



HAL
open science

Areal patterns and colexifications of colour terms in the languages of Africa

Guillaume Segerer, Martine Vanhove

► **To cite this version:**

Guillaume Segerer, Martine Vanhove. Areal patterns and colexifications of colour terms in the languages of Africa. *Linguistic Typology*, 2021, 10.1515/lingty-2021-2085 . halshs-03483348

HAL Id: halshs-03483348

<https://shs.hal.science/halshs-03483348v1>

Submitted on 16 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Areal patterns and colexifications of colour terms in the languages of Africa

Guillaume Segerer & Martine Vanhove

/gi'jom seʒe'βεε/ & /maʁ'tin va'nɔv/

LLACAN (CNRS, INALCO)

Abstract:

*Although colour terms are the semantic domain which has attracted the largest number of studies, notably from a typological point of view, searching the large online cross-linguistic lexical database of African languages (RefLex, Segerer & Flavier 2011-2019) pointed out several previously undetected areal colexification patterns and shared lexico-constructional patterns in a genetically balanced sample of 401 languages spoken Africa. In this paper, we illustrate several areal characteristics of colours terms along three scenarios: (i) the spread of an areal feature due to a common extra-linguistic setting (locust bean – *Parkia biglobosa* – as the lexical source of YELLOW); (ii) two convergence phenomena, one based on a shared lexico-constructional pattern including a term for WATER, and one based on shared colexifications (RED and RIPE vs. GREEN and UNRIPE); and (iii) an areal pattern of lexical diffusion of colour ideophones, a category which was considered difficult to borrow.*

Keywords: colour terms; colexification; shared lexico-constructional pattern; African languages; database; ideophones; borrowing; convergence phenomena; areal patterns

1. Introduction

Colour denomination is probably the semantic domain that has attracted the largest number of lexico-semantic, sociolinguistic, cognitive and/or crosslinguistic studies since Berlin & Kay's (1969) seminal article. In this study, they proposed a universalist approach, based on elicitation with the Munsell colour chart, to "basic" colour terms defined as "contrastive, nonoverlapping terms that are assumed to exhaustively partition color space." (Malt & Majid 2013: 585). These are identified by several criteria, among them, the fact that a so-called "basic" colour term must be mono-lexemic, must not be linked to a source term, and must not be a recent loan from a foreign language. Berlin & Kay further proposed a diachronic implicational hierarchy of colour lexicons. Their approach was followed by numerous studies and refinements,¹ among them the *World Color Survey* online database of colour terms in a sample of 110 languages. It was also criticised by many scholars, mainly for methodological and ethno-centric reasons (for an overview of the critical arguments, see e.g. Dimmendaal 1995, 2015; Foley 1997; Wierzbicka 2008; Malt & Majid 2013).

This paper focuses on colour terms in the languages of Africa. Its aim is, however, not to dwell upon this long-standing debate, but instead (i) to explore shared colexifications and lexico-constructional patterns and their areal distributions; (ii) to discuss an unexpected areal pattern of borrowing.

¹ For data, works and references developed in this approach, which include over ten African languages, see the *World Color Survey* project.

The term and concept of “colexification”, valid both in synchrony and diachrony, is an attempt to solve the issue of monosemy *vs* polysemy in lexical studies. It was introduced by François (2008) for typological purposes, and has been defined as follows:

A given language is said to colexify two functionally distinct senses if, and only if, it can associate them with the same lexical form. ... [the term is] purely descriptive, and neutral with respect to semantic and historical interpretation (François 2008: 170).

By “lexical form” the author specifies that it “may refer to a lexeme or a construction, or occasionally to a lexical root” (François 2008: 170, fn 4), i.e. colexification can be either strict or loose. Since then, the term has been largely adopted by the community of lexical and semantic typologists as an agnostic term to circumvent the difficulties (and impossibilities) that arise when working with large datasets and languages with no written histories, which is the case for most of the languages of our sample (see e.g. List et al. 2018: 279).

“Shared lexico-constructional patterns” are “expressions that constitute interlingual matches [which] show the same semantic and structural patterning, i.e. are made up of parts that match each other across the languages” (Koptjevskaja-Tamm & Liljegren 2017: 206-207). The authors use this expression as a more neutral term than “calque” “or translation loans” because “in most areal contexts it is difficult or even impossible to identify the model and the replica languages” (Koptjevskaja-Tamm & Liljegren 2017: 207).

