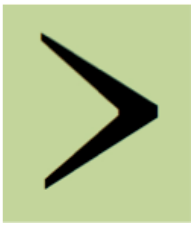
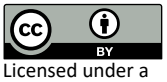


Papers in Historical Phonology
http://journals.ed.ac.uk/pihph
ISSN 2399-6714
Volume 7, 1–43
DOI: 10.2218/pihph.7.2022.7328





# Bangime: Secret language, language isolate, or language island? A computer-assisted case study

Abbie Hantgan
*CNRS-LLACAN*

Johann-Mattis List
*MPI-EVA, Leipzig*

**Abstract**

We report the results of a qualitative and quantitative lexical comparison between Bangime and neighboring languages. Our results indicate that the status of the language as an isolate remains viable, and that Bangime speakers have had different levels of language contact with other Malian populations—particularly Dogon and Mande-speaking peoples—at various points throughout their history. Bangime speakers, the Bangande, claim Dogon ancestry. The Bangande portray this connection to Dogon through borrowings from neighboring Dogon varieties and more rooted vocabulary from Dogon languages spoken to the east from whence the Bangande claim to have come. Evidence of multilayered long-term contact is clear: lexical items have permeated even core vocabulary. We hope that our findings will influence future studies on the reconstruction of the Dogon languages and other neighboring language varieties to shed light on the mysterious history of Bangime and its speakers.

## 1 Introduction

Bangime is a isolate spoken by an estimated 1,500 people in seven villages, situated in a cove at the western edge of a rocky cliff range known as the Bandiagara Escarpment in central-eastern Mali. *Bangime* translates as 'secret language' in many of the Dogon languages spoken along the eastern edges of the cliff range. Bangime speakers (correspondingly depicted as *Bangande*) identify themselves and their language as Dogon; however, their language is not in fact intelligible with any Dogon variety. For example, they are unaware that the terms used for their group and language are exonyms with the meaning of 'secret'. While the reason Bangande essentially do not have an endonym remains part of their mystery, their unfamiliarity with the lexical root [bang], meaning 'secret, hidden, furtive' is

due to the fact that the Dogon speakers in their immediate vicinity do not use the same lexical root, which alone suggests that Bangime-speakers once co-existed with Dogon-speaking populations who differ from those near them now.

This paper employs qualitative and quantitative methods for historical lexical comparison to explore the relationships between speakers of Bangime and the surrounding Dogon languages. In addition, we compare portions of both the Dogon and Bangime lexica with those from neighboring Mande, Atlantic, and Songhai languages in order to rule out any genetic affiliation between Bangime and languages beyond the Dogon escarpment. Placing Bangime in the context of other languages in Mali further supports its status as a linguistic isolate: not only is it disparate from the Dogon varieties, it also shares few, if any, features with surrounding languages—Fulfulde, Songhai, or those of the Mande group. However, our findings support contact with all of them at different stages in history.

This paper was born out of the Workshop on Linguistic Islands in Africa which brought researchers working on languages whose speakers were displaced from their linguistic homeland. Bangime, as will be argued as follows, could be considered to be a language island; it is also classified as a language isolate and deemed to be a secret language by those who do not understand it. This paper seeks to explore the different categorizations of the language through a comparison between Bangime and the languages that surround it.

Our quantitative studies are based on state-of-the-art methods for automatic word comparison across various languages (List et al. 2017). We propose a combination of conservative approaches that closely model the traditional comparative method (Ross & Durie 1996) while taking regular sound correspondences across all languages in the dataset into account, as in the LexStat method (List 2012a), and using simpler approaches that mainly pick up surface similarities between words, as in the Sound-Class-Based Alignment method (List 2012b). By combining these approaches and comparing their results directly, we can automatically identify potential layers of contact between Bangime and its neighboring varieties that reflect different contact periods. These layers can then be further analyzed qualitatively by comparing the automatic findings in detail.

In this study, we limit our analyses to lexical similarities but note that many grammatical features in Bangime, discussed in detail by Heath & Hantgan (2018), are also distinct from those found at least among Niger-Congo and Nilo-Saharan families. By investigating the socio-linguistic,

cultural, and historic implications of the findings presented here, we hope to suggest avenues for future research.

## 2 Background

### 2.1 Previous Research on Bangime

Bangime is considered to be one of only four undisputed West African isolates (Blench 2017: 167). Yet, this does not mean the door is closed on the debate. As Campbell (2016, 2017) advises, a thorough investigation is required before a language can be confirmed as an isolate. Bangime was once classified as Dogon (Gordon 2005, Williamson & Blench 2000); however, specialists of Dogon languages have long recognized the Bangime speech community as an outlier (Calame-Griaule 1956, Hochstetler et al. 2004, Plungian & Tembine 1994). While Bertho (1953: 413–414) went so far as to state that Bangime was distinct, not only from Dogon, but also from the nearby Fulfulde (Atlantic-Congo) and Bozo (Mande), Blench (2005) was the first researcher to suggest it was an isolate. According to the Dogon and Bangime Linguistics Project (https://dogonlanguages.org, Moran et al. 2016), within the variation attested among the now estimated 22 distinct Dogon languages, the lowest limit for mutual intelligibility based on lexical estimates is 32 percent (Prokhorov et al. 2012). In contrast, we now estimate that Bangime shares less than 20 percent of its core vocabulary with Dogon, and even these few lexical items, such as numerals, were likely borrowed long-ago from Dogon languages for the sake of identity-inclusion.

Bangime has been called by many other names in the literature. Among those discussed by Hantgan (2013: 5), one remained a mystery until further data from the Dogon languages were revealed. Blench (2005: 1) mentions an "intrusive -ri-" as it appears in /báŋeri mé/ (Calame-Griaule 1956: viii). Table 1 shows, in Dogon words for 'hide, conceal', the [-ri-] suffix and its allomorphs represent the Dogon causative, or transitive, morpheme.

In Bangime, the suffix [-mɛ] or [-jɛ] denotes a language from the name of the speakers (Hantgan 2013: 112), and yet [-jɛ] also corresponds with the medio-passive suffix among the Dogon languages. Thus, the word *bangime* could be seen as a mix between the Dogon root [bang] and the Bangime suffix [-mɛ], or simply an alternate (and often attested) pronunciation of the medio-passive form of the verb [bang-i-jɛ], meaning 'it is hidden' among the Dogon languages with this root.

| Language | IPA | Language | IPA |
|---|---|---|---|
| Ben Tey | bàŋgì-rí | Bunoge | jɔ́gɛ̀ |
| Gourou | bàŋà-ɾ̃á | Tiranige Diga | dʒíná-ŋgó |
| Jamsay | bàŋà-ɾ̃á | Mombo | dábú-rè |
| Tebul Ure | bàŋgì-rí | Penange | kúj-rɛ̀ |
| Togo Kan | bàŋú-ɾ̃ù | Donno-So | dʒɔ̀ɔ̀-rɔ́ |
| Tommo So | bàŋú-ndá | Nanga | dǎw-rí |
| Yanda Dom | báá-ndɛ́ | Najamba | síbí-r |
| Yorno So | bàŋá-rá | Toro Tegu | sútù |
| Toro-so | bàŋì-rí | Perge Tegu | súgú-ró |

**Table 1:** Dogon terms for concept 'hide, conceal something'

In either case, it is notable that the languages in the left columns of Table 1—those with forms that most closely resemble Calame-Griaule's term /báŋeri mé/—are those currently spoken furthest away from Bangime (Figure 1). Those in the right columns of the table, however, are Bangime's current closest neighbors. It is therefore likely that the name of the language and speakers was given at a time when and in a place where the ancestors of the Bangande and the now eastern Dogon were in contact.

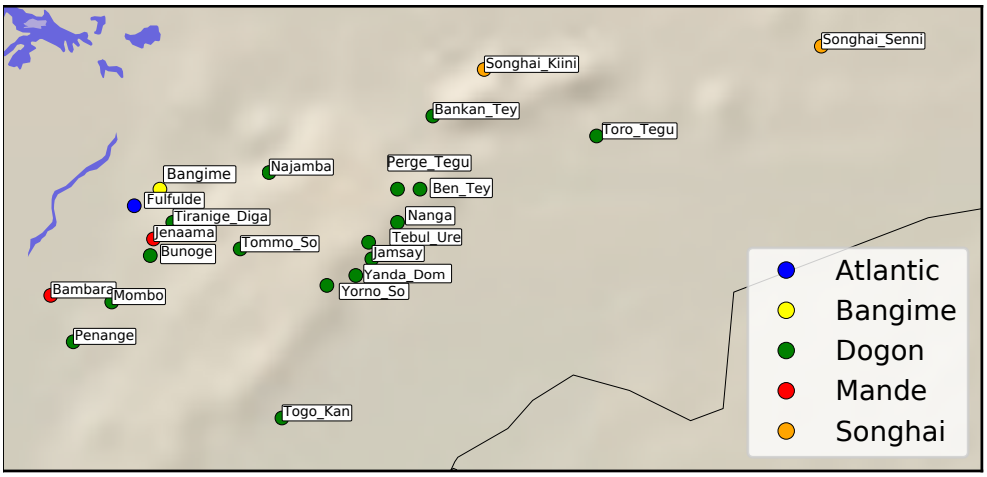



**Figure 1:** Bangime and its surrounding languages.

The map in Figure 1 shows where the other languages which influence Bangime are spoken.[1] Geographically, Mali encompasses three regions—Sahara, Sahel, and sub-Sahel—each with its own heterogeneous and independent ethnological-linguistic societies. Speakers of languages belonging to the Songhai family and the Tuareg branch of Afro-Asiatic live in the north; speakers of Dogon languages, the Fulani branch of Atlantic-Congo, and the Bozo-Soninke branch of Mande in the center; and speakers of Bambara (Mande) live primarily in the south and west, but some have migrated to other regions. The only genealogical overlap among these groups is between Bozo-Soninke and Bambara, both of which are Mande. Even with all this diversity, Bangime stands out as a linguistic isolate. Moreover, genetic studies demonstrate that its speakers represent a genetic deme (Babiker et al. 2018).

Bangime is spoken in seven villages at the end of the Gueou valley, which runs east perpendicular to the Niger river, directly into the central-western edge of the Bandiagara Escarpment. Two villages sit atop the mountain range, and the rest reside just below on the sandy soil. The valley stretches into the Sahelian plains which, approximately 250 kilometers to the north, emerge at Timbuktu in the Sahara. The Gueou valley's other inhabitants speak a Bozo variety called Jenaama. Most other Bozo languages are spoken along the Niger river.

Bangime speakers are in constant contact with speakers of Fulfulde, Bozo-Jenaama, and Dogon-Tiranige. In terms of wider contact areas, Bangande travel to the weekly markets in the Fulani city Konna (35 kilometers northwest) and the Soninke town Sambere (25 kilometers west), mainly to sell wild-grown and gathered fruits but also tobacco and other cultivated crops. Younger people travel to the regional capital Sevare-Mopti and the country's capital Bamako, where they learn to speak Bambara. Few Bangande travel north and thus rarely speak Songhai, though Songhai has made a historical impact on the Dogon languages, discussed in Section 2.2.

Added to this picture, Fulani herders and their livestock roam throughout the upper reaches of sub-Saharan Africa, building temporary camps—often in the mountains, to prevent their animals from destroying crops planted in the fields on top of the cliff and in the plains. Songhai groups live adjacent to the Dogon, directly south of the fabled city Timbuktu and further east along the single paved road that stretches south from the

[1] Coordinates were given for the ethnolinguistic areas closest to the seven Bangande villages; the data used can be found in the supplemental materials.

Malian capital, Bamako, to Gao. Since Bangime is unintelligible to the speakers of any neighboring languages, most Bangande are fluent in Fulfulde and use it as a language of wider communication with groups other than Fulani.

Although the single term, 'Dogon' implies a linguistic and cultural homogeneity, substantiated by descriptions of a unified Dogon language (Bendor-Samuel et al. 1989) and genetically cohesive people (Tishkoff et al. 2009), even early depictions of the inhabitants of the Bandiagara Escarpment illustrate their diversity and heterogeneity (Desplagnes 1907). According to the Dogon and Bangime Linguistics Project (Moran et al. 2016), Dogon as a group constitutes at least 22 separate languages and upwards of 60 dialects. Dogon languages are primarily spoken among villages located at varying levels of the rocky cliffs that make up the Bandiagara Escarpment. Some Dogon villages have moved to the plains within the past decade, and a few Dogon languages are spoken across the border with Burkina Faso. Each secluded cliff village speaks its own variety of Dogon, plus a reduced and often mixed Dogon variety, such as dialects of Tommo So or Jamsay, for the purposes of wider communication.

## 2.2 Historical Context

The Dogon were originally thought to have moved to the Bandiagara Escarpment from its eastern edges some 700–500 years ago. The ~500-year estimate was proposed by Griaule (1938: 28), who claimed to have uncovered nine Sigui masks in the village Ibi in 1933. The Sigui is an eastern-Dogon ritual that is performed every 60 years; as only one mask made for each festival, Griaule stated that the first mask was made in 1430 A.D. Griaule's work is largely disparaged today. Even putting the methodological concerns with this work aside, note that not all Dogon villages have a tradition of mask dancing. Bedaux (1972: 41) asserts that the Dogon did not originally constitute a single culturally homogeneous group; they may have migrated to the cliffs from various directions.

Early paleo-climatological investigations and archaeological evidence amassed by Macdonald (1997) and Bedaux (1972) place the date of Dogon occupation of the Bandiagara Escarpment from the 16$^{th}$ to 20$^{th}$ centuries. More recent developments estimate their arrival sometime between 1200-1400 AD (Mayor & Huysecom 2016, Mayor et al. 2005). The scenario presented by Mayor et al. (2005) depicts the Dogon fleeing the wider area's invaders sometime between 1230 and 1430 AD. In particular, there is linguistic evidence of the effect of the Songhai Empire

(c. 1460–c. 1591) on areas currently occupied by Dogon speakers. For example, most of the Najamba-speaking villages on the north-western portions of the Escarpment are still known by outsiders only by their Songhai names. Some include Koira Beiri [kòjrà béérì] 'big town', Ibissa [í-bìs-â] 'I have passed', Borko [bòrkìn] 'noble', and Tondifere [tóndì fèrè] 'stone brick'. However, contact was not unidirectional—for example, the Songhai village Kikara [kì-kárá], adjacent to where Dogon language Bankan Tey is spoken, translates to 'armpit' in Jamsay and many other Dogon varieties.

