# Soldiers, coffee and markets: On Baldi's *Dictionary of Arabic Loanwords in the Languages of Central and East Africa*

GIAN CLAUDIO BATIC

University of Naples L'Orientale
gcbatic@unior.it

The Dictionary of Arabic Loanwords in the Languages of Central and East Africa (hereafter DoAL) (2021) is Sergio BALDI's second magnum opus on Arabic loanwords in the languages of Africa. Preceded by Dictionnaire des emprunts arabes dans les langues de l'Afrique de l'Ouest et en Swahili (hereafter DEAr) (2008), the DoAL shifts the focus to the eastern and central quadrant of the African continent, also including and extending the data on Arabic loanwords in Swahili already present in the DEAr. From a general perspective and ideally, the DEAr and the DoAL should be considered as two parts of a single work in progress aimed at compiling all existing Arabic loanwords in the languages of the African continent. Of course, an ultimate goal of this kind clashes with the type of data available, so that any compilation work is inevitably limited by the nature of the sources. In the next section we will devote some space to the discussion of the actual representativeness of the DoAL in relation to the target languages and to their reality beyond the lexical descriptions we can provide of them.

The DEAr and the DoAL adopt the same indexing system for Arabic terms and present an isomorphic structure, so that the user can easily move from one volume to the other using the same numbering system and without having to go through new consultation guidelines. The end user will appreciate the structural consistency of the two dictionaries, in addition to the value given by the scientific compilation work that makes the DoAL – as it was for its predecessor – a reference work in the field of African lexicography.

This article has a twofold purpose: on the one hand it aims to provide critical observations of a theoretical and practical nature, and on the other hand it intends

to outline possible uses and lines of research centred on the lexico-graphic data provided by the DoAL.

Before starting the discussion on some aspects of the DoAL, it seems important to frame the work in the perspective of the Author, i.e. to state his proposed objectives. The DoAL

"[…] aims to provide a tool for researchers of African languages, to offer them a lexicon of these languages if they want to indicate Arabic loanwords in their work" (p. 1)

BALDI also states that his

"[…] objective was to collect all loanwords, such that this work is as complete as possible, though some words may not be commonly used by speakers of a given language that seem to be pure Arabic xenisms […]" (p. 5)

And it is precisely as a tool that we will look at the DoAL, trying to understand how it was designed and what uses it lends itself to.

## 1. Observations

Three observations can be made about the DoAL. They concern, in order, a) the languages examined, b) the problem of representativeness, and c) the nature of the sources.

### *1.1 Target languages*

In the DoAL the Author considers 57 languages (plus Arabic). The list of languages analysed (pp. 2-4) provides the names used by the authors of the individual lexical or lexico-grammatical descriptions. Since many of the sources consulted are – with varying degrees of value – pioneering or dated, the naming of languages is sometimes obsolete and misleading and differs from the standard denominations used in modern language catalogues such as, for example, the ISO e Glottolog databases and the SIL's *Ethnologue*. In some cases the Author indicates groups of languages (e.g. Batéké, i.e. Teke languages; Gbéa, i.e. Gbaya languages), and in other cases the name of the most widespread variant whereas the source deals with a specific dialect or sub-variant. In Table 1, for each target language, the name of the language used by the Author, the standard name used in scientific literature (if different from that provided by the Author) or the name of the language actually covered in the source, the Glottolog code (HAMMARSTRÖM et al. 2021), and the ISO 639-3 code are provided. The languages are sorted alphabetically according to their genetic affiliation (the top-level family is indicated).

