 fol. 74ʳ; BULAC MS.ARA.219bis di 1247.

156 Greenberg 1960; Wexler 1980, 539–540; van den Boogert 1997, 223–229.

Such spelling retentions and divergences are probably motivated by established orthographic practices specific to a teaching-learning circle, whereas the individual sensitivity of scribes does not seem to play a prominent role in the selection of spelling features. The orthography of some loanwords (e.g. OKb items 9, 11, 15; Soninke items 36, 43, 44, 49) display a high degree of consistency, even when written by different scribes. Such standardisation tendencies suggest that the writing of these items predated the manuscripts in which they are found and that the words in question had long ago been integrated into the target language.

Remarkably, the frequency of loanwords used in manuscripts does not seem to be a spelling factor, which is in stark contrast to frequency-conditioned standardisation of language-internal words.[157]

The spelling of loanwords borrowed with the Arabic definite article *al-* sometimes indicates a kind of graphemic hypercorrection. Thus, an otherwise integrated word with the initial vowel clipped, and, therefore, retaining only the consonant /l/, may be written with the whole article restored to the original (OKb Subsection 5.1.6.1; Soninke Subsection 5.2.4). At the same time, a reverse process has been observed, whereby a written word is spelled with the clipped article but the spoken counterpart attested in modern language retains the whole article (e.g. OKb item 29).

We have tried to use the terminology introduced in Table 1 in our discussion (in Section 5) of spelling types based on Table 2. The analysis presented allows us to clarify some of the terms. The major difference is between identical and divergent spelling, which is trivial. However, the term *copy-spelling* should probably be reserved to denote conscious copying of what the scribe sees in the original Arabic word. However, given our assumption that individual groups of teachers-learners constitute a stronger factor of orthographic conventions, we should acknowledge that copy-spelling, in the strict sense, is rather difficult to identify.

In light of the analysis presented in Section 5, we should also clarify the term *insertion* suggested by Kossmann.[158] It is useful to distinguish between the narrower and wider meaning of the term. In a narrow sense, the insertion is a word strictly used to denote an Arabic concept not available in the target language. In a wider sense, however, the insertion refers to the use of a word in the exegetical register associated with practices of translation. The *exegetical insertion* has then a higher potential to be passed onto the wider circles of speakers

---

[157] Bondarev and Dobronravin 2019.

[158] Kossmann 2013, 47; cited in Section 4.1.

through register-induced contacts. Due to this potential path of transmission from insertion to integrated borrowing, it is often hard to establish whether we are dealing with the (exegetical) insertion or the etymological spelling of an already integrated word. This is because of a lack of evidence of contemporaneous use (synchronic with our data) of the given word in the spoken language of the time when manuscripts were produced.

In summary, types of spelling presented in Table 2 and exemplified in Section 5 are difficult to subdivide into neat categories. However, our preliminary typology opens up a novel perspective on Arabic loans and we hope that it may also be a useful heuristic tool to look into the patterns of borrowing in addition to previous research based on common historical linguistic principles. These types can be used further in future research, involving more data, to investigate the characteristics of borrowing outlined in Table 3.

7 Conclusion

Our paper has examined different stages of the integration of Arabic loanwords into West African languages based on examples of spelling patterns identified in two distinct manuscript traditions of OKb and Soninke. Both languages, written in Arabic script (Ajami), were used for interlinear translations of Arabic texts. These interlinear Ajami writings exhibit a high degree of Arabic loanwords, some spelled etymologically, some deviating from the Arabic graphemic source to the point of complete dissimilarity. The question 'To what extent such spelling variation reflects the integration of Arabic loanwords into the target language?' was a primary motivation for our study. We started with an analysis of the spelling of Arabic loans in OKb and Soninke annotations by using the principle of gradient etymological spelling, introduced in Subsection 4.2. Different spelling types were, thus, examined following a scalar from identical to partial etymological (further subdivided into four subtypes) to dissimilar spelling. Based on this comparison of spelling patterns, we have sketched out some tendences for a preliminary typology of spelling of Arabic loanwords in interlinear Ajami writings (Section 6). However, our findings are far from being conclusive. Both OKb and Soninke data demonstrate that spelling alone is a weak diagnostic for the degree of integration of loanwords, but together with other linguistic and cultural factors, loanwords in Ajami open up an additional perspective on the chronology and paths of borrowing.