Dogon oral traditions describe a migration from Mande (Dieterlen 1955, Izard 1970, Marchal 1978), thought to be located in somewhere in the west of Mali or in present-day Guinea (Vydrin 2009), to a village called Kani-Gogouna, adjacent to Ibi, the village where Griaule found the Sigui masks. The Bangande also recount that they moved from Mande with the Dogon, settled in the village of Kani-Gogouna, and then relocated northwest, over and down the escarpment, to their current villages. However, a more likely scenario based on recent fieldwork (Hantgan 2013: 407-410) proposes that instead of traveling south-east and then over the arduous cliff range, the move went in a straight line east, through a series of villages located along the Gueou valley. Since the Bangande also recount many stories of war with the Jenaama-Bozo, they were probably driven back down the valley to finally settle in the current location at the edge of the escarpment.

According to a widely-known Malian legend, the Jenaama-Bozo, along with the other groups subsumed under the larger Mande branch, are said to have traveled with the Dogon from Mande but split due to a disagreement between brothers. Indeed, the name Dogon, pronounced by its speakers as [dɔ̀gɔ̀-nɔ́] or [dɔ̀gɔ̆-n] derives from the term [dɔ́gɔ́-nín] which means 'little brother' in many Mande languages. One of the consequences of the split between the brothers was a law forbidding Bozo and Dogon to intermarry. The Bozo ethnicity is essentially defined by its trade as fishermen. The Jenaama who live in the Gueou valley are technically outsiders: they do not fish but live off subsistence millet farming like the Bangande and Dogon, yet neither Bangande nor Dogon intermarry with them.

High above Bara, one of the Bangime-speaking villages, is a group of buildings carved directly into the rock face. Found throughout the escarpment, these constructions within the caves are currently used by Bangande and Dogon as both burial grounds and granaries for storing millet. The local populations believe they were housing built by a group known in most eastern Dogon languages as Tellem [tɛ́lɛ̀-m], which is translated

as 'those before' (Mayor et al. 2014).[2] First identified by archaeological excavations by Bedaux (1972), the Tellem are said to have occupied the cliffs from the 11$^{th}$ to 16$^{th}$ centuries AD. More recent work by Mayor & Huysecom (2016) depicts pre-Dogon populations as a cultural "mosaic" but with a distinct change in material cultural practices occurring around 750 years ago. They speculate that the populations who inhabited the escarpment prior to the Dogon either died or were subsumed into a new culture, that of the present-day Dogon, around 500–700 years ago.

Our results, given in Section 4, support evidence for the supposition that the Bangande were among the pre-Dogon groups; that they occupied the cliffs before the Dogon arrived. Further, we propose that there were separate waves of Dogon settlement, with indications of recent contact between Bangime-speakers and their immediate neighbors, and yet longer-term contact with Dogon groups that live to east of their current location. While we do not provide proof of certain dates, we suggest that the different Dogon waves occurred at different times in history, and perhaps from different directions. We explore these findings in the sections below.

## 3 Materials and Methods

All of the data were gathered independent of this study. The primary data were obtained through lexical elicitation for a Dogon-comparative wordlist (Heath et al. 2015), with the addition of data collected for Bangime (Heath & Hantgan 2018) and Jenaama (Heath 2016), as well as Mande language data from the RefLex database (Segerer & Flavier 2016). Fulfulde lexical data are drawn from Osborn et al. (1993), and Humburi Senni Songhai and Tondi Songwai Kiini data (abbreviated here as Sinni and Kiini) from Heath (2005, 2015), respectively.

The Dogon language data were gathered by a team of eight researchers as part of the US National Science Foundation (NSF)-funded Dogon and Bangime Linguistics Project. Lexical, grammatical, and geographic information about the Dogon languages may be found at The Dogon and Bangime Linguistics website (https://dogonlanguages.org, Moran et al. 2016). Language identifiers, classification hierarchies, and geographic data for the remainder of the languages used in the sample are drawn

[2] While the suffix [-m] can be assumed to be the animate noun class morpheme, used in Ben Tey and Bankan Tey, only the root [télɛ́] is listed in the Dogon comparative wordlist, with the meaning of 'almost'. Thus, we cannot confirm the translation of 'those before' however, we do note the similarity to the Mande root [téli] meaning 'quickly'.

from Ethnologue (https://www.ethnologue.com/, Simons & Fennig 2018) and Glottolog (https://glottolog.org, Hammarström et al. 2021).

### 3.1 Lexical Data Preparation

Given the diversity of the original sources, we had to normalize the data in various ways to render them comparable. Our first step was to convert each lexical spreadsheet to a file that could be read and interpreted by LingPy (List & Forkel 2022b), a Python library for automatic tasks in historical linguistics (https://lingpy.org), EDICTOR (List 2017), a web-based tool for creating, editing, and inspecting etymological datasets (https://digling.org/edictor/), and similar packages that are part of recently proposed tool chains for computer-assisted language comparison (Wu et al. 2020), developed in close collaboration with the Cross-Linguistic Data Formats (CLDF) initiative (https://cldf.clld.org, Forkel et al. 2018) and following the data curation principles developed in the Lexibank project (List, Forkel et al. 2022).

The basic formats have been discussed in detail in previous studies (List et al. 2018). They represent language data in a tabular form in which the first row serves as the header, and the consecutive rows describe the characteristics of individual words in the data. Each word (each row) is defined by a unique identifier (listed in the column *ID*), a language (listed in the column *DOCULECT*), a concept (*CONCEPT*), and the word form (in the form of a transcription, usually called *FORM*). While identifier, concept, language, and form are basic values *needed* to represent individual word forms that serve as translational equivalents in individual languages, the tabular format can be flexibly extended by adding additional columns that allow to attribute additional information to individual word forms.

In order to aggregate data collected from different sources, we had to deal with the multitude of glosses used to elicit similar meanings in different contexts. In order to do so, we used the tools provided by the Concepticon project (https://concepticon.clld.org, List et al. 2016, List, Rzymski et al. 2022) to systematically identify equivalent concepts across all datasets by linking all elicitation glosses to Concepticon Concept Sets. The linking procedure is semi-automatic: in the first pass, we used the Concepticon API to link all elicitation glosses automatically to one or more Concepticon concept sets (e.g., linking 'arm (of hand)' and 'arm (body)' to #1673 ARM). In a further step, we manually checked and corrected all automatically produced links.

The automated cognate detection methods included with the LingPy software package require the transcriptions to follow the standards of the

International Phonetic Alphabet (IPA 1999). Confusion can arise when this is not the case. For example, the Africanist tradition of deviates from the IPA in writing [j] as *y*, and [dʒ] as *j*. This means that the Africanist tradition will lead to a false representation of the glide [j] as a front vowel, and the affricate [dʒ] as a glide. Since both the distinction of vowels and consonants and glides and affricates play a crucial role for the automated cognate detection methods we employ, we had to modify the original transcriptions in order to make sure that we receive the best possible output of the computational analysis.

In order to harmonize the transcriptions, we used orthography profiles (Moran & Cysouw 2018), as implemented in the Segments package (Forkel et al. 2019). Orthography profiles provide help in two tasks. On the one hand, they segment transcriptions into units representing single sounds (that may well consist of more than one single character). On the other hand, they can convert these individual units into plain IPA transcriptions.

For the transcriptions, we followed the Broad IPA standard proposed by the Cross-Linguistic Transcription Systems (CLTS) reference catalog, (Anderson et al. 2018, List et al. 2021, https://clts.clld.org). CLTS is accepted by the cognate detection methods provided by LingPy. Table 2 gives an example of the conversion process: many of the Mande words listed in the RefLex database use an underscore tilde [~] to represent nasalization, those of the Dogon comparative spreadsheet use a superscript [$^{n}$] following the segment, and the Bangime data were transcribed with a tilde above the segment [~]. These differing orthographic transcriptions are regularized in the third column of the table.

| Language | Source Transcription | Tokenized Representation | Source Gloss | Concepticon Gloss |
|---|---|---|---|---|
| Bangime | k$\tilde{\tilde{\text{ii}}}$ | k ĩː $^{53}$ ~ | boat | BOAT |
| Mombo | kíː$^{n}$ | k ĩː $^{5}$ ~ | boat | BOAT |
| Tiranige Diga | kŭː$^{n}$ | k ũː $^{15}$ ~ | boat | BOAT |
| Jenaama | kū$^{n}$ | k ũ $^{3}$ ~ | boat | BOAT |
| Bambara | kúrṵ | k u $^{5}$ r u $^{5}$ ~ | boat (skiff) | BOAT |

**Table 2:** Illustration of data conversion in our workflow.

Diacritics and tonal markings on vowels and sonorants were converted to a number corresponding to the 5-tonal step pattern used for tone representation in Chinese and other South-East Asian languages, followed by a tilde in the cases of nasalized segments. Once properly defined, orthography profiles were used to automatically tokenize unsegmented transcriptions in the varying orthographies and convert them to our desired target transcription system. The tokenized representation is indicated by

| List | Source | Coverage |
|---|---|---|
| Blust-2008-210 | Greenhill et al. 2008 | 125 (58%) |
| Gregersen-1976-217 | Gregersen 1976 | 122 (56%) |
| Matisoff-1978-200 | Matisoff 1978 | 118 (59%) |
| Swadesh-1955-100 | Swadesh 1955 | 72 (72%) |
| Swadesh-1952-200 | Swadesh 1952 | 116 (58%) |
| Tadmor-2009-100 | Tadmor 2009, (Leipzig-Jakarta) | 69 (69%) |

**Table 3:** Overlap with popular basic vocabulary lists.

inserting spaces between graphemes (consisting of one or more characters) to indicate the start of a new unit that could phonetically or phonologically be perceived as a sound on its own.

After we created compatible transcriptions and translations, we combined our various wordlists into one multilingual comparative file that essentially contains a (new) unique identifier, a language name without spaces or special characters, both the original IPA transcription and the tokenized form produced by our orthography profile, original glosses, and concept identifiers from the Concepticon concept sets. This resulted in a wordlist consisting of 315 concepts translated into 38 of the estimated 68 Malian languages (Simons & Fennig 2018). To make sure that languages were equally represented, showing high mutual coverage and as few missing translations per concept set as possible, we further extracted a selection of 22 languages and 300 concepts (see the Appendix for the entire list of 300 concepts). Table 3 illustrates the overlap of our 300-item concept list with other well-known concept lists. The overlap with the larger lists is above 50 percent and even higher with the smaller lists.

Our initial selection procedure aimed to achieve a good mix of cultural and cross-linguistically interesting concepts, while selecting language data that would elucidate the ancestry of Bangime, with the overall goal of providing a balanced sample of high coverage and interesting concepts.

## 3.2 Lexical Data Comparison

### 3.2.1 Identifying Layers of Contact

Using automatic methods for cognate detection on languages that we know are unrelated such as Bangime and its neighbors, can produce many false positives that reflect neither recent contact nor ancient relations. To address this problem, we propose a new workflow for

automatic word comparison inspired by general approaches to exploratory data analysis (Morrison 2014). Our main idea is not to restrict our analysis to one method only but to take advantage of the methods LingPy offers for automatic cognate detection, which are quite different in their underlying basic models and strategies.

The LexStat-Infomap approach (List et al. 2017) has been shown to outperform earlier approaches, coming quite close to expert judgments on cognates in multilingual wordlists.[3] Its strategy closely resembles the classical comparative method, searching the data for *regular sound correspondences* before assigning words to common cognate sets. Regular sound correspondences usually reflect deep genetic signals, but can also result from intensive language contact. That is, the family of LexStat approaches tends to single out sporadic borrowings to some degree (List 2012a), but borrowings can be easily confused with cognates where language contact is intensive (List 2014). We expect that applying the LexStat-Infomap approach will propose cognate sets that reflect (a) a true genetic signal among closely related languages, and (b) ancient layers of contact intense enough to surface (potentially weakly) as regular sound correspondences.

In contrast to the conservative and highly sophisticated LexStat-Infomap approach, LingPy also offers simpler methods that are especially useful for quick data exploration, especially when the number of languages and concepts is large. The Sound-Class-Based Alignment (SCA) approach (List 2012b) derives pairwise word similarity scores from pairwise phonetic alignments between all word pairs in a given concept slot without taking regular sound correspondences into account. As a result, it may select true genetic cognates, or it may assign to the same cognate sets words that are similar only due to spurious borrowings or coincidental similarities. We can use this seeming disadvantage to our advantage when dealing with complex linguistic situations like the one we encounter with Bangime.

By comparing the findings for the conservative, yet rather accurate LexStat-Infomap approach with the SCA method, we can systematically search for discrepancies between our *genotypic* and *phenotypic* (Lass 1997) cognate detection approaches. As a rule of thumb, we can say that when both algorithms identify certain words as cognates, they generally

[3] The method by Jäger et al. (2017) seems to outperform LexStat-Infomap on certain datasets, but according to Rama et al. (2018), the difference is minimal and may be in favor of LexStat.

are, notwithstanding certain erroneous judgments that arise for several reasons. We expect to find most of these cases in languages already known to be related. However, if the SCA method identifies certain words as cognate, and LexStat-Infomap does not, we may assume we are dealing with either chance similarities or rather recent instances of borrowing.[4] The more cases of borrowing we find in specific language pairs, the stronger the argument for recent borrowing.

Although our approach is relatively simple, we think it offers several improvements over previously proposed approaches to the automatic identification of borrowings. In contrast to Mennecier et al. (2016) and van der Ark et al. (2007), for example, who use a version of the edit distance (Levenshtein 1965) to search for borrowings between unrelated language varieties only, or phylogeny-based approaches (List 2015, List et al. 2014) that can only be applied to related languages, our approach can be applied to both related and unrelated languages. Since we generally assume that LexStat-Infomap is sufficient to detect very clear cases of cognates, we are also confident that cognates it does not accept but SCA does are probably true instances of borrowings.