GIAN CLAUDIO BATIC
**Soldiers, coffee and markets**

| TOP-LEVEL FAMILY | LANGUAGE | LANGUAGE NAME | GLOTTOCODE | 639-3 |
|---|---|---|---|---|
| Afro-Asiatic | Rendille | – | rend1243 | rel |
| Atlantic-Congo | Ankole | Nyankole | nyan1307 | nyn |
| | Lunyankole, Runyankore [1] | | | |
| Atlantic-Congo | Batéké | Teke languages | kasa1251 | – |
| Atlantic-Congo | Bemba | – | bemb1257 | bem |
| Atlantic-Congo | Bende | – | bend1258 | bdp |
| Atlantic-Congo | Ciluba | Luba-Lulua | luba1249 | lua |
| Atlantic-Congo | Digo[2] | – | Digo1243 | dig |
| Atlantic-Congo | Gbéa | Gbaya languages | gbay1279 | – |
| Atlantic-Congo | Gikuyu | Kikuyu | kiku1240 | kik |
| Atlantic-Congo | Gmbwaga | Ngbaka Ma'bo | ngba1284 | nbm |
| Atlantic-Congo | Ngh'wele | Kwere | gwer1238 | gwr |
| Atlantic-Congo | Haya | – | haya1250 | hay |
| Atlantic-Congo | Ila | – | ilaa1246 | ilb |
| Atlantic-Congo | Jita | – | jita1239 | jit |
| Atlantic-Congo | Kamba | – | kamb1297 | kam |
| Atlantic-Congo | Kikongo[3] | – | (core1256)[4] yomb1244 | kon[5] |
| Atlantic-Congo | Kiluba | Luba-Katanga | luba1250 | lub |
| Atlantic-Congo | Kinyarwanda | – | kiny1244 | kin |
| Atlantic-Congo | Kiw'oso | Bosho | (mach1266) | (jmc) |
| Atlantic-Congo | Kuria | – | kuri1259 | kuj |
| Atlantic-Congo | Lega | (Beya-Musange dialect) | lega1253 | – |
| Atlantic-Congo | Lhukonzo | Konzo | konz1239 | koo |
| Atlantic-Congo | Lingala | Kinshasa Lingala | ling1263 | lin |
| Atlantic-Congo | Lomongo | Mongo | mong1338 | lol |
| Atlantic-Congo | Lotuxo | Otuho | otuh1238 | lot |
| Atlantic-Congo | Luena | Luvale | luva1239 | lue |
| Atlantic-Congo | Luganda | Ganda | gand1255 | lug |
| Atlantic-Congo | Lunyoro | Nyoro | nyor1246 | nyo |

[1] The Author lists the Nyankole language under three different names: Ankole, Lunyankole and Runyankore.

[2] Although found in the DoAL (e.g. entry 1862, **'askāri** 'military, army'), the language is not present in the initial list (pp. 2-4).

[3] The Author is considering the Kikongo (i.e. Kongo) cluster. At least one of BENTLEY's two works (1887, 1895) on the Kikongo (i.e. Kongo) variant spoken in M'banza-Kongo (San Salvador) has been consulted (the author is listed as one of the sources for Kikongo, p. 3), but the reference is missing.

[4] Out-dated.

[5] Inclusive code (kon > kng, ldi, kwy, yom).

GIAN CLAUDIO BATIC
**Soldiers, coffee and markets**

| | | | | |
|---|---|---|---|---|
| Atlantic-Congo | Macua | Makhuwa | makh1264 | vmw |
| Atlantic-Congo | Matengo | – | mate1258 | mgv |
| Atlantic-Congo | Ndogo | – | ndog1248 | ndz |
| Atlantic-Congo | Ndonga | – | ndon1254 | ndo |
| Atlantic-Congo | Ngh'wele | Kwere | gwer1238 | gwr |
| Atlantic-Congo | Ngombe | – | ngom1268 | ngc |
| Atlantic-Congo | Nyakyusa | Nyakyusa-Ngonde | nyak1260 | nyy |
| Atlantic-Congo | Pokomo | – | poko1261 | pkb |
| Atlantic-Congo | Sango | – | sang1238 | sag |
| Atlantic-Congo | Shona | – | shon1251 | sna |
| Atlantic-Congo | Sukuma | – | suku1261 | suk |
| Atlantic-Congo | Swahili | – | swah1253 | swh |
| Atlantic-Congo | Xhosa | – | xhos1239 | xho |
| Atlantic-Congo | Zande | – | zand1248 | zne |
| Atlantic-Congo | Zulu | – | zulu1248 | zul |
| Austronesian | Malagasy | Kibosy Kiantalaotsy-Majunga | bush1250 | buc |
| Central Sudanic | Madi | Ma'di | madi1260 | mhi |
| Khoe-Kwadi | Khoi-Khoin | Nama | nama1264 | naq |
| Kuliak (Nilo-Saharan) | Ik | – | ikkk1242 | ikx |
| Nilotic | Acooli | Acoli | acol1236 | ach |
| Nilotic | Alur | – | alur1250 | alz |
| Nilotic | Anywa | Anuak | anua1242 | anu |
| Nilotic | Ateso | Teso | teso1249 | teo |
| Nilotic | Bari | – | bari1284 | bfa |
| Nilotic | Dholuo | Luo | luok1236 | luo |
| Nilotic | Dinka | Dinka languages | dink1262 | din |
| Nilotic | Pokot | Pökoot | poko1263 | pko |
| Nilotic | Shilluk | – | shil1265 | shk |