The Ajami data should, nevertheless, be taken with caution. This is because the learned environment in which Ajami writings were produced has its own lexical base, often independent from the lexical history of the spoken languages. Although the Ajami texts in our corpus are translations bound to the source texts, with a high potential of Ajami–Arabic interplay, orthographic choices seem to be largely dictated by the conventions established within the individual learned circles, which transmit such conventions across time and manuscripts. Such circles are usually clustered around certain scholars,[159] who could be responsible for an orthographic consistency different from the other learned circles of the same speech community. The multiple and parallel channels of this kind of individualised transmission must have amplified the divergent history of Arabic loanwords – the process that has so far been unrecognised as a factor behind multiple forms of loanwords in West African languages.

Acknowledgements

---

159 See Ogorodnikova 2023 on individual distinct learning groups.

Abbreviations

1, 2, 3 = 1st, 2nd, 3rd person
1s, 2s, 3s = 1st person singular, etc.
1p, 2p, 3p = 1st person plural, etc.
AAN = Aliou Ndiaye, Adéane, manuscript
ABSTR = abstract
ADV = adverbial (operator)
AFP = argument focus perfective
ANTP = antipassive
Arabe 402 = the manuscript in the Bibliothèque nationale de France (N334 Arabe 402)
AUX = auxiliary
BL = British Library
BmT = Bibliothèque municipale de Tours
BnF = Bibliothèque nationale de France
BULAC = Bibliothèque Universitaire des Langues et Civilisations
COPEQ = equative copula
DEM = demonstrative
DNN = N.N., Dembancane, manuscript
DO = direct object
EAP = Endangered Archives Programme, the British Library
FOC = focus marker
FUT = future tense
GEN = genitive
GER = gerundive

INACP = incomplete aspect
IPFV = imperfective
IMV = imperative
JRL = John Rylands Library
MAAO = Musée national des Arts d'Afrique et d'Océanie
NEG = negative
NMAG = agent noun
NMLZ = nominalization
OKb = Old Kanembu
PL = plural
PFV = perfective
PGL = Palace Green Library
2ShK = the 'Shetima Kawo' manuscript
SBD = subordination marker
SG = singular
SJ = subject marker
SUBJ = subjunctive
TCD = Trinity College, Dublin
T.Kano = the 'Tahir Kano' manuscript
1YM = the 'Yerima Mustafa' manuscript
Q.2:72 etc. = chapter and verse of the Qur'an
UbL = Universitätsbibliothek Leiden
UPenn = University of Pennsylvania
ZOC = O.C., Ziguinchor, manuscript

References

Badawi, Elsaid M. and Muhammad Abdel Haleem (2008), *Arabic–English Dictionary of Qur'anic Usage*, Leiden: Brill.

Baldi, Sergio (1999), *Ancient and New Arabic Loans in Chadic* (University of Leipzig Papers on Africa, Languages and Literature, 7), Leipzig: Institut für Afrikanistik, Universität Leipzig.

Baldi, Sergio (2002), 'On Arabic Loans in Kanuri', *Studi Magrebini*, 25: 45–57.

Baldi, Sergio (2003), 'Some Statistical Analysis on Arabic Loans in Kanuri', in Robert Nicolai and Peter Zima (eds), *Lexical and Structural Diffusion: Interplay of Internal and External Factors of Language Development in the West African Sahel*, Nice: Sophia Antipolis, 9–15.

Baldi, Sergio (2005), 'Les emprunts berbères, surtout d'origine arabe, dans les langues de l'Afrique Occidentale', *Studi Magrebini*, n.s., 3: 103–118 <https://unora.unior.it/retrieve/handle/11574/35563/1636/10Baldic.pdf> (accessed on 8 November 2023).

Baldi, Sergio (2007), 'Kanuri', in Kees Versteegh et al. (eds), *Encyclopedia of Arabic Language and Linguistics*, vol. 2, Leiden: Brill, 552–555.

Baldi, Sergio (2008), *Dictionnaire des emprunts arabes dans les langues de l'Afrique de l'Ouest et en swahili*, Paris: Karthala.

Baldi, Sergio (2020), 'Arabic Influence in West Africa: An Overview', *Folia Orientalia*, 57: 11–23.