Note that we do not expect the cognate sets proposed by LexStat-Infomap for the unrelated languages given in 4 to reflect true, deep cognates. Given the spuriousness of these findings, LexStat-Infomap is probably capturing ancient layers of contact, which is all the more interesting given the unknown history of Bangime. Hence, our approach offers an automatic *stratification* analysis (Lee & Sagart 2008): by applying methods with different degrees of conservatism, we can extract different layers of shared words. The ones LexStat-Infomap recognizes represent the oldest layer, and the ones identified only by SCA represent more recent layers.

### 3.2.2 Shared Vocabulary Statistics

We wanted to know the degree to which the languages of different genetic origins in our sample shared words, especially with respect to Bangime. We viewed the problem as similar to admixture analyses in population genetics (Pritchard et al. 2000), although analyses in biology automatically determine which genes are most likely to represent a certain ancestral population. We are in a much more comfortable position; thanks to

[4] As our dataset shows, in recent borrowings, donor and recipient word tend to resemble each other more than older borrowings do since borrowings are usually nativized over time and adjusted to the phonotactic system of the target language.

classical approaches to linguistic reconstruction, we often have an independent account on the words that were used in a given proto-language, so we do not need sophisticated algorithms to determine which words in our Bangime sample are shared with other language families. Instead, we can simply consider all inferred cognate sets in our data for a given language or group of languages and count how many are shared exclusively between the given language or group and the other languages and groups in the sample. Again, by contrasting LexStat-Infomap and SCA results when calculating the inferred cognate set statistics, we can better assess the degree to which the 'admixture' of a given language or group of languages differs.

### 3.2.3 Implementation

Our approach is implemented in the form of Python scripts, which are available in the supplementary material. They use the parameters indicated in List et al. (2017) for the LexStat-Infomap and the SCA approach to cognate detection. The output data are given in the supplemental materials[5] in tabular form for manual inspection (e.g., by using the EDICTOR tool) or as matrices that can be fed to phylogenetic software like SplitsTree (Huson 1998) to compute splits networks (e.g., with the NeighborNet algorithm, Bryant & Moulton 2004) and as plots that visualize specific aspects of the data.

## 4 Results

### 4.1 Wordlist Statistics

Table 4 summarizes the coverage statistics for each of the 22 languages selected for our study. Sub-groupings for the Mande and Songhai languages are based on Glottolog (Hammarström et al. 2021). The Dogon grouping sub-classification follows Moran & Prokić (2013) and Prokhorov et al. (2012), except that our findings, indicated by the asterisk in the table and discussed below, align Najamba with the western rather than the eastern group.

We note that, based on the genetic relationship to Jenaama, a variety of Soninke might have been a more appropriate choice for an additional Mande language spoken in the area of Bangime, but the available Soninke

[5] All of the raw data, graphs, and the code used to produce our results can be found in the following Github repository: https://github.com/lingpy/language-island-paper.

| Language ID | Language name | Items | Coverage | Subgroup | Source |
|---|---|---|---|---|---|
| 1 | Bambara | 212 | 0.71 | Western Mande | Dumestre 2011 |
| 2 | Bangime | 300 | 1.00 | Isolate | Hantgan & Heath 2016 |
| 3 | Bankan Tey | 297 | 0.99 | Eastern Dogon | Heath et al. 2015 |
| 4 | Ben Tey | 276 | 0.92 | Eastern Dogon | Heath et al. 2015 |
| 5 | Bunoge | 272 | 0.91 | Western Dogon | Heath et al. 2015 |
| 6 | Fulfulde | 282 | 0.94 | Northern Atlantic | Osborn et al. 1993 |
| 7 | Jamsay | 286 | 0.95 | Eastern Dogon | Heath et al. 2015 |
| 8 | Jenaama | 237 | 0.79 | North-western Mande | Heath 2016 |
| 9 | Mombo | 292 | 0.97 | Western Dogon | Heath et al. 2015 |
| 10 | Najamba | 293 | 0.98 | *Eastern Dogon | Heath et al. 2015 |
| 11 | Nanga | 299 | 1.00 | Eastern Dogon | Heath et al. 2015 |
| 12 | Penange | 280 | 0.93 | Western Dogon | Heath et al. 2015 |
| 13 | Perge Tegu | 296 | 0.99 | Eastern Dogon | Heath et al. 2015 |
| 14 | Songhai Senni | 267 | 0.89 | Eastern Songhai | Heath 2015 |
| 15 | Songhai Kiini | 249 | 0.83 | Eastern Songhai | Heath 2005 |
| 16 | Tebul Ure | 286 | 0.95 | Eastern Dogon | Heath et al. 2015 |
| 17 | Tiranige Diga | 290 | 0.97 | Western Dogon | Heath et al. 2015 |
| 18 | Togo Kan | 288 | 0.96 | Eastern Dogon | Heath et al. 2015 |
| 19 | Tommo So | 292 | 0.97 | Eastern Dogon | Heath et al. 2015 |
| 20 | Toro Tegu | 297 | 0.99 | Eastern Dogon | Heath et al. 2015 |
| 21 | Yanda Dom | 295 | 0.98 | Eastern Dogon | Heath et al. 2015 |
| 22 | Yorno So | 294 | 0.98 | Eastern Dogon | Heath et al. 2015 |

**Table 4:** Coverage and sources of our data

wordlist did not cover our selected concepts well enough to be included in the sub-sample. Further, Bambara has a somewhat surprising influence on the Bangime lexicon as found in our sample and explored below.

Using the selected languages and concepts with the coverages shown in Table 4, we performed both methods, SCA and LexStat-Infomap, on the condensed 300-item wordlist. Both models require thresholds that determine at which level of similarity or distance words are considered to be cognate. For our analysis, we employed the thresholds reported by List et al. (2017), who determined on empirical data of six language families, manually coded for cognates by experts, which thresholds yield the best results on average. This yielded a threshold of 0.55 for the LexStat-Infomap approach and 0.45 for the SCA approach.

## 4.2 Shared Similarities

Bangime is undoubtedly a language isolate, but it is not an insular language; we expect to see clear effects of contact with its neighbors. The most likely impact is from the Dogon languages, yet questions remain about when and which varieties left their mark. The LexStat-Info method shows us the deeper levels of contact. We can then compare these findings to the surface levels SCA brings out.

### 4.2.1 LexStat

The dark blue patterning along the rows adjacent to Bangime in the heat map (Fig. 2) indicates that the LexStat method finds practically no notable cognates between Bangime and the other languages in the sample. We still see more similarities between Bangime and Dogon than between Dogon and Fulfulde or Songhai. That is, while Bangime is certainly not genetically related to any of the surrounding languages, it has notable lexical affiliations, specifically with the Dogon languages. We explore the implications of this finding in detail below.

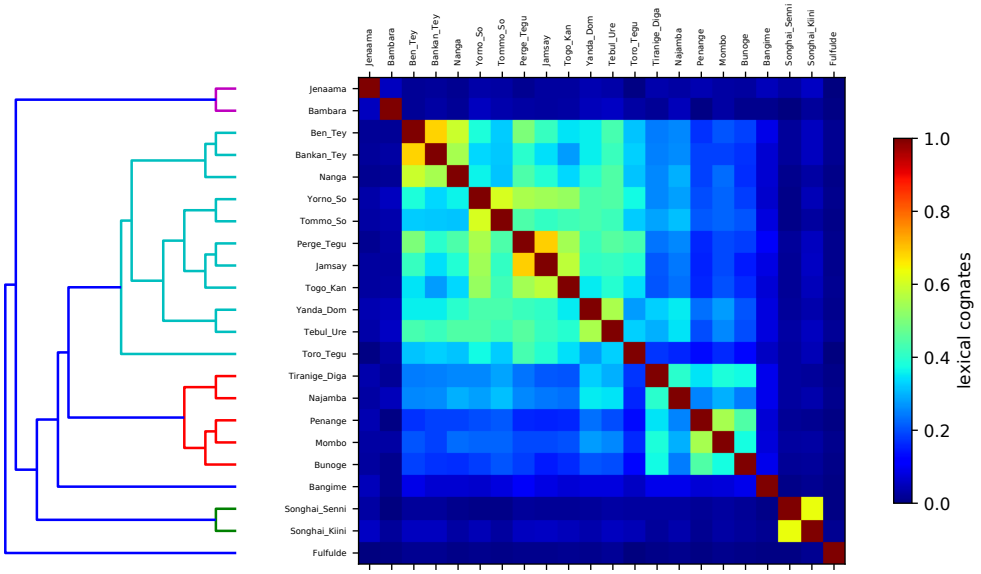



**Figure 2:** Heat Map generated from the cognate percentages inferred by the LexStat-Infomap approach. The cognates chosen by the LexStat-Infomap approach which appear here as lighter colors on the spectrum towards red in the middle, where each language is compared with itself, can be interpreted either as true cognates or deeper levels of contact.

The tree structuring the heatmap was calculated with the help of the UPGMA algorithm (Sokal & Michener 1958) applied to the distance matrix derived from potential cognate words identified by the LexStat-Infomap approach. Since the tree was mainly calculated for the purpose of easing the inspection of shared "cognate" percentages, it should be taken with some care and not used as a statement regarding the detailed phylogenies of some of the language families in our sample. While it captures major subgroups well, it is not suprising that it differs, for example, from the study by Prokhorov et al. (2012) with respect to the internal grouping of the Dogon languages.

While Fulfulde may appear as somewhat of an outlier, we attribute its diversion to the lack of related languages in the sample; Maasina Fulfulde is the only Atlantic, and also non-tonal, language spoken in Mali. In many instances, both computational methods rejected similar words between Fulfulde and the other languages in the sample due to the lack of tones in Fulfulde, even though many borrowings are known to be found among the languages with Fulfulde as the source. Jenaama and Bambara are both considered to be part of the Mande group, albeit distantly related. Data from languages more closely related to Jenaama such as Soninke and others among the Bozo subgroup can be viewed in the larger dataset from which the subset was drawn but due to coverage gaps these languages were omitted for the purposes of the current study.

Figure 3 shows percentages of shared vocabulary as judged by both methods. Examining the specific relationships between Bangime and the other languages in the sample shows that, despite the fact that Bangime is the smallest single language represented, it is the least homogeneous. The graphs confirm our expectations that Bangime has been heavily influenced by both Dogon and Mande languages.

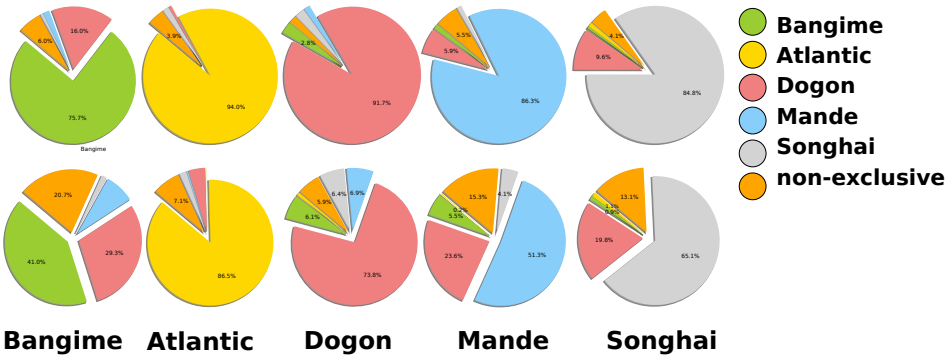



**Figure 3:** Comparing exclusively shared cognates across language families in our sample for LexStat-Infomap (top row) and SCA (bottom row). The "non-exclusive" group refers to cases where inferred cognate sets are attested in more than two families.

The first row of pie charts was produced based on the cognates found by the LexStat-Info method. As discussed in Section 3, the lexical similarities found by LexStat are not necessarily true cognates, though they are likely to reflect long-term contact between languages. We see that even the conservative estimate shows that only 75 percent of Bangime lexical items are unique and the generous estimate posits a mere 41 percent. Not surprisingly, the largest contribution of affinities in the lexicon are from the Dogon languages. Again, although we find no evidence to support a

genetic relationship between Bangime and the other languages used in the sample, shown in the following subsection, the differences between LexStat and SCA results are striking, splitting recent borrowings and ancestral contact into different categories of language contact.

### 4.2.2 SCA

The overall SCA-generated results represented by the heat map in Figure 4 are similar to those shown in the LexStat-generated heat map. Here, they have a lighter shade overall, probably portraying more recent stages of contact through lexical borrowing.

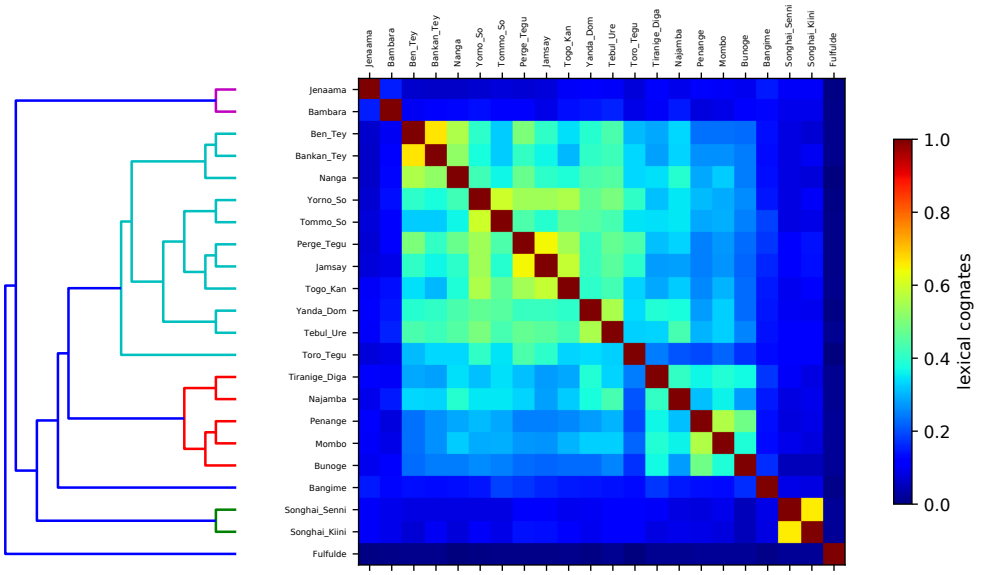



**Figure 4:** Heat Map generated from the cognate percentages inferred by the SCA approach. The "cognates" selected by the SCA approach which appear here as lighter colors on the spectrum towards red in the middle, where each language is compared with itself, can be interpreted as borrowings rather than leading us towards genetic affinities.