Table 1 – Languages analysed in the DoAL

### *1.2 Representativeness*

The second issue to be addressed emerges – almost immediately – from the title and concerns the very object of the lexicographic compilation and its place in the real world. Since dictionaries are inevitably partial – albeit in many cases accurate and valuable – representations of the lexical inventory of a language, the DoAL could actually be labelled as a 'dictionary of Arabic loanwords *in the dictionaries* of the languages of Central and East Africa': that is, it is a refined product of partial and cultural products that cannot convey to us the ultimate truth, that is, the adherence of the data to the lexical reality as it manifests itself in the community of speakers. The use of compilations as intermediaries between the data and the

final product (i.e. the DoAL) is in itself unavoidable when operating on a large scale, but it is worth remembering that the existence of the loanwords dealt with in the DoAL is anchored not so much to the languages as to their available representations (lexical descriptions, dictionaries, glossaries). Certainly, it can be argued that any compilation or survey work of this kind is a product of products, or indeed that research is often a high-refined product built on sources (data collected and somehow processed) and not on raw data.

### *1.3 Nature of sources*

In addition to the structural limitation of lexical descriptions, the inhomogeneity of the available sources must be pointed out. This inhomogeneity acts as much on a synchronic as on a diachronic level. It happens in fact that relatively recent lexical descriptions set themselves different goals of completeness, thus describing a current language but with a different degree of depth, or that descriptions far back in time are used alongside recent or less recent descriptions. The qualitative difference of the sources makes them difficult to compare: does HULSTAERT's (1952) impressive *Dictionnaire lɔmɔ́ngo-français* have the same value as an article devoted to loans, such as KABUTA's (1998) for the Ciluba (i.e. Luba-Lulua) language? Of course, the issue has not escaped the Author, even with regard to publications that fall into the same category:

> "The reliability of the works I consulted varies: the Swahili dictionary of Johnson is less reliable than that of father Sacleux with regard to quoting loanwords in general and Arabic in particular" (p. 5)

The Author worked with the sources he had at his disposal and was able to consult, being forced in many cases to rely on a single lexical description to find Arabic loans in a given language. This limitation, in the landscape of publications devoted to the lexical description of minority languages, is physiological. We should not be surprised if for Swahili, the Author lists ten sources, while for languages such as Lotuxo (i.e. Otuho) and Sukuma only one. To what extent is the abundance of sources for some languages and the scarcity of sources for others reflected in the final compilation? If for Swahili (and, in the DEAr, for Hausa) we can work with various sources that present a satisfactory degree of exhaustiveness, for other languages (most of them) we are somehow forced to rely on the quality of the only available source, resetting to zero all the parameters that usually guide us in the selection of sources (in no particular order: period of publication, type of publication, prestige of the publishing house, reputation of the author, number of entries, etc.).