For the purposes of this paper, we looked at a very large sample of languages spoken in Africa (see Section 2), and in Section 5 have made some comparisons with other languages of the world.

This paper is organised as follows: Section 2 presents our data, database and methods for extracting colour-related terminology, briefly discussing the limits and advantages of such a large-scale lexical research. Section 3 illustrates an example of an areal feature spreading due to a common extra-linguistic setting, that of the locust bean as the lexical source of YELLOW. In Section 4 we discuss a convergence phenomenon for colour naming based on a shared lexico-constructional pattern including a term for WATER. Section 5 studies a convergence phenomenon in shared colexifications, strict and loose, namely that of the polysemy of RED and RIPE as opposed to GREEN and UNRIPE. Section 6 investigates an areal pattern of borrowings of colour ideophones, a word category that Africanists have long believed to be difficult to borrow across languages that already have a category of ideophones.

2. Data and methods

This research follows a previous study of the lexical sources, patterns of colexification, and patterns of lexicalisation of colour terms in a large sample of African languages² (Seeger and Vanhove 2019), for which we extracted data from bilingual dictionaries, several articles dedicated to colour systems and chapters in grammars, or received first-hand lexical data from colleagues. For retrieval and practical reasons, in this previous study, as well as for the present study, we had to limit the study to eleven colour terms (BLACK, WHITE, RED, YELLOW, GREEN, BLUE, GREY, BROWN, PURPLE, ORANGE and PINK).³ We searched the data via lexemes’ translations into European languages (Dutch, English, French, German, Portuguese).

² For the previous research our sample amounted to 350 languages, excluding Semitic and Berber languages. As mentioned below, in the present study the sample comprises 401 languages, including Semitic and Berber.

³ These correspond to the eleven so-called “basic” colour terms of Berlin & Kay (1969).

Table 1 provides the various (still recognisable) sources of these colour terms, ordered by broad semantic domains of origin and number of tokens. As expected from Berlin and Kaye's research, BLACK, WHITE and RED being most often monomorphemic unrecognisable words (a part of their definition of "basic" colour terms), their semantic source is most often unknown, hence the low figures associated with them in the table. The figures for the sources of PURPLE, ORANGE and PINK are also low but for a different reason: they are seldom mentioned in dictionaries (also probably rarely lexicalised).

Domain	Source	Black	White	Red	Yellow	Green	Blue	Grey	Brown	Orange	Purple	Pink	TOTAL	
VEGETAL KINGDOM 380	locust-bean				88				1	1			90	
	leaf					70	2		2				74	
	grass					44			1				45	
	plant	2	4	5	9	7	3	1	2	1	2		36	
	tree	2		5	12	1		1	7	2	4		34	
	fruit	3			5	2	3		4	10	2		29	
	flower		1	1	6						3	4	15	
	saffron				15									15
	kola			3	3					3	3			12
	algae					7	2							9
	corn				5	1								6
	mould					5			1					6
	other	1			1	2	1	1	1	1	1	1		9
MINERALS 110	ashes		2				2	30	1				35	
	soil		7	5	10				7	1	2	3	35	
	dust		1					5		1			7	
	charcoal	5											5	
	gold				5								5	
	soot	2						1					3	
	other	1	1	3	7		2	4	1			1	20	
ANIMALS 71	bird	3	1	4	2	5	5		2	1	1	1	25	
	mammal		1	2	1		1	10	4	1		1	21	
	reptile	1		1	2	6	1	2	1				14	
	fish	3	2	1			1		2				9	
	insect	1				1							2	
FOOD 56	fat				10					2			12	
	milk		6		4		2						12	
	coffee								9				9	