Compared with the LexStat-Infomap heat map above, the SCA generated heat map here is far lighter in color, illustrating the tendency of SCA to reflect much more recent stages of language contact. While the Dogon languages' affiliations remain largely the same in both figures, Bangime is more clearly aligned with its neighboring languages such as those in the Dogon western group and the Mande language Jenaama, than was shown in the darker borders of the LexStat-Infomap diagram.

Interestingly, we see that Jenaama and Bambara are still not considered to be as closely related as, for example, Songhai Kiini and

Songhai Senni. This finding is reasonable in that the two Mande groups have neither shared proximate ancestry nor frequent language contact, whereas the two Songhai groups have both. Thus, both of the computational methodologies we used found very similar end-results in terms of language relatedness, with the locus of the largest difference between methods being found in the Bangime results. The congruence of the comparison between the two methodologies, shown in Figure 5 shows that the largest discrepancy with respect to similarities identified by the LexStat-Infomap approach as compared to the SCA approach can be found in Bambara, a Mande language, and Tommo So, a Dogon variety. Whether this suggest a stronger direct contact between the two varieties with Bangime, cannot be answered in this study and would require more research specifically on the internal history of the Dogon languages and the contact history of Dogon, Bangime, and Mande languages. However, it may be interesting to note that many young Bangande now speak Bambara as a result of traveling to the southern part of the country for work and returning, and thus the language is having an inevitable effect on their speech, especially if they depart the village at a young age.

The influence of Tommo So may be explained by three factors: (1) all Dogon and Bangande perform songs in Tommo So, regardless of whether the people speak the language; (2) legend describes the Dogon migration arriving first at the currently Tommo So-inhabited village Kani-Gogouna, so the language may express a type of ancestral lineage; and (3) Tommo So is centrally spoken, and many use it as a lingua-franca, especially in the region surrounding Bandiagara. Although more extensive research is needed, it seems likely that the difference in the conservative LexStat-Infomap approach and the shallower SCA approach reflects some recent contact dynamics among these languages.

Considering the Bangande-projected Dogon identity, we are not entirely surprised to find that Dogon languages have lexically influenced even core Bangime vocabulary—specifically—as we will see in the next section, lower numerals, body parts, and culturally significant items such as farming tools, crops, and words associated with the West African caste system. However, what is somewhat surprising is the geographical distance at which we find evidence of this contact. We explore specifics in the next section.

### 4.3 Areal Comparisons

As explained in Section 3, both the SCA and LexStat methods may indicate borrowings as well as true cognates, but the latter tends to find cognates

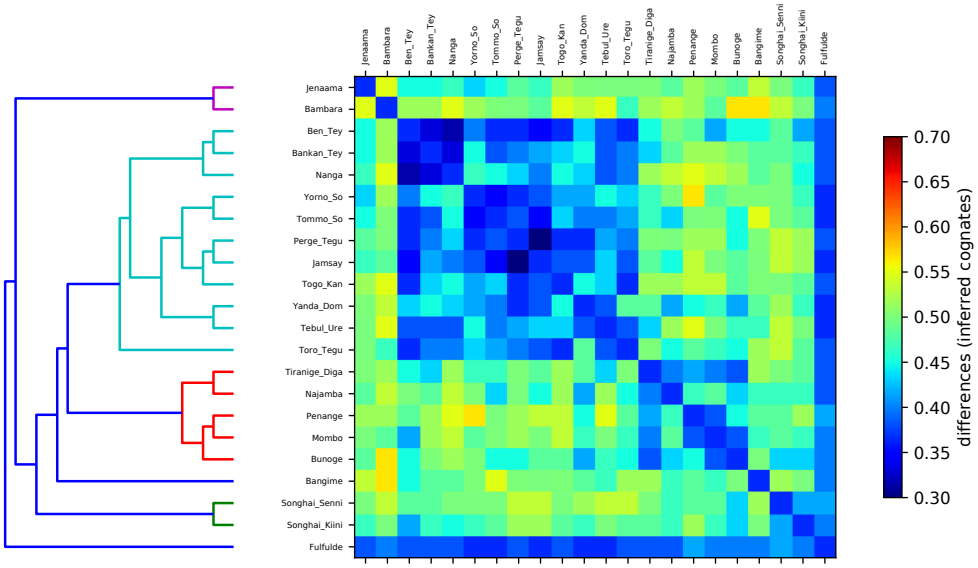



**Figure 5:** Comparing the discrepancies between cognate percentages inferred by LexStat-Infomap and SCA. If no difference is encountered, we set the value in the heatmap to 0.5. If the proportions differ, we first subtract the LexStat-Infomap value from the SCA value and then add this value to our baseline of 0.5. As a result, comparing one language with itself will always yield the baseline, and values larger than 0.5 indicate to which degree SCA infers more cognates for the respective language pair, while values less than 0.5 indicate that SCA infers less cognates. We suspect that those cases in which SCA infers a larger number of more cognates point to recent borrowings between the varieties.

rather than borrowings, and if it identifies borrowings as cognates, these may be substantiated by *layers of contact* which are reflected in the form of regular sound correspondences in the borrowed words. Here, we focus on those instances when both methods select a word as 'cognate' with Bangime. In this way, we can concentrate on long-term contact between the Bangande and surrounding speakers. As Campbell (2017) points out, one way to find the lost ancestors of a language isolate is through areal comparisons within a Sprachbund. Almost all instances of LexStat cognates are shared with the SCA method, though obviously the reverse does not hold true. Since Bangime is an isolate, it does not have cognates per se with surrounding languages, but it certainly shares vocabulary worth exploring in further detail.

First, we illustrate the distribution of numerals Bangime shares with the languages in our sample (Table 5) because we note that this area of the language's core vocabulary greatly resembles languages from different groups spoken in Mali. Check marks indicate where both the SCA

and LexStat-Info method selected the concept and language has having a shared form ('cognate') with Bangime. Check marks in parentheses signal forms that were selected only by the SCA method and thus represent shallower matches that are not substantiated by recurring correspondences. The gray shading highlights the most prevalent languages and concepts. The original word forms for these tables are given in the appendices; otherwise all of the data used for this study are available in the supplemental materials and can be viewed with the help of the EDICTOR tool (List 2017), by loading the file D_subset-300-22.tsv-cognates.tsv, shared as part of our supplemental material.

| DOCULECT | ONE | TWO | THREE | FOUR | FIVE | SIX | SEVEN | EIGHT | NINE | TEN |
|---|---|---|---|---|---|---|---|---|---|---|
| Bangime (Isolate) | ✓ | | ✓ | (✓) | (✓) | ✓ | (✓) | ✓ | (✓) | |
| Bambara (Mande) | | | | (✓) | | | | ✓ | | |
| Bankan Tey (Dogon) | | | ✓ | (✓) | | ✓ | | | (✓) | |
| Ben Tey (Dogon) | | | ✓ | (✓) | | ✓ | | | (✓) | |
| Bunoge (Dogon) | ✓ | | ✓ | (✓) | | ✓ | | (✓) | (✓) | |
| Fulfulde (Atlantic) | | | ✓ | (✓) | | | | | | |
| Jamsay (Dogon) | ✓ | | ✓ | (✓) | | ✓ | | | | |
| Jenaama (Mande) | | | | | | | | ✓ | | |
| Mombo (Dogon) | | | ✓ | | | ✓ | | (✓) | (✓) | |
| Najamba (Dogon) | | | ✓ | | | ✓ | | (✓) | | |
| Nanga (Dogon) | | | ✓ | (✓) | | ✓ | (✓) | | (✓) | |
| Penange (Dogon) | | | ✓ | | | ✓ | | (✓) | (✓) | |
| Perge Tegu (Dogon) | ✓ | | ✓ | (✓) | | ✓ | | | | |
| Songhai Kiini | | | | | | | (✓) | | | |
| Songhai Senni | | | | | | | (✓) | | | |
| Tebul Ure (Dogon) | ✓ | | ✓ | | | ✓ | | | | |
| Tiranige Diga (Dogon) | | | ✓ | | | ✓ | | (✓) | (✓) | |
| Togo Kan (Dogon) | ✓ | | ✓ | | (✓) | ✓ | | (✓) | (✓) | |
| Tommo So (Dogon) | ✓ | | ✓ | (✓) | (✓) | ✓ | | | (✓) | |
| Toro Tegu (Dogon) | ✓ | | ✓ | (✓) | | ✓ | | | | |
| Yanda Dom (Dogon) | | | ✓ | | | ✓ | | ✓ | | |
| Yorno So (Dogon) | ✓ | | ✓ | (✓) | | ✓ | | | (✓) | |

**Table 5:** Shared numerals from one to ten.

Both methods identify matches for the Bangime numerals *tà:rù* THREE and *ké:rè* SIX with a large number of the other varieties, spanning several language families in the region. Bangime's most common source languages are found in the eastern Dogon subgroup; only one western neighboring language, Bunoge, shares four of the ten lower numerals with Bangime, as compared to five eastern Dogon languages. For SEVEN the SCA method identifies a match with the distantly spoken Songhai languages. EIGHT is matched with Mande and the two Dogon varieties Najamba and Yanda Dom by both methods. The only numerals where

the methods do not propose any similarities of numerals in Bangime with the other varieties in our sample are TWO and TEN. That certain numerals could have been borrowed from neighboring varieties is quite possible, given the shared cultural roles which, for example, the numerals ONE through FIVE play for the Bangande and Dogon: the traditional week has five days with a different village's market rotating on one day of the five-day week. This traditional method of keeping track of the days is being lost in favor of the seven day week, now common among all cultures in the region, but the date of many performances and events crucially still relies on the five-day week.

Among other core vocabulary items, we witness, although less often, similarities for certain body parts as shown in Table 6.

| DOCULECT | BEARD | CHEST | JAW | JOINT | LUNG | VAGINA |
|---|---|---|---|---|---|---|
| Bangime (Isolate) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bambara (Mande) | | | | | ✓ | |
| Bankan Tey (Dogon) | | (✓) | | | ✓ | ✓ |
| Ben Tey (Dogon) | | | | | ✓ | (✓) |
| Bunoge (Dogon) | | | (✓) | | ✓ | |
| Fulfulde (Atlantic) | | | | | | |
| Jamsay (Dogon) | | ✓ | ✓ | ✓ | ✓ | (✓) |
| Jenaama (Mande) | ✓ | | | | | |
| Mombo (Dogon) | | (✓) | ✓ | ✓ | ✓ | |
| Najamba (Dogon) | | | | ✓ | ✓ | ✓ |
| Nanga (Dogon) | | | | ✓ | | |
| Penange (Dogon) | | (✓) | | ✓ | | |
| Perge Tegu (Dogon) | | ✓ | | ✓ | ✓ | ✓ |
| Songhai Senni | | ✓ | | | | ✓ |
| Tebul Ure (Dogon) | | | | ✓ | ✓ | |
| Tiranige Diga (Dogon) | | | (✓) | ✓ | | |
| Togo Kan (Dogon) | | | (✓) | | ✓ | |
| Tommo So (Dogon) | | | | | | |
| Toro Tegu (Dogon) | | | | ✓ | | |
| Yanda Dom (Dogon) | | | | ✓ | ✓ | |
| Yorno So (Dogon) | | (✓) | | ✓ | ✓ | |

**Table 6:** Shared body parts in the data.

Blench (2005: 15) states that the Bangime words for 'ear' and 'mouth' are roots commonly found among Niger-Congo languages, yet neither of our analyses finds similar word forms in our sample across the languages of Mali. Among numerals and body parts, many terms from Bangime are identified as being similar to Jamsay, including the word *gɔ̀ndɛ̀* for the concept CHEST, also found in Songhai. The methods also identify Bangime *dɔ̀kɛ́-rɛ́* JOINT and *pógó pógò* LUNG as shared with most of the

Dogon languages. While the former seems less likely, since most forms identified only coincide in the initial consonant, the latter has a clear counterpart in Bambara *fógó*$^{n}$ *fògó*$^{n}$. In addition, manual inspection revealed some very similar matches between body part terms in Bangime and words with a shifted meaning in Dogon varieties. Examples are shown in Table 7. These items were not identified by the automatic approaches, since these only compare words with the same meaning.

| Bangime | Bondu-So (Dogon) |
|---|---|
| SKIN | HEAD |
| kíndʒē | kíŋgè |
| | |
| Bangime | Tiranige Diga (Dogon) |
| HAIR | SKIN |
| kwìì | gwíí |
| | |
| Bangime | Toro-Tego (Dogon) |
| WING (shoulder) | FOOT OR LEG |
| kúwó | kúwó |

**Table 7:** Similar word forms for body parts with shifted meanings.

There is a precedent for switched meanings in Bangime. Hantgan (2013) describes portions of the Bangime lexicon with purposely reversed meanings such as the naming of plants used in traditional medicines by reversed colors than they appear to be. Even though practically no one outside the Bangande community speaks Bangime, they claim that the use of reverse meanings further protects them from potential eavesdropping.

Additional terms identified as shared between Bangime and varieties from the other language families in our sample are shown in Table 8. Words for COW and MEAT, listed in Table 8, along with HUNDRED, GOLD, FORGIVE, and MOSQUE which can be found in the supplemental materials, are likely attributed to the wide distribution Fulfulde lexical items, some of which are originally from Arabic, while CASSAVA, with a likely source from Bambara and other Mande varieties, is a recently introduced crop, selected by our methods as being a shared term.