## 2. Loanword mapping

The DoAL is an extremely useful tool for mapping the distribution of loans and their diffusion centres. The distribution of the languages analysed in the DoAL is shown in Figure 1 (the coordinates used to build the maps are taken from HAMMARSTRÖM et al. 2021):
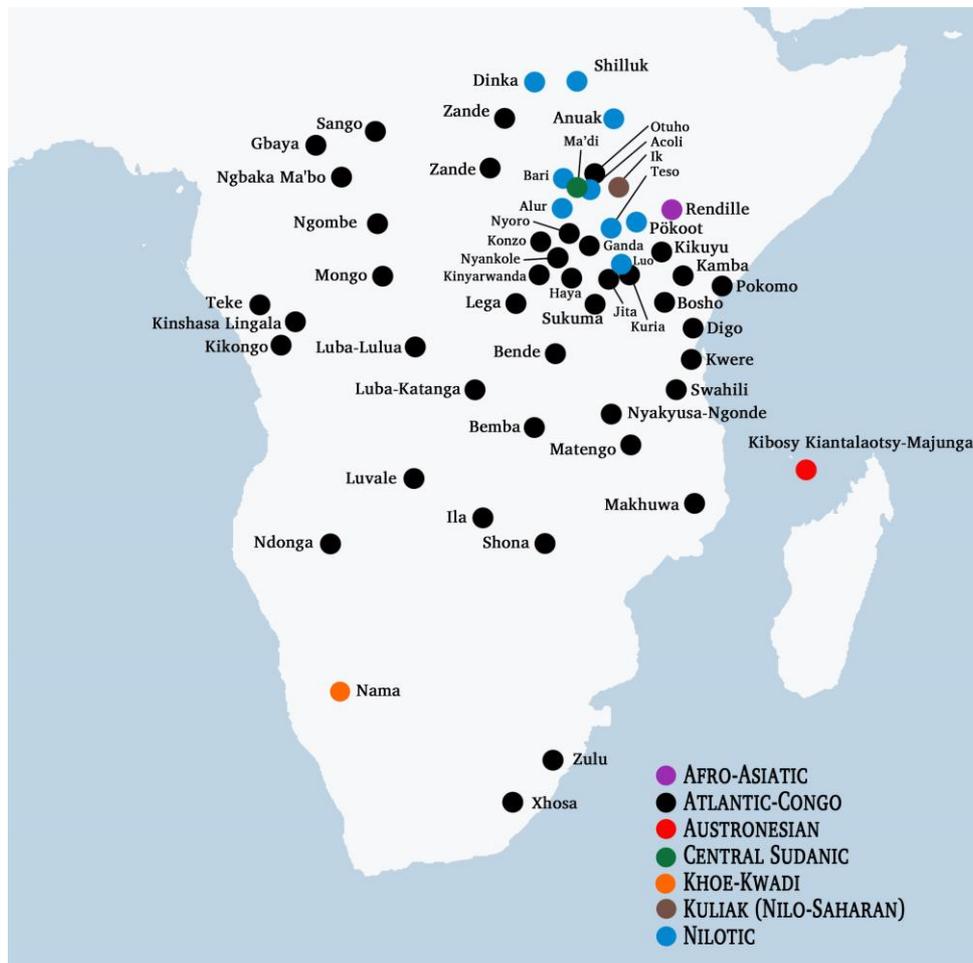


Figure 1 – Distribution and genetic affiliation of the languages examined in the DoAL

In order to analyse the diffusion paths, or at least the major centres of diffusion, it will be necessary to consider 'successful' Arabic loans, i.e. loanwords integrated into a significant number of languages. Figures 2-7 (see pages 127-129) show the distributions of the following Arabic loanwords: **sāʻa** 'while, hour, clock' (#1398),

**sūq** 'market' (#1402), **'askāri** 'military, army' (#1862), **qalam** 'pen' (#2335), **qahwa** 'coffee' (#2350), and **māl** 'wealth, money' (#2684).[6]

Where the Arabic term has been integrated through another language (as indicated in the DoAL), a line connects the intermediate donor language to the language(s) that borrowed the term. In this case, the ISO codes of the intermediate donor language and the recipient languages are indicated.

As can be seen, Swahili plays a prominent role as a second donor language. Two observations can be made in this regard:

(a) the influence of Arabic acts primarily on the east coast. Arabic terms are integrated into Swahili and later, through Swahili, into other languages in contact with the Swahilophone world;

(b) the spread of Arabic loanwords through Swahili follows a north-west direction. The range of Swahili as a donor language finds its limit in the belt that includes north-central Kenya and the area bordering the northern and western shores of Lake Victoria. A first analysis based exclusively on the mapping of recipient languages allows us to formulate the hypothesis of the existence of two centres of diffusion: a southern centre corresponding to the Swahili-speaking area and exerting its influence towards the north and north-west, and a northern centre corresponding to the Nilotic and Afro-Asiatic area (although the latter is hardly represented in the DoAL) acting towards the south. This observation is quite evident if we consider the close contiguity with the Arabic language and the Arab-Islamic world of both the Swahilophone coastal area and the region corresponding to present-day Sudan and South Sudan. It is also important to stress the role played by the contact of the Nilotic and Afro-Asiatic languages with Sudanese Arabic (apd; suda1236) in Sudan, South Sudan and Eritrea.