Bathily, Mody (2017), *Le grand dictionnaire: soninké–français, français–soninké, s.l.: s.n.*

Bondarev, Dmitry (2013a), 'Language Change Induced by Written Codes: A Case of Old Kanembu and Kanuri Dialects', in Esther-Miriam Wagner, Ben Outhwaite and Bettina Beinhoff (eds), *Scribes as Agents of Language Change* (Studies in Language Change, 10), Boston: De Gruyter Mouton, 291–323.

Bondarev, Dmitry (2013b), 'Qur'anic Exegesis in Old Kanembu: Linguistic Precision for Better Interpretation', in Tal Tamari and Dmitry Bondarev (eds), *Qur'anic Exegesis in African Languages = Journal of Qur'anic Studies,* 15/3: 56–83.

Bondarev, Dmitry (2014a), 'Multiglossia in West African Manuscripts: A Case of Borno, Nigera', in Jörg B. Quenzer, Dmitry Bondarev and Jan-Ulrich Sobisch (eds), *Manuscript Cultures: Mapping the Field* (Studies in Manuscript Cultures, 1), Berlin: De Gruyter, 113–155.

Bondarev, Dmitry (2014b), 'Old Kanembu and Kanuri in Arabic Script: Phonology through the Graphic System', in Mumin and Versteegh 2014, 107–142.

Bondarev, Dmitry (2017), 'Islamic Education and Ample Space Layout in West African Manuscripts', in Brigaglia and Nobili 2017, 105–142.

Bondarev, Dmitry (2021), 'A Typology of West African Ajami Manuscripts: Languages, Layout and Research Perspectives', in Jörg B. Quenzer (ed.), *Exploring Written Artefacts: Objects, Methods, and Concepts* (Studies in Manuscript Cultures, 25), Berlin: De Gruyter, 707–728.

Bondarev, Dmitry (2022), 'Grammar and Art of Translation as Expression of Muslim Faith: Translational Practices in West Africa', in Hephzibah Israel (ed.), *Routledge Handbook on Translation and Religion*, London: Routledge, 415–431.

Bondarev, Dmitry (2024), 'Shifting Regimes, Reshaping Manuscripts: Disappearance of Annotated Qurʾans in Borno', in Christiane Gruber and Bihter Esener (eds), *Regime Change: New Horizons in Islamic Art and Visual Culture* (Gingko Library Art Series), London: Gingko, 32–46.

Bondarev, Dmitry (forthcoming), *Old Kanembu: A Saharan language of the Qur'anic glosses.*

Bondarev, Dmitry and Nikolay Dobronravin (2019), 'Standardisation Tendencies in Kanuri and Hausa Ajami Writings', in Dmitry Bondarev, Alessandro Gori and Lameen Souag (eds), *Creating Standards: Interactions with Arabic Script in 12 Manuscript Cultures* (Studies in Manuscript Cultures, 16), Berlin: De Gruyter, 237–269.

Bondarev, Dmitry and Abba Tijani (2014), 'Performance of Multilayered Literacy: Tarjumo of the Kanuri Muslim Scholars', in Kasper Juffermans, Yonas Mesfun Asfaha and Ashraf K. Abdelhay (eds), *African Literacies: Ideologies, Scripts, Education*, Newcastle-upon-Tyne: Cambridge Scholars Publishing, 118–146.

Brigaglia, Andrea and Mauro Nobili (eds) (2017), *The Arts and Crafts of Literacy: Manuscript Cultures in Muslim Sub-Saharan Africa* (Studies in Manuscript Cultures, 12), Berlin: De Gruyter.

Chonai, Hasan (1999), *Gruppa teda-daza (tsentral'nosaharskaya semya yazikov) i yeyo genetichesckiye vzaimodeystviya (etimologichesckiy i fonologichesckiy aspekt)* Группа теда-даза (центральносахарская семья языков) и ее генетические взаимодействия (этимологический и фонологический аспект) ('Genealogical Relationships in Teda-Daza

Language Group (Central Saharan Linguistic Family): Phonology and Etymology'), un-published PhD thesis, Russian State Humanitarian University.

Creissels, Denis (2016), 'Phonologie segmentale et tonale du soninké (parler du Kingi)', *Mandenkan*, 55: 3–174.

Cultru, Prosper (1913), *Premier voyage du Sieur de La Courbe fait à la Coste d'Afrique en 1685, Publié avec une carte de Delisle et une introduction*, Paris: Edouard Champion / Emile Larose.