### 4.4 Group-Specific Comparisons

Words from Mande languages Bambara and Jenaama are shown in Figure 5 and Table 9 to be particularly prevalent in Bangime.

| DOCULECT | CASSAVA | COW | GRASS | HORN | MAIZE | MARROW | MEAT |
|---|---|---|---|---|---|---|---|
| Bangime (Isolate) | ✓ | ✓ | ✓ | (✓) | ✓ | ✓ | ✓ |
| Bambara (Mande) | ✓ | | | | | | |
| Bankan Tey (Dogon) | ✓ | ✓ | | | (✓) | ✓ | ✓ |
| Ben Tey (Dogon) | ✓ | ✓ | | | | | ✓ |
| Bunoge (Dogon) | ✓ | ✓ | ✓ | (✓) | | | ✓ |
| Fulfulde (Atlantic) | ✓ | (✓) | | | | | |
| Jamsay (Dogon) | ✓ | (✓) | | (✓) | ✓ | | ✓ |
| Jenaama (Mande) | ✓ | ✓ | | | | ✓ | |
| Mombo (Dogon) | ✓ | ✓ | | (✓) | | | ✓ |
| Najamba (Dogon) | ✓ | ✓ | | (✓) | | ✓ | ✓ |
| Nanga (Dogon) | | (✓) | | (✓) | | ✓ | ✓ |
| Penange (Dogon) | | ✓ | ✓ | (✓) | | | ✓ |
| Perge Tegu (Dogon) | ✓ | (✓) | | (✓) | (✓) | | ✓ |
| Songhai Kiini | ✓ | | | (✓) | | ✓ | |
| Songhai Senni | ✓ | | | | | ✓ | |
| Tebul Ure (Dogon) | ✓ | ✓ | | (✓) | | | ✓ |
| Tiranige Diga (Dogon) | ✓ | ✓ | | (✓) | | ✓ | ✓ |
| Togo Kan (Dogon) | ✓ | (✓) | | (✓) | | | ✓ |
| Tommo So (Dogon) | ✓ | ✓ | | (✓) | | | ✓ |
| Toro Tegu (Dogon) | ✓ | (✓) | | (✓) | | | ✓ |
| Yanda Dom (Dogon) | ✓ | ✓ | | (✓) | | | ✓ |
| Yorno So (Dogon) | | ✓ | | (✓) | | | ✓ |

**Table 8:** Shared agricultural vocabulary items.

The impact of the Mali Empire on Bangime, with Mande rule of the western portions of Mali from the 1200's to 1600's A.D., seems to be evidenced by shared word forms with specific cultural meanings found in Bangime and the languages spoken to the west of the Bangande-speaking region. Compare, for example, NOBLE *hɔ́rɔ́ⁿ* in Bambara with *hɔ̄ɔ̄rɔ̀* in Bangime and SLAVE *kɔ̀mê* in Bambara with *kɔ̀mɛ̀* in Bangime. Additionally note that while NOBLE is selected by both algorithms as a shared form, SLAVE is rejected by the LexStat method due to a lack of regularity. In the latter case, the form for the concept SLAVE in the Dogon language Tiranige Diga *kɔ́mɛ́* is the only one that is selected by both methods as being shared with Bangime.

While most researchers assume that the Dogon villages came to be occupied as a result of escaping from prior empires, Nunn & Puga (2012) provide statistical support for the unique benefits of living in rocky terrain in Africa in order to avoid the repercussions of the slave trades from 1400 and 1900 A.D. As explained in Hantgan (2013), the Bangande separate themselves into two classes: slave and royal, with depictions of slave raids happening in the villages and surroundings told through oral his-

| DOCULECT | NOBLE | SLAVE | TONGS | FIGHT | FOAM | NEW | LOUSE | WORK |
|---|---|---|---|---|---|---|---|---|
| Bangime (Isolate) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Bambara (Mande) | ✓ | (✓) | | ✓ | | ✓ | | (✓) |
| Bankan Tey (Dogon) | | (✓) | ✓ | | | ✓ | | (✓) |
| Ben Tey (Dogon) | | (✓) | | | | ✓ | | |
| Bunoge (Dogon) | | ✓ | | | ✓ | | ✓ | ✓ |
| Fulfulde (Atlantic) | | | | | | | | |
| Jamsay (Dogon) | | (✓) | | | | ✓ | | (✓) |
| Jenaama (Mande) | ✓ | (✓) | | ✓ | ✓ | | ✓ | |
| Mombo (Dogon) | | (✓) | | | ✓ | ✓ | ✓ | ✓ |
| Najamba (Dogon) | | (✓) | | | | ✓ | | (✓) |
| Nanga (Dogon) | | (✓) | | | | ✓ | | (✓) |
| Penange (Dogon) | | (✓) | | | ✓ | ✓ | ✓ | |
| Perge Tegu (Dogon) | | (✓) | | | | ✓ | | (✓) |
| Songhai Senni | | (✓) | | | | | | |
| Songhai Kiini | | (✓) | | | | | | |
| Tebul Ure (Dogon) | | (✓) | | | | ✓ | | (✓) |
| Tiranige Diga (Dogon) | | ✓ | | | ✓ | ✓ | ✓ | ✓ |
| Togo Kan (Dogon) | ✓ | (✓) | | | | ✓ | | (✓) |
| Tommo So (Dogon) | | (✓) | | | | ✓ | | (✓) |
| Toro Tegu (Dogon) | | (✓) | ✓ | | | ✓ | | (✓) |
| Yanda Dom (Dogon) | | (✓) | | | ✓ | ✓ | ✓ | (✓) |
| Yorno So (Dogon) | | (✓) | | | | ✓ | ✓ | (✓) |

**Table 9:** Bangime vocabulary shared with Dogon, and Mande languages.

tories. At least at the western side of the Bandiagara Escarpment where Bangime is spoken, the results of language contact appear to be recent. Specific concepts within the domain of castes, central to Malian cultural hierarchies, such as NOBLE and SLAVE are likely from Mande languages, passed into Bangime and western Dogon languages, yet not the Dogon languages of the east.

Table 10 shows basic and cultural concepts which are predominantly shared with Eastern Dogon varieties in our sample. As mentioned above in Section 4, throughout the Dogon and Bangande communities, songs are almost always performed in Tommo So, independent of whether or not the person singing speaks or understands the language. It might therefore be possible that Bangime has indeed borrowed the term for SONG, *ŋúɥⁿɛ̂ː* , from Dogon varieties. That we find a very basic and supposedly stable concept like STAR, Bangime *tōrɛ̀*, in this list may look very surprising at first sight, but it may also reflect attempts by the Bangande to further portray their projected Dogon identity. The Dogon were made famous by the assertion of Griaule & Dieterlen (1965) that the Dogon Sigui ritual celebrates the presence of a star that is unable to be viewed without a telescope. Although it is unlikely that Dogon possess the ability to view

| DOCULECT | SONG | STAR | ROOF | GO DOWN | HELP | PUSH | SHOW |
|---|---|---|---|---|---|---|---|
| Bangime (Isolate) | ✓ | ✓ | (✓) | ✓ | (✓) | ✓ | ✓ |
| Bambara (Mande) | | | | (✓) | | | |
| Bankan Tey (Dogon) | ✓ | | (✓) | ✓ | (✓) | ✓ | ✓ |
| Ben Tey (Dogon) | ✓ | ✓ | | ✓ | (✓) | ✓ | ✓ |
| Bunoge (Dogon) | ✓ | | | | (✓) | | (✓) |
| Fulfulde (Atlantic) | | | (✓) | | | | |
| Jamsay (Dogon) | ✓ | ✓ | (✓) | | (✓) | | |
| Jenaama (Mande) | | | | | | | |
| Mombo (Dogon) | (✓) | | | ✓ | | | ✓ |
| Najamba (Dogon) | ✓ | | (✓) | ✓ | (✓) | ✓ | ✓ |
| Nanga (Dogon) | ✓ | | (✓) | ✓ | (✓) | ✓ | ✓ |
| Penange (Dogon) | (✓) | | | ✓ | (✓) | | ✓ |
| Perge Tegu (Dogon) | ✓ | ✓ | (✓) | ✓ | (✓) | ✓ | ✓ |
| Songhai Senni | | | | | | | |
| Songhai Kiini | | | | | | | |
| Tebul Ure (Dogon) | ✓ | | | ✓ | (✓) | ✓ | ✓ |
| Tiranige Diga (Dogon) | (✓) | | ✓ | ✓ | (✓) | ✓ | ✓ |
| Togo Kan (Dogon) | ✓ | ✓ | | ✓ | (✓) | ✓ | ✓ |
| Tommo So (Dogon) | ✓ | | (✓) | ✓ | (✓) | ✓ | ✓ |
| Toro Tegu (Dogon) | ✓ | | | ✓ | (✓) | (✓) | ✓ |
| Yanda Dom (Dogon) | ✓ | | (✓) | ✓ | (✓) | ✓ | ✓ |
| Yorno So (Dogon) | (✓) | | | ✓ | (✓) | ✓ | ✓ |

**Table 10:** Basic and cultural vocabulary items in Bangime shared predominantly with Dogon varieties.

the particular star in question, stars are an integral part of daily Dogon life, providing navigational guidance in the barren plains and details by which they plan events such as rituals and plantings. Not all of the remaining items in the table show convincing similarities between Bangime word forms and word forms in neighboring languages. Thus, ROOF in Bangime (*kùmbè*), for example, seems to be a clear false positive in both approaches, resulting from the wrong mismatch of the initial *kù* in Bangime with similar *kV* as second element in some Dogon languages in our sample. Further and detailed investigation of the automatically identified items will be needed in order gain a deeper understanding of the complex areal processes that shaped the linguistic diversity in Mali and impacted on the observed lexical structure of Bangime.

## 5 Conclusion

Our study seems to confirm that Bangime is a language isolate, yet the effects of both recent and distant contact, crucial to solving the mysteries surrounding and positing a reasonable history of these “secret” people,

are difficult to observe using automatic cognate detection methods alone. As a result, any deeper conclusions drawn from the data would require careful manual inspection of the wordlists informed by additional knowledge of Bangande culture and their own understanding of their ethnological and linguistic identity.

While Bangande consider themselves Dogon, and Dogon consider themselves to be the young brothers of the Mande, our automated analyses could not provide any lexical evidence to suggest that any of these groups are related to each other. More likely, the Bangande escaped whatever caused the annihilation of their ancestors and came to settle in the secluded Geou valley prior to Dogon settlement. As the influence of the western Dogon languages seems relatively more recent than that of the eastern groups, the Dogon peoples seem to have come to the Bandiagara Escarpment from different directions and settled at different times. No dating or archaeological research has been conducted in the area where the Bangande reside, but if they did live in the cliffs before the Dogon, then perhaps they are "those who came before"—the Tellem.

But if the Bangande have been not only in contact, but practically immersed in the surrounding Dogon cultures for at least half a millennium then it is that much more surprising that they have managed to keep their language intact. Samar & Bhatia (2017: 62), point out that '[they] are not sure if there is such a case where two compatible languages have been in contact for more than, say, 500 years, without any of them dying'. If so, then Bangime represents an interesting and vanishingly rare counterexample. Not only is it not dead, it is thriving; around 1,500 speakers use it on a daily basis in their homes and with their children.

Given the neighboring Dogon groups' disdain for the Bangande, we can propose that a reason why Bangande have not mixed genetically with their Dogon neighbors (cf. Babiker et al. 2018) is because Dogon village men do not accept to marry Bangande wives; though the opposite holds true. In traditional West African society, arranged marriages forbid the joining of certain groups to one another, such as between the Bozo and the Dogon. However, given the geographic proximity, language contact is inevitable. Bangime is a "secret language" to its neighbors; linguists classify it as an isolate; and since it is likely currently spoken in a remote location, at a great remove from its original, ancestral speakers, it is a language island. In identifying the words that have washed up on its shores over time, we may someday be able to track its passage, explain its practices, and gain from its knowledge and experiences. Future interdisciplinary research should be pursued, aided by computational resources that can help to decipher the linguistic, genetic, and anthropological clues.

**Acknowledgments**

This research was funded by the ERC Starting Grant 715618 "Computer Assisted Language Comparison" (JML, https://digling.org/calc, 2017–2022). This paper came out of a presentation at the Workshop on Linguistic Islands in Africa which took place in Bad Münster am Stein-Ebernburg from June 14-18, 2017. The study was originally submitted for the inclusion into a special issue of a journal. After having been accepted in 2018, the special issue was not processed for years. As a result, we decided to withdraw the original article and instead share an updated and slightly shortened version of the study again for open peer review with Papers in Historical Phonology. We note that our approaches shared here have in part been superseded by novel techniques by now, notably Hantgan et al. (2022) and List & Forkel (2022a). We hope, however, that the techniques described in this study are still of interest for scholars working on language isolates or languages heavily influenced from language contact.

We are indebted to the organizers of the workshop, Prof. Dr. Rainer Voßen and Dr. Ulrike Zoch, as well as the workshop participants, for making this possible. We would like to express our upmost gratitude to Russell D. Gray for hosting Abbie Hantgan as a visiting researcher at the Max Planck Institute for the Science of Human History in Jena while this study was taking place. We also thank those who were by then members of the Linguistic and Cultural Evolution department and its affiliates, in particular Hiba Babiker, as well as Christoph Rzymski, Simon J. Greenhill, and Robert Forkel. The idea to compare languages in a pairwise fashion was inspired by discussions with Adam Powell, who reported an idea for genetic data, which turned out to be quite different from what we did in the end, but which gave the initial idea to shift the focus in the methodology from cognate sets to cognate sets (or sets of potentially borrowed words) shared among pairs and across families. We also gratefully acknowledge

the continued support of the Dogon and Bangime Linguistics project, particularly PI Jeffrey Heath and Steven Moran whose initial work on comparative Dogon fed the initial stages of this project.

### Associated Material

The supplementary material contains the Python code along with the data that are needed to replicate the analyses discussed in this study along with usage instructions. It has been curated on GitHub at https://github.com/lingpy/language-island-paper and archived with Zenodo at https://doi.org/10.5281/zenodo.5716590 (version 4.0). The CLDF version of the dataset has been curated on GitHub at https://github.com/lexibank/hantganbangime and archived with Zenodo at https://doi.org/10.5281/zenodo.5126441.