The intermediate donor language is not always Swahili. Figures 2 and 6 indicate the case of the Kinshasa Lingala and Sango languages: the Arabic loanwords **sā'a** 'while, hour, clock' and **qahwa** 'coffee' are supposed to have entered the Sango language via Kinshasa Lingala (in the case of **qahwa**, the Lingala term is not indicated in the DoAL).

---

[6] The six loanwords were selected to exemplify an informative mapping and provide elements for further analysis. Other terms that would have proved equally significant are, among others, **bunduqīya / bundūg / bundug** 'rifle, gun' (#317), **ḫaima** 'tent' (#869), and **sukkar** 'sugar' (#1314).
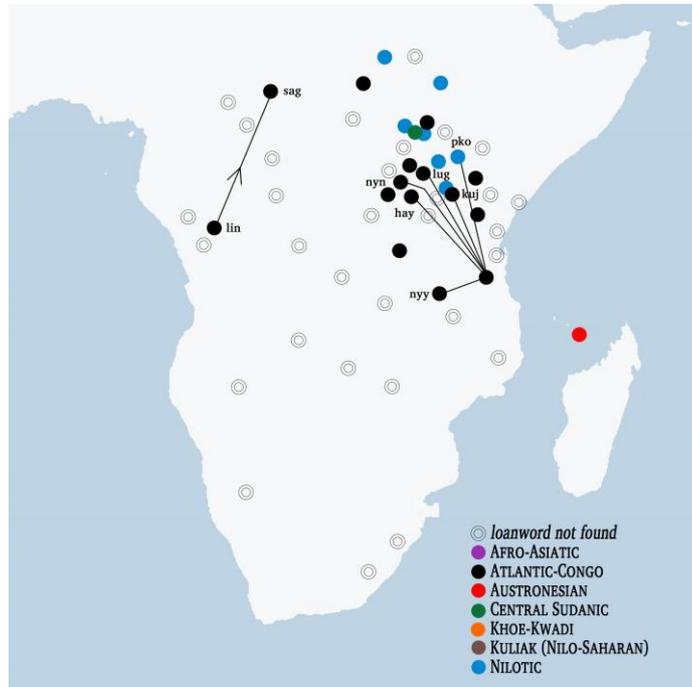
Figure 2 – **sāʻa** 'while, hour, clock' (#1398)
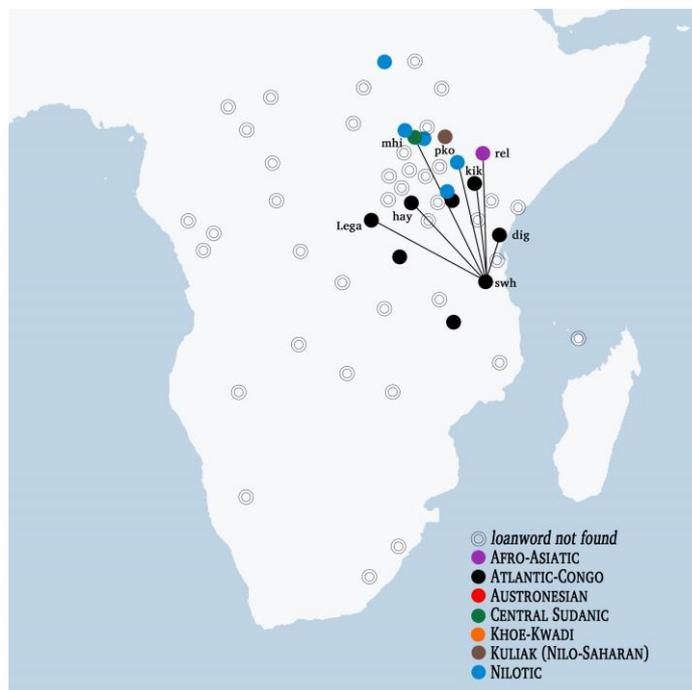