Cuoq, Joseph (1984), *Histoire de l'Islamisation de l'Afrique de l'Ouest des origines à la fin du XVIe siècle,* Paris: Geuthner.

Dantioko, Oudiary Makan (2003), *Dictionnaire soninké–français*, Bamako: Jamana.

De Moraes Farias, Paulo F. (2003), *Arabic Medieval Inscriptions from the Republic of Mali: Epigraphy, Chronicles, and Songhay-Tuareg History* (Fontes Historiae Africanae, New Series, 4), Oxford: Oxford University Press / New York: The British Academy.

Diagana, Ousmane Moussa (2013), *Dictionnaire soninké–français (Mauritanie)* Paris: Karthala.

Diagana, Ousmane Moussa (1995), *La langue soninké: morphosyntaxe et sens à travers le parler de Kaédi (Mauritanie)*, Paris: Harmattan.

Diagana, Seydina-Ousmane (1992), *Contact de langues: Approche sociolinguistique des emprunts du soninké au français, à l'arabe et au pulaar*, PhD thesis, Université de Paris V.

Dimmendaal, Gerrit J. (2011), *Historical Linguistics and the Comparative Study of African Languages,* Amsterdam: John Benjamins.

Dramé, Djibril (2021), 'Le soninké dans les manuscrits islamiques: Corrélation entre les correspondances graphématiques et phonémiques', *Mandenkan*, 66: 187– 203.

Dramé, Djibril (2022), *Mapping Dialectal Variation in Soninke Islamic Manuscripts*, PhD thesis, University of Hamburg.

Greenberg, Joseph H. (1960), 'Linguistic Evidence for the Influence of the Kanuri on the Hausa', *The Journal of African History,* 1/2: 205–212.

Gutelius, David (2000), 'Newly Discovered 10th/16th Century Ajami Manuscript in Niger and Kel Tamasheq History', *Saharan Studies Association Newsletter*, 8/1–2: 6.

Heath, Jeffrey (1998), *Dictionnaire songhay–anglais–français*, vol. 1: *Koyra Chiini*, Harmattan: Paris.

Heath, Jeffrey (2006), *Dictionnaire touareg du Mali: tamachek–anglais–français*, Paris: Khartala.

Hiskett, Mervyn (1984), *The Development of Islam in West Africa*, London: Longman.

Hutchison, John P. (1981), *The Kanuri Language: A Reference Grammar*, Madison: University of Wisconsin.

Insoll, Timothy (2003), *The Archaeology of Islam in Sub-Saharan Africa*, Cambridge: Cambridge University Press.

Kossmann, Maarten (2005), *Berber Loanwords in Hausa* (Berber Studies, 12), Cologne: Rüdiger Köppe.

Kosmann, Maarten (2013), *The Arabic Influence on Northern Berber* (Studies in Semitic Languages and Linguistics, 67), Leiden: Brill.

Levtzion, Nehemia and Randall L. Pouwels (eds) (2000), *The History of Islam in Africa*, Athens, OH: Ohio University Press / Oxford: James Currey / Cape Town: David Philip.

Lukas, Johannes (1957), *Die Sprache der Tubu in der Zentralen Sahara* (Deutsche Akademie der Wissenschaften zu Berlin, Institut für Orientforschung, 14), Berlin: Akademie-Verlag.

McLaughlin, Fiona (forthcoming), 'Ajami Writing Practices in Atlantic-Speaking Africa', in Friederike Lupke (ed.), *The Oxford Guide to the Atlantic Languages* (Oxford Guides to the World's Languages), Oxford: Oxford University Press; for an online prepublication version,

see <https://people.clas.ufl.edu/fmcl/files/AjamiCIRCRED.pdf> (accessed on 20 September 2023).

Maschler, Yael (1994), 'Metalanguaging and Discourse Markers in Bilingual Conversation', *Language in Society*, 23/3: 325–366 <doi:10.1017/S0047404500018017>.

Mumin, Meikal and Kees Versteegh (eds) (2014), *The Arabic Script in Africa: Studies in the Use of a Writing System* (Studies in Semitic Languages and Linguistics, 71), Leiden: Brill.

Newman, Paul (2000), *The Hausa Language: An Encyclopedic Reference Grammar* (Yale Language Series), New Haven: Yale University Press.

Ngom, Fallou (2017), 'West African Manuscripts in Arabic and African Languages and Digital Preservation', in Thomas Spear (ed.), *Oxford Research Encyclopedia of African History*, <https://doi.org/10.1093/acrefore/9780190277734.013.123>.