### Author contact details

*Abbie Hantgan*
CSPC, INALCO CNRS UMR 8135 LLACAN
Langage, langues et cultures d'Afrique noire
Campus CNRS
Villejuif 94800
France

abbie.hantgan-sonko@cnrs.fr

*Johann-Mattis List*
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
Leipzig, 04103
Germany

mattis_list@eva.mpg.de

## References

Anderson, Cormac, Tiago Tresoldi, Thiago Costa Chacon, Anne-Maria Fehn, Mary Walworth, Robert Forkel & Johann-Mattis List. 2018. A Cross-Linguistic Database of Phonetic Transcription Systems. *Yearbook of the Poznań Linguistic Meeting* 4(1). 21–53.

Babiker, Hiba, Floyd Reed & Jeffrey Heath. 2018. *The genetic identity behind the masks: Bangande and Dogon of Western Africa*. Unpublished manuscript.

Bedaux, R. M. A. 1972. Tellem, reconnaissance archéologique d'une culture de l'Ouest africain au Moyen-Age: Recherches architectoniques. *Journal de la Société des Africanistes* 42(2). 103–185.

Bendor-Samuel, John, Elizabeth Olsen & Anne White. 1989. Dogon. In John Bendor-Samuel (ed.), *The Niger-Congo languages*, 169–177. Lanham MD/New York/London: University Press of America.

Bertho, Jacques. 1953. La place des dialectes dogon (Dogõ) de la falaise de Bandiagara parmi les autres groups linguistiques de la zone soudanaise. *Bulletin de l'Institut Francais d'Afrique Noire* 15(1). 405–441.

Blench, Roger. 2005. *Bangime, a language of unknown affiliation in northern Mali*. online. http://www.rogerblench.info/Language/20data/Isolates/Bangime/20wordlist/20paper.pdf.

Blench, Roger. 2017. African language isolates. In Lyle Campbell (ed.), *Language isolates*, 162–192. London & New York: Routledge.

Bryant, David & V. Moulton. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2). 255–265.

Calame-Griaule, Geneviève. 1956. Les dialectes Dogon. *Africa* 26(1). 62–72.

Campbell, Lyle. 2016. Language isolates and their history, or, what's weird, anyway. *Berkeley Linguistics Society* 36(01). 16–31.

Campbell, Lyle. 2017. Introduction. In Lyle Campbell (ed.), *Language isolates*, 1–18. London & New York: Routledge.

Desplagnes, L. 1907. *Le plateau central Nigerien*. Paris: La Rose.

Dieterlen, Germaine. 1955. Mythes et organisation sociale au Soudan français. *Journal de la Société des Africanistes* 25. 39–76.

Dumestre, Gérard. 2011. *Dictionnaire bambara–français*. Paris: Karthala.

Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5(180205). 1–10.

Forkel, Robert, Steven Moran, Johann-Mattis List, Simon J Greenhill, Lucas Ashby, Kyle Gorman & Gereon Kaiping. 2019. *Segments. Unicode Standard tokenization routines and orthography profile segmentation [Software Library, Version 2.1.3]*. Geneva: Zenodo.

Gordon, Jr., Raymond G. (ed.). 2005. *Ethnologue: Languages of the World, Fifteenth edition*. http://www.ethnologue.com/15.

Greenhill, Simon J., Robert Blust & Russell D. Gray. 2008. The Austronesian Basic Vocabulary Database: From bioinformatics to lexomics. *Evolutionary Bioinformatics* 4. 271–283.

Gregersen, Edgar A. 1976. The glottochronological performance of African languages. *Cahiers de l'Institut de Linguistique de Louvain* 3(5-6). 107–146.

Griaule, Marcel. 1938. *Masques Dogons*. Paris: Institut d'Ethnologie.

Griaule, Marcel & Germaine Dieterlen. 1965. *Le renard pâle*. Paris: Institut d'Ethnologie.

Hammarström, Harald, Martin Haspelmath, Robert Forkel & Sebastiaon Bank. 2021. *Glottolog. Version 4.4*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://glottolog.org.

Hantgan, Abbie. 2013. *Aspects of Bangime phonology, morphology, and morpho-syntax*. Indiana University dissertation.

Hantgan, Abbie, Hiba Babiker & Johann-Mattis List. 2022. First steps towards the detection of contact layers in Bangime: A multi-disciplinary, computer-assisted approach [version 2; peer review: 2 approved]. *Open Research Europe* 2(10). 1–27.

Hantgan, Abbie & Jeffrey Heath. 2016. *Bangime lexicon*. Unpublished wordlist.

Heath, Jeffrey. 2005. *Tondi Songway Kiini (Songhay, Mali): Reference grammar and TSK-English-French dictionary*. Stanford, California: CSLI Publications.

Heath, Jeffrey. 2015. *Dictionary Humburi Senni (Songhay of Hombori, Mali) - English - French*. http://hdl.handle.net/2027.42/117646.

Heath, Jeffrey. 2016. *Jenaama lexicon*. Unpublished wordlist.

Heath, Jeffrey & Abbie Hantgan. 2018. *A grammar of Bangime*. Berlin, Boston: De Gruyter Mouton.

Heath, Jeffrey, Laura McPherson, Kirill Prokhorov & Steven Moran. 2015. Dogon comparative wordlist. http://cdstar.shh.mpg.de/bitstreams/EAEA0-C97A-A1D2-2E76-0/a.xls.

Hochstetler, J., J.A. Lee Durieux & E.I.K. Durieux-Boon. 2004. Sociolinguistic survey of the Dogon language area. *Journal of Language Survey Reports*.

Huson, Daniel H. 1998. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* 14(1). 68–73.

IPA (ed.). 1999. *Handbook of the International Phonetic Association: A guide to the use of the international phonetic alphabet*. Cambridge: Cambridge University Press.

Izard, M. 1970. *Introduction à l'histoire des royaumes mossi*. Paris & Ouagadougou: Eds du CNRS-CVRS.

Jäger, Gerhard, Johann-Mattis List & Pavel Sofroniev. 2017. Using support vector machines and state-of-the-art algorithms for phonetic alignment to identify cognates in multi-lingual wordlists. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Long Papers*, 1204–1215. Valencia: Association for Computational Linguistics.

Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.

Lee, Yeon-Ju & Laurent Sagart. 2008. No limits to borrowing: The case of Bai and Chinese. *Diachronica* 25(3). 357–385.

Levenshtein, V. I. 1965. Dvoičnye kody s ispravleniem vypadenij, vstavok i zameščenij simvolov. *Doklady Akademij Nauk SSSR* 163(4). 845–848.

List, Johann-Mattis. 2012a. Lexstat. Automatic detection of cognates in multilingual wordlists. In *Proceedings of the EACL 2012 Joint Workshop of Visualization of Linguistic Patterns and Uncovering Language History from Multilingual Resources*, 117–125. Stroudsburg.

List, Johann-Mattis. 2012b. SCA: Phonetic alignment based on sound classes. In Marija Slavkovik & Dan Lassiter (eds.), *New directions in logic, language, and computation*, 32–51. Berlin & Heidelberg.

List, Johann-Mattis. 2014. Investigating the impact of sample size on cognate detection. *Journal of Language Relationship* 11. 91–101.

List, Johann-Mattis. 2015. Network perspectives on Chinese dialect history. *Bulletin of Chinese Linguistics* 8. 42–67.

List, Johann-Mattis. 2017. A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. System Demonstrations*, 9–12. Valencia: Association for Computational Linguistics. http://edictor.digling.org.

List, Johann-Mattis, Cormac Anderson, Tiago Tresoldi & Robert Forkel. 2021. *Cross-Linguistic Transcription Systems [dataset, Version 2.2.0]*. Jena: Max Planck Institute for the Science of Human History. https://clts.clld.org.

List, Johann-Mattis, Michael Cysouw & Robert Forkel. 2016. Concepticon. A resource for the linking of concept lists. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Tenth International Confer-*

*ence on Language Resources and Evaluation*, 2393–2400. European Language Resources Association (ELRA).

List, Johann-Mattis & Robert Forkel. 2022a. Automated identification of borrowings in multilingual wordlists [version 3; peer review: 4 approved]. *Open Research Europe* 1(79). 1–11.

List, Johann-Mattis & Robert Forkel. 2022b. *LingPy. A Python library for quantitative tasks in historical linguistics [Software, Version 2.6.8]*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://lingpy.org.

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch & Russell D. Gray. 2022. Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data* 9(316). 1–31.

List, Johann-Mattis, Simon J. Greenhill & Russell D. Gray. 2017. The potential of automatic word comparison for historical linguistics. *PLOS ONE* 12(1). 1–18.

List, Johann-Mattis, Shijulal Nelson-Sathi, Hans Geisler & William Martin. 2014. Networks of lexical borrowing and lateral gene transfer in language and genome evolution. *Bioessays* 36(2). 141–150.

List, Johann-Mattis, Christoph Rzymski, Simon J. Greenhill, Nathanael E. Schweikhard, Annika Tjuka & Robert Forkel. 2022. *Concepticon. A resource for the linking of concept lists [Dataset, Version 2.6.0]*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://concepticon.clld.org/.

List, Johann-Mattis, Mary Walworth, Simon J. Greenhill, Tiago Tresoldi & Robert Forkel. 2018. Sequence comparison in computational historical linguistics. *Journal of Language Evolution* 3(2). 130–144.

Macdonald, Kevin C. 1997. More forgotten tells of Mali: An archaeologist's journey from here to Timbuktu. *Archaeology International* 1. 40–42. http://doi.org/10.5334/ai.0112.

Marchal, Jean-Yves. 1978. Vestiges d'occupation ancienne au Yatenga (Haute-Volta): Une reconnaissance du pays Kibga. *Cahiers ORSTOM. Série Sciences Humaines* 15(4). 449–484.

Matisoff, James A. 1978. *Variational semantics in Tibeto-Burman. The 'organic' approach to linguistic comparison*. Philadelphia: Institute for the Study of Human Issues.

Mayor, A. & E. Huysecom. 2016. "Toloy", "Tellem", "Dogon": Une réevaluation de l'histoire du peuplement en Pays dogon (Mali). In M. Lafay & E. Coulibaly F. Le Guennec-Coppens (eds.), *Regards scientifiques sur l'afrique depuis les indépendances*, 333–350. Paris: Karthala.

Mayor, A., E. Huysecom, A. Gallay, M. Rasse & A. Ballouche. 2005. Population dynamics and paleoclimate over the past 3000 years in the Dogon Country, Mali. *Journal of Anthropological Archeology* 24. 25–61.

Mayor, A., E. Huysecom, S. Ozainne & S. Magnavita. 2014. Early social complexity in the Dogon Country (Mali) as evidenced by a new chronology of funerary practices. *Journal of Anthropological Archeology* 34. 17–41.

Mennecier, Phillipe, John Nerbonne, Evelyne Heyer & Franz Manni. 2016. A Central Asian language survey. *Language Dynamics and Change* 6(1). 57–98.

Moran, Steven & Michael Cysouw. 2018. *The Unicode Cookbook for Linguists: Managing writing systems using orthography profiles*. Berlin: Language Science Press. http://langsci-press.org/catalog/book/176.

Moran, Steven, Robert Forkel & Jeffrey Heath (eds.). 2016. *Dogon and Bangime linguistics*. Jena: Max Planck Institute for the Science of Human History. http://dogonlanguages.org/.

Moran, Steven & Jelena Prokić. 2013. Investigating the relatedness of the endangered Dogon languages. *Literary and Linguistic Computing* 28(4). 676–691.

Morrison, David A. 2014. Phylogenetic networks: A new form of multivariate data summary for data mining and exploratory data analysis. *WIREs Data Mining and Knowledge Discovery* 4(4). 1–17.

Nunn, Nathan & Diego Puga. 2012. Ruggedness: The blessing of bad geography in Africa. *The Review of Economics and Statistics* 94(1). 20–36.

Osborn, Donald Zhang, Joseph I. Donahoe & David J. Dwyer. 1993. *A Fulfulde (Maasina)-English-French lexicon: A root-based compilation drawn from extant sources followed by English-Fulfulde and French-Fulfulde listings = Lexique Fulfulde (Maasina)-Anglais-francais*. East Lansing: Michigan State University Press. //catalog.hathitrust.org/Record/002789711.

Plungian, Vladimir A. & Issiaka Tembine. 1994. Vers une description sociolinguistique du pays Dogon: Attitudes linguistiques et problèmes de standardisation. In Gérard Dumestre (ed.), *Stratègies communicatives au Mali: Langues régionales, bambara, francaise*, 163–195. Paris: Didier Erudition.

Pritchard, Jonathan K., Matthew Stephens & Peter Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155. 945–959.

Prokhorov, Kirill, Jeffrey Heath & Steven Moran. 2012. Dogon classification. In *Proto-Niger-Congo: comparison and reconstruction interna-*

*tional congress*. INALCO, Paris. https://llacan.cnrs.fr/fichiers/nigercongo/fichiers/Dogon%5C%20classification.ppt.

Rama, Taraka, Johann-Mattis List, Johannes Wahle & Gerhard Jäger. 2018. Are automatic methods for cognate detection good enough for phylogenetic reconstruction in historical linguistics? In *Proceedings of the North American Chapter of the Association of Computational Linguistics*, 393–400.

Ross, Malcom & Mark Durie. 1996. Introduction. In Mark Durie (ed.), *The comparative method reviewed. Regularity and irregularity in language change*, 3–38. New York: Oxford University Press.

Samar, Reza Ghafar & Tej K. Bhatia. 2017. Predictability of language death: Structural compatibility and language contact. *Language Sciences* 62. 52–65.

Segerer, Guillaume & Sébastien Flavier. 2016. *RefLex: Reference Lexicon of Africa*. Version 1.1. Paris, Lyon. http://reflex.cnrs.fr/.

Simons, Gary F. & Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the World, Twenty-first edition*. http://www.ethnologue.com/18.