Figure 3 – **sūq** 'market' (#1402)
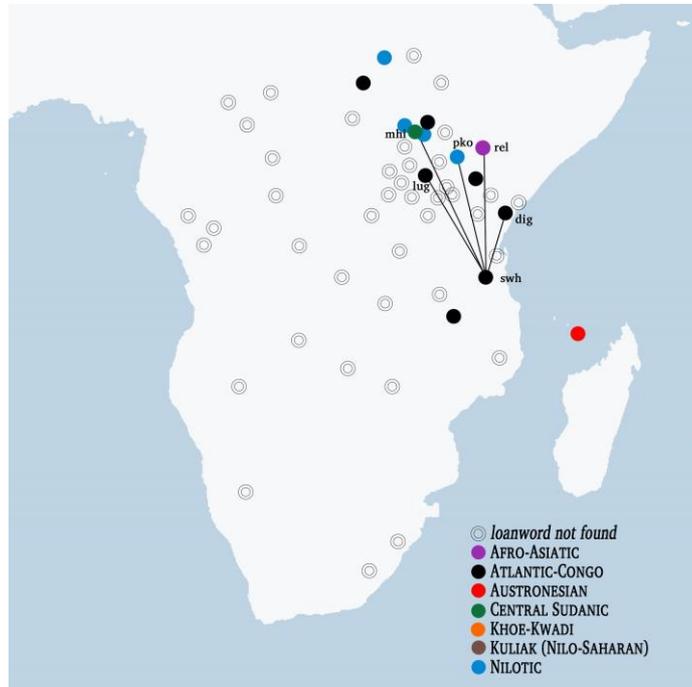
Figure 4 – **'askāri** 'military, army' (#1862)
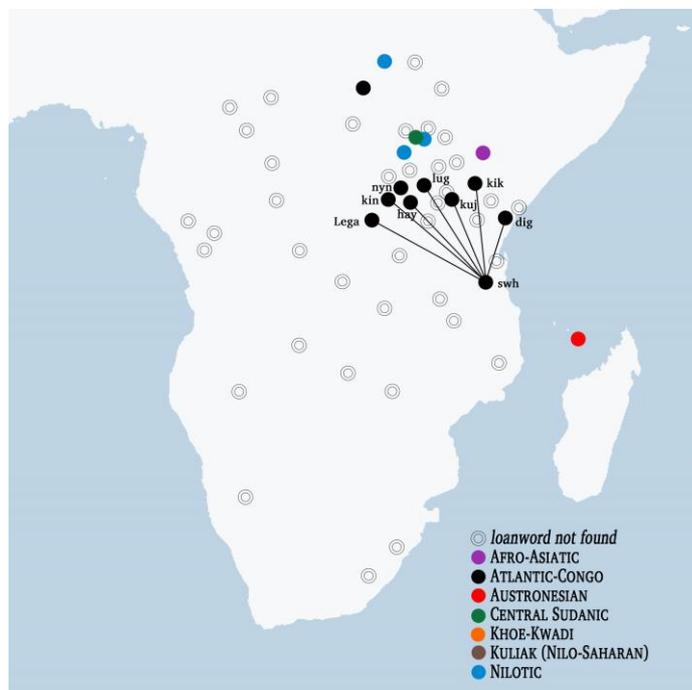