Nixon, Sam (ed.) (2017), *Essouk-Tadmekka: An Early Islamic Trans-Saharan Market Town* (Journal of African Archaeology Monograph Series, 12), Leiden/Boston: Brill.

Ogorodnikova, Darya (2017), '*Ajamī* Annotations in Multilingual Manuscripts from Mande Speaking Areas: Visual and Linguistic Features', *Islamic Africa*, 8: 111–143.

Ogorodnikova, Darya (2021), '"I Heard It from My Teacher": Reflections on Transmission of Knowledge in Islamic Manuscripts from Senegambia and Mali', in Stefanie Brinkmann, Giovanni Ciotti, Martin Delhey and Stefano Valente (eds), *Education Materialized: Reconstructing Teaching and Learning Contexts through Manuscripts* (Studies in Manuscript Cultures, 23), Berlin: De Gruyter, 127–150.

Ogorodnikova, Darya (2023), *Soninke Ajami Manuscripts and Islamic Education: A Reconstruction of Scribal Practices*, PhD thesis, Universität Hamburg.

Salvaing, Bernard (2020), 'Islam in Sub-Saharan Africa', in Fallou Ngom, Mustapha H. Kurfi and Toyin Falola (eds), *The Palgrave Handbook of Islam in Africa*, Cham: Springer International Publishing.

Salvaing, Bernard and John Hunwick (2003), 'Writers of Guinea', in John Hunwick (ed.), *Arabic Literature of Africa*, vol. 4: *The Writings of Western Sudanic Africa* (Handbuch der Orientalistik, 4), Leiden: Brill, 491–529.

Souag, Lameen (2010), 'Ajami in West Africa', *Afrikanistik Aegyptologie Online*, <http://www.afrikanistik-aegyptologie-online.de/archiv/2010/2957> (accessed on 22 August 2023).

Souag, Lameen (2015), 'Archaic and Innovative Islamic Prayer Names around the Sahara', *Bulletin of the School of Oriental and African Studies*, 78/2: 357–374.

Souag, Lameen (2016), 'Language Contact in the Sahara', in Thomas Spear (ed.), *Oxford Research Encyclopedia of African History*, <https://doi.org/10.1093/acrefore/9780199384655.013.141>.

Tamari, Tal (2016), 'Styles in Islamic Education: Perspectives from Mali, Guinea, and The Gambia', in Robert Launay (ed.), *Islamic Education in Africa: Writing Boards and Blackboards*, Bloomington / Indianapolis: Indiana Univ. Press, 29–60.

Tamari, Tal (2017), 'Bamana Texts in Arabic Characters: Some Leaves from Mali', in Brigaglia and Nobili 2017, 207–278.

Tamari, Tal and Dmitry Bondarev (2013), 'Introduction and Annotated Bibliography', in Tal Tamari and Dmitry Bondarev (eds), *Qur'anic Exegesis in African Languages = Journal of Qur'anic Studies*, 15/3: 1–55.

Van den Boogert, Nico (1997), *The Berber Literary Tradition of the Sous, with an Edition and Translation of 'The Ocean of Tears' by Muhammad Awzal (d. 1749)*, Leiden: Nederlands Instituut voor het Nabije Oosten.

Vydrine, Valentin (1998), 'Sur l'écriture Mandingue et Mandé en caractères arabes (mandinka, bambara, soussou, mogofin)', *Mandenkan*, 33: 1–87.

Vydrine, Valentin (1999), *Manding–English Dictionary: Maninka, Bamana*, vol. 1, Saint Petersburg: Dimitry Bulanin.

Vydrin, Valentin (2014), 'Ajami Scripts for Mande Languages', in Mumin and Versteegh 2014, 199–224.

Wexler, Paul (1980), 'Problems in Monitoring the Diffusion of Arabic into West and Central African Languages', *Zeitschrift der Deutschen Morgenlandischen Gesellschaft*, 130: 522–556.

Zappa, Francesco (2011), 'When Arabic Resonates in the Words of an African Language: Some Morphological and Semantic Features of Arabic Loanwords and Calques in Bambara', in Guiliano Lancioni and Lidia Bettini (eds), *The Word in Arabic* (Studies in Semitic Languages and Linguistics, 62), Leiden: Brill, 229–249.