Sokal, Robert. R. & Charles. D. Michener. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* 28. 1409–1438.

Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96(4). 452–463.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137.

Tadmor, Uri. 2009. Loanwords in the world's languages. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 55–75. Berlin & New York: de Gruyter.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, Agnes A. Awomoyi, Jean-Marie Bodo, Ogobara Doumbo, Muntaser Ibrahim, Abdalla T. Juma, Maritha J. Kotze, Godfrey Lema, Jason H. Moore, Holly Mortensen, Thomas B. Nyambo, Sabah A. Omar, Kweli Powell, Gideon S. Pretorius, Michael W. Smith, Mahamadou A. Thera, Charles Wambebe, James L. Weber & Scott M. Williams. 2009. The genetic structure and history of Africans and African Americans. *Science (New York, N.Y.)* 5930(324). 1035–1044.

van der Ark, René, Philippe Mennecier, John Nerbonne & Franz Manni. 2007. Preliminary identification of language groups and loan words in Central Asia. In *Proceedings of the RANLP Workshop on Acquisition and Management of Multilingual Lexicons*, 13–20.

Vydrin, V. 2009. On the problem of the Proto-Mande homeland. *Journal of Language Relationships*. 107–142.

Williamson, Kay & Roger Blench. 2000. Niger–Congo. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*, 11–42. Cambridge University press: Cambridge.

Wu, Mei-Shin, Nathanael E. Schweikhard, Timotheus A. Bodt, Nathan W. Hill & Johann-Mattis List. 2020. Computer-assisted language comparison: State of the art. *Journal of Open Humanities Data* 6(2). 1–14.

# Appendix

## Shared Numerals 1–5

| DOCULECT | ONE | TWO | THREE | FOUR | FIVE |
|---|---|---|---|---|---|
| Bangime (Isolate) | tòrè | ʒìndò | tà:rù | (nè:$^{n}$) | (núndī) |
| Bambara (Mande) | kélé$^{n}$ | fìlá | sàbá | (ná:ní) | dú:rú |
| Bankan Tey (Dogon) | tùmá | jŏj | tà:ní | (nìŋŋĕj$^{n}$) | nùmmŭj$^{n}$ |
| Ben Tey (Dogon) | tùw$^{n}$ɔ́ | jĕj | tà:nú | (nĭ:j$^{n}$) | nùmŭj$^{n}$ |
| Bunoge (Dogon) | tó:lè | dè:gà | tá:ndù | (nê:w$^{n}$) | nɔ́:mɔ̀ |
| Fulfulde (Atlantic) | go'o | ɗiɗi | tati | (naji) | joji |
| Jamsay (Dogon) | túrú | lĕj | tă:n | (năj$^{n}$) | nŭ:j$^{n}$ |
| Jenaama (Mande) | kéẁ$^{n}$ | pēndē | sìkèẁ$^{n}$ | nàtàw$^{n}$ | kɔ̀:gɔ̀w$^{n}$ |
| Mombo (Dogon) | tá:ŋgù jè: | né:ŋgá | tá:ndì | ké:dʒɔ́ | nú:mò |
| Najamba (Dogon) | kúndé | nô:j | tà:ndî: | ké:dʒɛ̀j | nùmî: |
| Nanga (Dogon) | tùmâ | wŏj | tà:ndĭ: | (nɔ̆j$^{n}$) | nìmĭ: |
| Penange (Dogon) | tà:ŋgà | nègà | tà:ndì | kèdʒɔ̀ | nɔ̀:m |
| Perge Tegu (Dogon) | túrú | lĕj | tà:lú | (năj$^{n}$) | nùmĭ: |
| Songhai Senni | fɔ́: | híŋká | hínzâ | tá:cí | gú: |
| Songhai Kiini | fó: | híŋká | — | — | gú: |
| Tebul Ure (Dogon) | túrɛ̀: | lé:rú | tà:ndú | kédɛ́ | nŭm |
| Tiranige Diga (Dogon) | tò:mà | nì:ŋgà | tá:ndí | cè:dʒɔ̀ | nú: |
| Togo Kan (Dogon) | túrú | lɔ́j | tă:n | (năj) | (núné:) |
| Tommo So (Dogon) | túrú | néé | tàándú | (năj) | (ǹnɔ́) |
| Toro Tegu (Dogon) | túrú | lĕj | tá:r$^{n}$ú | (năj) | nŭ:j |
| Yanda Dom (Dogon) | tùmá | nɔ́: | tá:ndù | cézɔ́ | nûm |
| Yorno So (Dogon) | túrú | léj | tă:n | (năj$^{n}$) | nùmɔ́r$^{n}$ɔ́ |

## Shared Numerals 6–10

| DOCULECT | SIX | SEVEN | EIGHT | NINE | TEN |
|---|---|---|---|---|---|
| Bangime (Isolate) | ké:rè | (kí:jè) | ságì | (tɛ̀gɔ̀) | kùrɛ̀ |
| Bambara (Mande) | wɔ́:rɔ́ | wóló$^{n}$wùlá | sé:gí$^{n}$ | kɔ̀nɔ̀$^{n}$tɔ́$^{n}$ | tá$^{n}$ |
| Bankan Tey (Dogon) | kúròj | síj$^{n}$ɔ̀j$^{n}$ | gá:rà-j | (tè:sú-m) | pɛ́:rú |
| Ben Tey (Dogon) | kúròj | súj$^{n}$ɔ̀j$^{n}$ | gá:rà-j | (tè:sǐ-m) | pɛ́rú |
| Bunoge (Dogon) | kúléw$^{n}$ | sɔ́:w$^{n}$ | (sé:léw$^{n}$) | (tó:wà) | kóbéw$^{n}$ |
| Fulfulde (Atlantic) | dʒe:gon | dʒeɗɗi | dʒetti | dʒe:naj | sappo |
| Jamsay (Dogon) | kúrój | sûj$^{n}$ | gá:rà | lá:rúwà | pɛ́rú |
| Jenaama (Mande) | tùùmì$^{n}$ | jìjènì | sɛ̄kī | kàpì | cɛ̄ɛ̄m |
| Mombo (Dogon) | kúléj$^{n}$ | sɔ́:lì | (sé:lè) | (tó:wà) | pɛ́:lù |
| Najamba (Dogon) | kúlèj | swɛ̂j | sá:gì: | twâj | pídʒɛ́lì |
| Nanga (Dogon) | kúrê | súj̀ê | gá:rè | (tè:sǐ:) | pɛ́:rù |
| Penange (Dogon) | kùlè:nì | sɔ̀:lì | (sè:lè) | (tò:wà) | pè:l |
| Perge Tegu (Dogon) | kúró:j | sûj$^{n}$ | gá:rà | lá:rwà | pɛ́rú |
| Songhai Senni | ʔíddù | (ʔí:jê) | ʔíjákù | à-fɔ́:-sí: à-wój | wój |
| Songhai Kiini | ʔíddû | (ʔí:jè) | já:hâ | jággâ | wój |
| Tebul Ure (Dogon) | kúlé | sɔ́ | gà gárà | gà gárà-bà | pɛ́lú |
| Tiranige Diga (Dogon) | kùlèj$^{n}$ | sɔ́:j | (sé:lé) | (tó:wá) | pìjɔ̀lù |
| Togo Kan (Dogon) | kúré: | sɔ̂: | (sìlâ:) | (tùwâ:) | pɛ́rú |
| Tommo So (Dogon) | kúlój | sɔ́j | gá gìrà | (túwwɔ́) | pɛ́lú |
| Toro Tegu (Dogon) | kúréj | sój$^{n}$ | gá:rà | lá:rà | pɛ́:rú |
| Yanda Dom (Dogon) | kúlé | swɛ́: | sá:gè | twâ: | píjél |
| Yorno So (Dogon) | kúlòj | sɔ̂j | gá gàrà | (tùwɔ́) | pél |

## Shared Body Parts (1)

| DOCULECT | BEARD | CHEST | JAW |
|---|---|---|---|
| Bangime (Isolate) | sèmbò | gɔ̀ndɛ̀ | kāw$^{n}$ɔ̀ |
| Bambara (Mande) | bò$^{n}$bò$^{n}$ sí: | dísí | dákóló |
| Bankan Tey (Dogon) | bɛ̌j$^{n}$ | (gèndè kû:) | àndì: cìr$^{n}$ɛ̌j$^{n}$ |
| Ben Tey (Dogon) | bɛ̌j$^{n}$ | dʒɛ̀lɛ̀ kû: | áŋgí |
| Bunoge (Dogon) | ʔɔ̀ndɔ́ kùlɛ̀ | bɛ́mbɔ̀ | (ŋá:ŋà:) |
| Fulfulde (Atlantic) | waare | beccal | balaken kewal |
| Jamsay (Dogon) | bɛ̀w kúrɔ́ | gɔ̀ŋɔ́ | ɛ̀w$^{n}$ɛ́ |
| Jenaama (Mande) | ʃìmbò cí | kēndē | dèŋgè |
| Mombo (Dogon) | bé:$^{n}$ kùlɛ̀ | (gɔ̂:$^{n}$) | áw$^{n}$à |
| Najamba (Dogon) | bɛ̂:ŋgò | pélɛ̀-ŋgè | dʒà dʒà-gâ: |
| Nanga (Dogon) | bè bɛ̂: | gòmbó-rô | ăŋ |
| Penange (Dogon) | dɔ́mdʒɔ́ kùlɛ̀ | (gɔ̀:$^{n}$) | ágá-lá |
| Perge Tegu (Dogon) | bè bɛ̌w | gɔ̀w$^{n}$ɔ́ | (áŋú) |
| Songhai Senni | kà:bè | gàndè | gà-gáb-ò |
| Tebul Ure (Dogon) | bèjà: kúlà | gwá:$^{n}$ | àŋgă: |
| Tiranige Diga (Dogon) | jéná kùlè-ŋgé | bémbé | (áná) |
| Togo Kan (Dogon) | bɛ̌ | gɔ̀ŋɔ́ | ɛ̀w$^{n}$ɛ́ |
| Tommo So (Dogon) | bɛ̀ɛ́ | gèŋŋú | àŋgàà dúú |
| Toro Tegu (Dogon) | bɛ̌w | cèlkú | kà pèŋgù-ró |
| Yanda Dom (Dogon) | bìjà kùlà | dʒèŋù | àŋgà |
| Yorno So (Dogon) | bɛ̌: | (gɔ̌:$^{n}$) | àŋgă:$^{n}$ |

## Shared Body Parts (2)

| DOCULECT | JOINT | LUNG | VAGINA |
|---|---|---|---|
| Bangime (Isolate) | dɔ̀kɛ́-rɛ́ | pógó pógò | númbɛ́ |
| Bambara (Mande) | kóló túgúdá | fógóⁿ fògóⁿ | kɔ́dá |
| Bankan Tey (Dogon) | díːráj | pùsù pàsá | dúmbá |
| Ben Tey (Dogon) | díjáj | pùːsù pàːsá | (dʒéré) |
| Bunoge (Dogon) | palu-giʔel | pùsù pùsú | — |
| Fulfulde (Atlantic) | dìgú | wuucere | faata |
| Jamsay (Dogon) | wáɲɔ̄wⁿ | pùːdʒú pàːdʒú | (dɛ̀mɛ́) |
| Jenaama (Mande) | dígé-lé | — | pùwɔ̀ |
| Mombo (Dogon) | dígìn-gè | pùdʒù pádʒù | sébì |
| Najamba (Dogon) | dígá | bú bùdʒûː | dúmbú |
| Nanga (Dogon) | dìgì-lè | kɛ̀ndɛ̀ | kéndʒí |
| Penange (Dogon) | dìgú | pùgùdʒáⁿ pùgùdʒáⁿ | sùgù |
| Perge Tegu (Dogon) | dóbó | pùːsù pàːsá | dɛ̀mbɛ́ |
| Songhai Senni | dóːbú | kùmbò | dɔ̀fɛ̀ |
| Tebul Ure (Dogon) | dígú-gɔ́ | pùdù páːdù | èndɛ́ |
| Tiranige Diga (Dogon) | dìgí | àl pús | — |
| Togo Kan (Dogon) | mùŋú | púːdʒù páːdʒù | íːɲí |
| Tommo So (Dogon) | dìgú | kìndɛ̀ bùúdú bààdù | dɛ̀mmɛ̀ bòndó |
| Toro Tegu (Dogon) | dìgú | kùpàːⁿ sú | pɔ̀rɔ́ |
| Yanda Dom (Dogon) | dìgù dígù | púːzù páːzù | càː |
| Yorno So (Dogon) | dìgú | bŭːdʒù bàːdʒù | dɛ̀mɛ̀ bònɔ́ |

## Shared Agriculture Vocabulary (1)

| DOCULECT | CASSAVA | COW | GRASS | HORN |
|---|---|---|---|---|
| Bangime (Isolate) | bànàŋ̀kùù | nàà | gūdʒ̀è | (sìrà) |
| Bambara (Mande) | bànàⁿkú | mìsí | bíⁿ | bíɲɛ́ |
| Bankan Tey (Dogon) | bànàŋkú | nằːm | bèrú | kíjà |
| Ben Tey (Dogon) | bànàkùː béré | nằːm | bèrú | cíjà |
| Bunoge (Dogon) | bànàŋkú | ná | kɔ́dʒì | (kélè) |
| Fulfulde (Atlantic) | bananku | (nagge) | huɗo | buutu tuwal |
| Jamsay (Dogon) | bànàkûːⁿ | (nàŋà-náː) | ɔ̀jɔ́ | (círɛ́) |
| Jenaama (Mande) | bànàⁿkuⁿ | nà | sɔ̀gù | būwōⁿ |
| Mombo (Dogon) | bálákûːⁿ | nâː | bélé-ŋgè | (kélɛ́) |
| Najamba (Dogon) | bànàŋkûː-ŋgò | nɛ̌ː | sɔ̌ːmɛ̀-ŋgò | (kélɔ̀ː) |
| Nanga (Dogon) | bànjî | (nàŋá) | bèrí | (kírâ) |
| Penange (Dogon) | dàlmà-dʒògó | nàː | kɔ́jí | (kélɛ́) |
| Perge Tegu (Dogon) | bànàkùrù-wòsó | (nàŋá) | ɔ̀jɔ́ | (kírɛ́) |
| Songhai Kiini | bánáŋkù | kòpsò | — | (hìlí) |
| Songhai Senni | bánáŋkû | háw | sùbò | — |
| Tebul Ure (Dogon) | bàrⁿàkúːⁿ | nằː | jɔ̀wá | (cílá) |
| Tiranige Diga (Dogon) | bànáŋkúː | náː | bél-ŋge | (kɛ̀ːlɛ̀-ŋgè) |
| Togo Kan (Dogon) | bànàkùː báːgá | (nàʔⁿà) | gú gúrú | (kírɛ́) |
| Tommo So (Dogon) | bànànkù tímɛ́ | nàá | kèrú | (kélɛ́) |
| Toro Tegu (Dogon) | bànàŋkú | (nàŋá) | sɔ̀lɔ́ | (círá) |
| Yanda Dom (Dogon) | bànàkùl táná | nàː | ɔ̀à | (célá) |
| Yorno So (Dogon) | wòdʒó | nằː | dɔ̀gɔ́ | (kílè) |