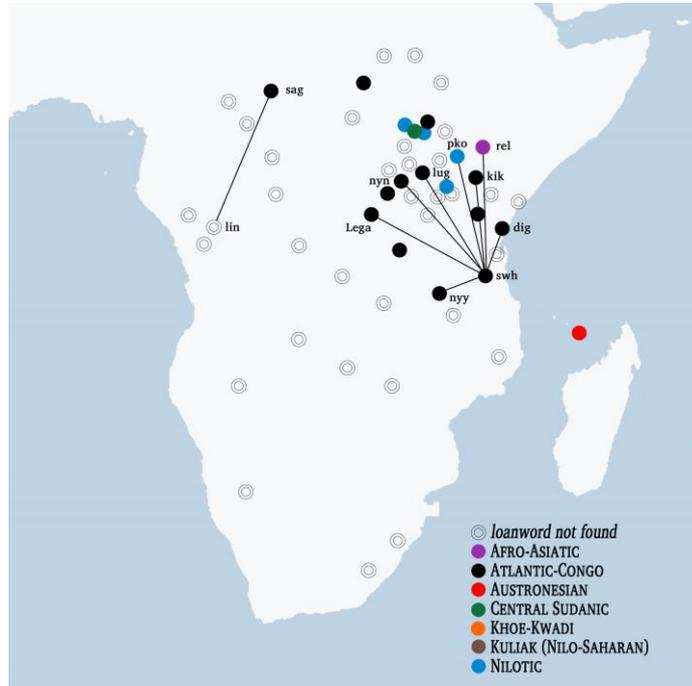Figure 5 – **qalam** 'pen' (#2335)
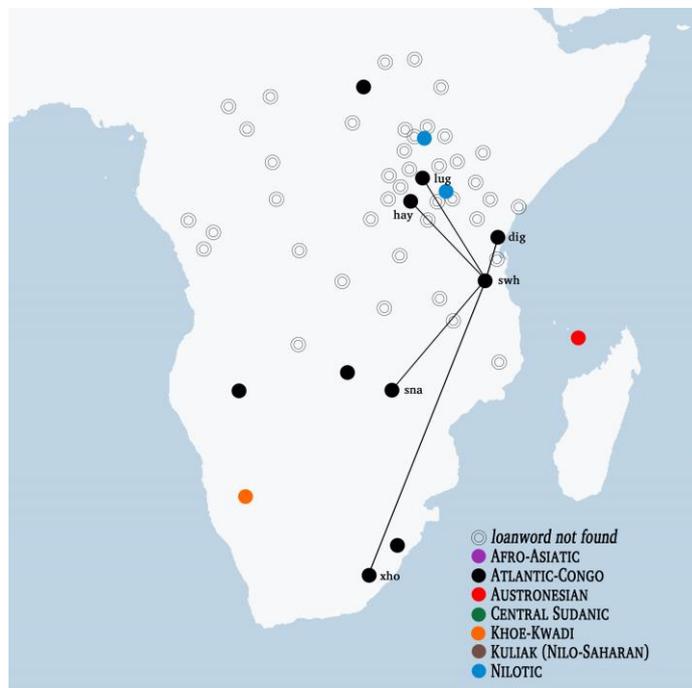
Figure 6 – **qahwa** 'coffee' (#2350)



Figure 7 – **māl** 'wealth, money' (#2684)

GIAN CLAUDIO BATIC
**Soldiers, coffee and markets**

The influence of a language as a vehicle of loanwords or the extension of material and immaterial objects resulting from direct or indirect contact with Arab-Islamic culture can be measured not so much by mapping individual loanwords, but rather by mapping entire lexical fields. The mapping of loanwords related to money (spending money, changing money, obtaining money as a result of a business transaction, etc.) and currency (coin(s), silver, dinar, drachma, piaster, etc.), for instance, would give us the situation illustrated in Figure 8:
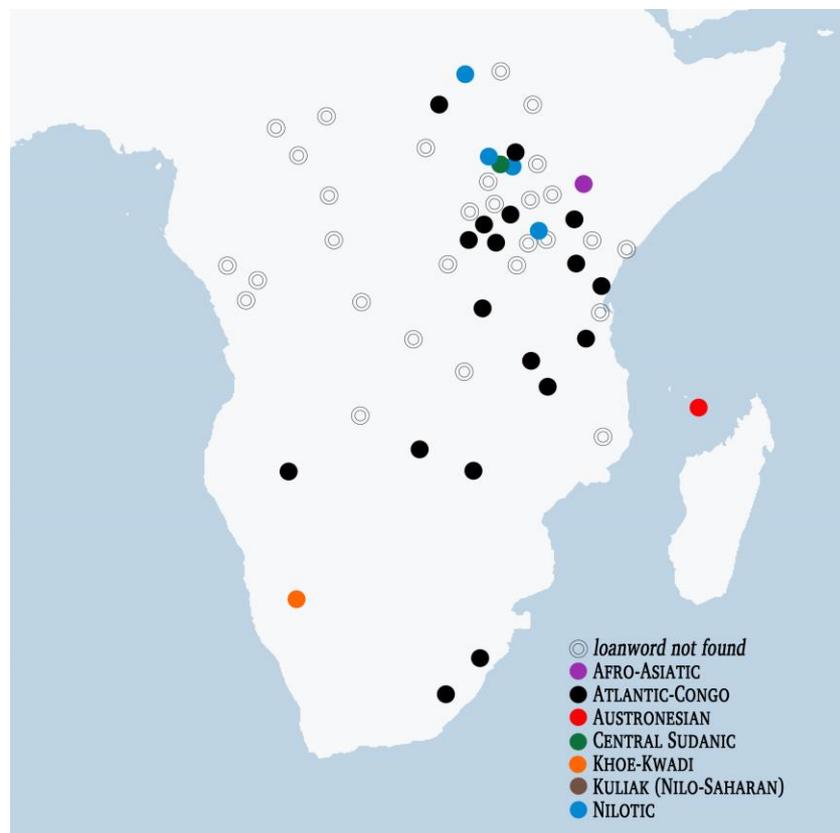