## Shared Agriculture Vocabulary (2)

| DOCULECT | MAIZE | MARROW | MEAT |
|---|---|---|---|
| Bangime (Isolate) | bìrɔ̀ ndɔ́$^{n}$ | dòndì | ŋàw$^{n}$ |
| Bambara (Mande) | kàbá | sɛ́mɛ́ | sogo |
| Bankan Tey (Dogon) | (bàrà èmbɛ̌j) | nɔ̀:nî: | nàmâ: |
| Ben Tey (Dogon) | bàr mɛ̌j$^{n}$ | bú:sǎ-m | nàw$^{n}$â: |
| Bunoge (Dogon) | dílímà | búgè | námà |
| Fulfulde (Atlantic) | mbamm baari | mbuso | teew |
| Jamsay (Dogon) | pòrò: ɲú: | búsâm | nɔ̀w$^{n}$ɔ́ |
| Jenaama (Mande) | kèndè | dòndì | tō |
| Mombo (Dogon) | èmbɛ̀ dúlúmá | bùzé | námà |
| Najamba (Dogon) | mà dèmbáŋgó | nòndî: | nàmâ: |
| Nanga (Dogon) | jàr$^{n}$à mbɛ́ | nɔ̀ndí | nàmâ |
| Penange (Dogon) | dàlmà jògó | nù: | nàmà |
| Perge Tegu (Dogon) | (bàrà ɛ̀:mbɛ́) | búsâm | nàw$^{n}$á |
| Songhai Kiini | kòtkòlò | lòndò | hâm |
| Songhai Senni | màlàŋ hàmbó | lòndí | hámó |
| Tebul Ure (Dogon) | sàr$^{n}$à èmbɛ́ | nɔ́: | nàmá |
| Tiranige Diga (Dogon) | dùmà jò:gé | nó:ndí | námá |
| Togo Kan (Dogon) | màɲèr$^{n}$ èmá | sìjè-i$^{n}$ bɔ́:nɔ́ | nɔ̀w$^{n}$ɔ́ |
| Tommo So (Dogon) | sànà èmmɛ́ | bɔ́ndú | nàmá |
| Toro Tegu (Dogon) | tɔ̀r$^{n}$í jⁿàr$^{n}$à | síwó | nàw$^{n}$á |
| Yanda Dom (Dogon) | sànà èmè | bɔ̀ndú-m | nàmà |
| Yorno So (Dogon) | sàr$^{n}$à èmɛ́ | bɔ́nɔ̀ | nàw$^{n}$á |

## Bangime vocabulary shared with Dogon, and Mande languages (1)

| DOCULECT | NOBLE | SLAVE | TONGS | FIGHT |
|---|---|---|---|---|
| Bangime (Isolate) | hɔ̄ɔ̄rɔ̀ | kɔ̀mɛ̀ | kámbá-rà | kóré |
| Bambara (Mande) | hɔ́rɔ́$^{n}$ | (kòmê) | bàlá$^{n}$ | kɛlɛ |
| Bankan Tey (Dogon) | dímò-m | (gùndǎ-m) | kámbò | zàjá |
| Ben Tey (Dogon) | bà:mà-jí-m | (gùlǔ-m) | ɛ̌w | dʒàdʒá |
| Bunoge (Dogon) | hɔ́:lɔ̀ | kɔ́mbɛ̀ | kàbàlà | ɲɔ́ŋɛ̀ |
| Fulfulde (Atlantic) | dimo | maccu-ɗo | kampaaje | haɓu-de |
| Jamsay (Dogon) | ɔ̀ɣɔ̀ în | (gùnón | ɛ̌w | dʒɛ̀dʒɛ́ |
| Jenaama (Mande) | hɔ̄ɔ̄rɔ̄w$^{n}$ | (kɔ̀ɲɔ́$^{n}$) | kùɥgè kámbō | kɛ̄lɛ̄ |
| Mombo (Dogon) | dímò | (gúndà) | kábá lámá | dʒáw$^{n}$-j$^{n}$è |
| Najamba (Dogon) | nò: nálà | (gùndɛ́) | ɛ̌:gò | dʒá:ŋí-j |
| Nanga (Dogon) | kɔ́:là | (gùndá) | ê: | dʒòrí-jé |
| Penange (Dogon) | hɔ́:lɔ́ | (gùndà) | kábálá | dʒájⁿ |
| Perge Tegu (Dogon) | ɔ̀gɔ̀ î:$^{n}$ | (gùlɔ̃n) | êw | dʒàjá |
| Songhai Senni | bòrkìnò | (kɔ̀w$^{n}$ɔ́:) | — | wàŋgù |
| Songhai Kiini | bòrkǐn | (tàmò) | — | gùrdʒèj |
| Tebul Ure (Dogon) | ɲì dímè | (gùndɔ́) | — | zèdʒí |
| Tiranige Diga (Dogon) | ɔ:lá | kɔ́mɛ́ | è: | jáŋí-jó |
| Togo Kan (Dogon) | hɔ́:rɔ́$^{n}$ | (gùnɔ́) | ɛ̌w | dʒàj-î: |
| Tommo So (Dogon) | ɔ́lɔ́ nɔ́ | (gùnnɔ̀-nɔ́) | èwɛ́ | dʒàwí-jé |
| Toro Tegu (Dogon) | bɔ̀rcì ní | (gùnɔ́) | kâm | zɔ̌ŋ |
| Yanda Dom (Dogon) | dímè | (gùnò) | èw | zá:ní-j$^{n}$é |
| Yorno So (Dogon) | — | (gùnɔ̀-nɛ́) | ɛ̌w | dʒàjɛ́: |

## Bangime vocabulary shared with Dogon, and Mande languages (2)

| DOCULECT | FOAM | NEW | LOUSE | WORK |
|---|---|---|---|---|
| Bangime (Isolate) | pūrɛ̀ɛ̀ | kárá | sɛ̀mà | wārì |
| Bambara (Mande) | kà$^{n}$gá | kúrá | — | báárá |
| Bankan Tey (Dogon) | kù kù:rú | kándà | tó:r$^{n}$ú-m | bírá: |
| Ben Tey (Dogon) | júgúrɔ̀ | kálà | kɔ̀tɔ̀ kɔ̃:-m | — |
| Bunoge (Dogon) | pùlà | —- | sémɔ̀ | wàlè |
| Fulfulde (Atlantic) | ƒuufo | hesu-de | temba | golu-de |
| Jamsay (Dogon) | pútɔ̀ | kàná | kàrùm-kă: | bírɛ́ |
| Jenaama (Mande) | púlā | tōē | sɔ̀mɔ̀ | kāj$^{n}$ |
| Mombo (Dogon) | púlɔ́ | kándá | sémɛ̀ | wálé |
| Najamba (Dogon) | bùdʒɛ̀ búdʒɛ̀-ŋgò | kàndă: | nígídʒí | bírɔ́: |
| Nanga (Dogon) | músî | kándà | kɔ̀rɔ̀ŋ kɔ̃: | bírá |
| Penange (Dogon) | púlá | kàndà | sɛ̀mɛ̀ | káj$^{n}$ |
| Perge Tegu (Dogon) | jɔ̀gɔ̀rɔ̀ pútɔ̀ | kàlá | kɔ̀rɔ̀m kɔ̃: | bírɛ́ |
| Songhai Senni | kù:fò | — | gán ó | gój-ó |
| Songhai Kiini | kòbú | tàw | gá:ní | gój-ó |
| Tebul Ure (Dogon) | júbdá | kàndá | kɔ̀rɔ̀ŋ gɛ̆: | bírá: |
| Tiranige Diga (Dogon) | pùlá | kándá | sémɔ́ | wàlè |
| Togo Kan (Dogon) | jùgɔ́rɔ́ | kàná | kàŋ kă: | bírɛ́ |
| Tommo So (Dogon) | júbúdɔ́ | kàndá | — | bírɛ́ |
| Toro Tegu (Dogon) | kùgíj | kàlá | tór$^{n}$ú | bírá |
| Yanda Dom (Dogon) | bùŋá | kàndà | sɛ̀mɛ̀ | bìdɛ́ |
| Yorno So (Dogon) | júgɔ̀rɔ̀ | kàná | sèw$^{n}$ɛ́ | bírɛ́ |

**Basic and cultural vocabulary items in Bangime shared predominantly with Dogon varieties (1)**

| DOCULECT | SONG | STAR | ROOF |
|---|---|---|---|
| Bangime (Isolate) | ŋúɥ$^{n}$ê: | tōrè: | (kùmbè) |
| Bambara (Mande) | — | kwe | sá$^{n}$fɛ́lá |
| Bankan Tey (Dogon) | nùmá | bàndà kɔ̀:rɔ̆j | (tèmbè kû:) |
| Ben Tey (Dogon) | nùw$^{n}$ɔ́ | tóró | — |
| Bunoge (Dogon) | núŋɔ̀ | dʒɔ́ŋ gúlɛ̀ | gá:ŋgù |
| Fulfulde (Atlantic) | dʒimol | albalda | (tiba) |
| Jamsay (Dogon) | nûŋ | tóró | (tèmè kú:$^{n}$) |
| Jenaama (Mande) | sùwōgù | tɔ́ɲɔ̄w$^{n}$ | bɛ̀ndɛ̀ |
| Mombo (Dogon) | (nɔ̂:) | áŋ kólɛ̀ | sà:súgú |
| Najamba (Dogon) | ŋwànă: | àndʒì | (dèbí) |
| Nanga (Dogon) | nùŋá | wà: kɔ̀rɔ́ | (dɛ̆w) |
| Penange (Dogon) | (nɔ̀:) | ɔ̀dʒɔ̀ gɔ̀lɔ̀ | dógó |
| Perge Tegu (Dogon) | núŋgɔ̀ | tóró | (tèmbè kú:) |
| Songhai Senni | dò:nò | — | — |
| Songhai Kiini | dò:ní | — | féféndé |
| Tebul Ure (Dogon) | nùŋă: | tóndó: | ládúgɔ́ |
| Tiranige Diga (Dogon) | (nwɛ́:) | kɔ̀lɔ̀ŋ kólɔ̀w$^{n}$ | tèmbè |
| Togo Kan (Dogon) | núŋú | tóró | kù$^{n}$ mɔ̀rú |
| Tommo So (Dogon) | núdʒɔ́ | tòndòlòj | (dèmbɛ́) |
| Toro Tegu (Dogon) | nùŋɔ́ | sémrĕj | dómó kù |
| Yanda Dom (Dogon) | nùŋà | tóndóló | (dèbù) |
| Yorno So (Dogon) | (né:) | tònòlòj | dă: |

**Basic and cultural vocabulary items in Bangime shared predominantly with Dogon varieties (2)**

| DOCULECT | GO DOWN | HELP | PUSH | SHOW |
|---|---|---|---|---|
| Bangime (Isolate) | sūuɥè | (bɔ̄gɔ̀) | tɛ̀mbɛ̀ | tɛ̄rɛ̀ |
| Bambara (Mande) | (dʒìgíⁿ) | — | dígí | jɛ̀lɛ́ⁿ |
| Bankan Tey (Dogon) | sídʒé | (bàrí) | dàmbí | kɛ́rí |
| Ben Tey (Dogon) | sídʒé | bàrí) | dàmbí | cɛ́ːrí |
| Bunoge (Dogon) | — | (bánnè) | dúndʒú-rè | (tégó-mì) |
| Fulfulde (Atlantic) | dʒippaa-de | wallu-de | duɲu-de | hollu-de |
| Jamsay (Dogon) | súgó | (bàrá) | — | |
| Jenaama (Mande) | dʒàwⁿ | — | cɔ̄bì | à-wādʒī |
| Mombo (Dogon) | sígé | — | túmbjêː | táːrè |
| Najamba (Dogon) | súgí | (bǎr) | dàmbí | téːrɛ́ |
| Nanga (Dogon) | sígé | (bǎːrí) | dàmbí | kéːrí |
| Penange (Dogon) | sígé | (bárgè) | túmbú-gè | táːrè |
| Perge Tegu (Dogon) | sígé | (bàrá) | dàmbá | tóːró |
| Songhai Senni | zùmbù | fàːbà | tàŋà | bándʒ-édʒndí |
| Songhai Kiini | zùmbù | fàːbó | tàŋ | — |
| Tebul Ure (Dogon) | súgó | (bàrá) | dàmbí | tágá |
| Tiranige Diga (Dogon) | sígó | (báró) | dámbó | táːró |
| Togo Kan (Dogon) | súgó | (bàrá) | dèmé | tóːrù |
| Tommo So (Dogon) | súgó | (bàrá) | dàmbá | táárá |
| Toro Tegu (Dogon) | sígú | (bàrú) | (dǎm) | tótú |
| Yanda Dom (Dogon) | súwó | (bàdú) | dàmbú | dámdɛ́ |
| Yorno So (Dogon) | súgó | (bàrá) | dàmá | táːrá |