Figure 8 – Distribution of the money-currency lexical field

The lexical field considered consists of 19 Arabic terms distributed across 26 languages. The most receptive language is, predictably, Swahili (16 loans), followed by Kibosy Kiantalaotsy-Majunga (4), and by the Acoli, Ma'di and Diga languages (3 loans each). All other languages do not have more than two loanwords.

The DoAL can be easily used to conduct this kind of analysis by resorting to the Arabic, English and French indices at the end of the volume. Of particular interest is the comparison of purely lexical data (such as those obtained from the DoAL)

with historical data concerning the diffusion of material and immaterial culture elements of Arab and Arab-Islamic origin pertaining, for example, to trade, administrative practices, and religion.

We know that the Author has "consulted thousands of dictionaries" (p. 1). Unfortunately, we do not know to what extent the non-reporting of loanwords in certain languages or areas is due to the absence of adequate sources or to the absence of loans. This limits the ability to carry out a complete mapping of loanwords: it is not possible to determine, relying solely on the DoAL, which languages (apart from the problem of source reliability discussed above) have demonstrated a degree of impermeability to the penetration of Arabic loanwords (and indirectly to Arab and Arab-Islamic culture).

## 3. A conclusive remark

The main aim of this brief discussion was to formulate critical observations and explore some potentialities offered by the DoAL.

The DoAL is the result of years of research in the field of Arabic loans in African languages. A solid work of bibliographic research as well as a rigorous compilation, the volume promises no more than what it gives – and what it gives is a lot. Indeed, the DoAL presents itself as a solid basis for exploring the dynamics and patterns of penetration and integration of Arabic terms in African societies, which makes it a valuable tool for anyone wishing to approach the topic of loanwords from a lexical, historical, anthropological, semantic, or semantic-cognitive perspective.

GIAN CLAUDIO BATIC
**Soldiers, coffee and markets**

## References

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

BALDI, Sergio (2021). *Dictionary of Arabic loanwords in the languages of Central and East Africa*, Leiden & Boston: Brill.

BENTLEY, W. Holman (1887). *Dictionary and grammar of the Kongo language, as spoken at San Salvador, the ancient capital of the old Congo empire, West Africa*. London: Baptist Missionary Society.

BENTLEY, W. Holman (1895). *Appendix to the dictionary and grammar of the Kongo language, as spoken at San Salvador, the ancient capital of the old Congo empir*e, West Africa, London: Baptist Missionary Society.

HAMMARSTRÖM, Harald and Robert FORKEL, Martin HASPELMATH, Martin and Sebastian BANK (2021). *Glottolog 4.5*, Leipzig: Max Planck Institute for Evolutionary Anthropology, <https://doi.org/10.5281/zenodo.5772642>, last access on 1/3/2022, <http://glottolog.org>.

HULSTAERT, Gustaaf (1952). *Dictionnaire lɔmɔ́ngo-français (lonkundo)*, Tervuren: Annales du Musée Royal du Congo Belge.

KABUTA, Ngo Semzara (1998). "Loanwords in Cilubà", *Lexikos* 8: 37-64.