	juice			3				1				4	
	egg				3							3	
	other	1	1	6	2			4	1		1	16	
BODY 53	blood			14	1			2	1		1	19	
	excreta				12			1	1		1	15	
	body parts	1	3	3		1	3	3	1			15	
	body fluids				2	1						3	
	other			1								1	
ARTEFACTS 41	fabric	3	2		1	8		2			2	18	
	dye				1	6		1	1	1		10	
	'laundry blue'					5	1					6	
	other	1	2		1			2		1		7	
ATMOSPHERIC ELEMENTS 28	sky		1			15						16	
	mist	1					3	1				5	
	cloud		1									1	
	hail		1									1	
	ice		1									1	
	other			1		2		1				4	
OTHER 36	disease				6							6	
	cosmetics			2				1				3	
	other	1	3	1	4	1	1	4	6	1	4	1	27
TOTAL		Black	White	Red	Yellow	Green	Blue	Grey	Brown	Orange	Purple	Pink	
		26	39	60	223	157	63	67	73	30	23	14	

Table 1: Semantic sources of colour terms in all the sources for the languages of Africa (revised and expanded from Table 4 in Segerer & Vanhove 2019: 296-297)

Of course we are aware of the limitations of such an approach (for a full discussion see Segerer & Vanhove 2019). There are indeed many more colour-related terms than these eleven notions. It is well known that such a limited set blurs the native visual sense systems (see e.g. Dimmendaal 1995, 2015; Foley 1997; Dedrick 1998; Moñino 2004), which often rely on “striking visual contrasts and patterns” as Wierzbicka (2008: 415) puts it for Warlpiri, or culturally salient objects (e.g. cattle, dyeing techniques) (see e.g. Dimmendaal 1995, 2015; Malt & Majid 2013), rather than colours of the spectrum as thought of in a European culture. Nevertheless, considering these systems would have led us far beyond the reasonable page limits of an article, not to mention that most of the languages of our sample are far from benefitting from detailed descriptions of the native colour and visual sense systems. Our method of investigation also is admittedly biased by the translation equivalents, not to mention that we rarely know anything about the lexicographers’ method of investigation, nor how the speakers precisely partition the colour spectrum. Nevertheless, together with Evans (2012: 9), commenting on Berlin and Kaye’s methodology, we believe that our quantitative approach of these eleven notions is also “a necessary step(...) if cross-linguistically comparable data are to be obtained.”

For the purposes of the present paper, we compiled a sample of 401 languages corresponding to 434 sources, meaning that a few languages are represented by more than one source, either because the different sources show significant variation, or because they represent different dialects. We tried, as far as possible, to keep a balance between the different families and groups as shown in the following list:

- 289 Niger-Congo (NC), approx. 17% of NC languages: 16 Adamawa, 17 Atlantic-Bak, 22 Atlantic-North, 106 Benue-Congo, 11 Dogon, 19 Gur, 2 Ijoid, 6 Kordofanian, 8 Kru, 15 Kwa, 36 Mande, 8 Mel, 19 Ubangi, 4 Unclassified (Gola, Limba, Sua, Bangeri-me)
- 65 Afro-Asiatic (AA), approx. 17% of AA languages spoken in Africa: 5 Berber, 24 Chadic, 18 Cushitic, 8 Omotic, 10 Semitic
- 42 Nilo-Saharan (NS), approx. 18% of NS languages: 1 Berta, 14 Central Sudanic, 16 Eastern Sudanic, 3 Komuz, 1 Maban, 2 Saharan, 1 Unclassified (Shabo)
- 3 Khoisan (K), approx. 11% of K languages: 2 Khoe-Kwadi, 1 Non-Khoe⁴
- 6 “Varia”: 4 Songhay, Sandawe, Laal.⁵

The sample consists of two sets: (i) 238 languages were extracted from the *RefLex* lexical database for African languages (Segerer & Flavier 2011-2019), in which we could conduct semi-automatic searches; (ii) in order to get a better coverage of the genetic diversity, we added a second set, which contains another 163 languages, and for which we had to perform manual searches (for details see Segerer & Vanhove 2019). The geographical distribution of the language sample is shown in Figure 1.

⁴ Khoisan languages are no longer considered a valid genetic unit (Güldemann ed. 2018). This label is used for sake of convenience, as it is still widespread in the linguistic community.

⁵ The full list of references, languages, and the data are available online at <http://reflex.cnrs.fr/Couleurs/>. This database is still growing. The data used for this paper are those available as of July, 2019.

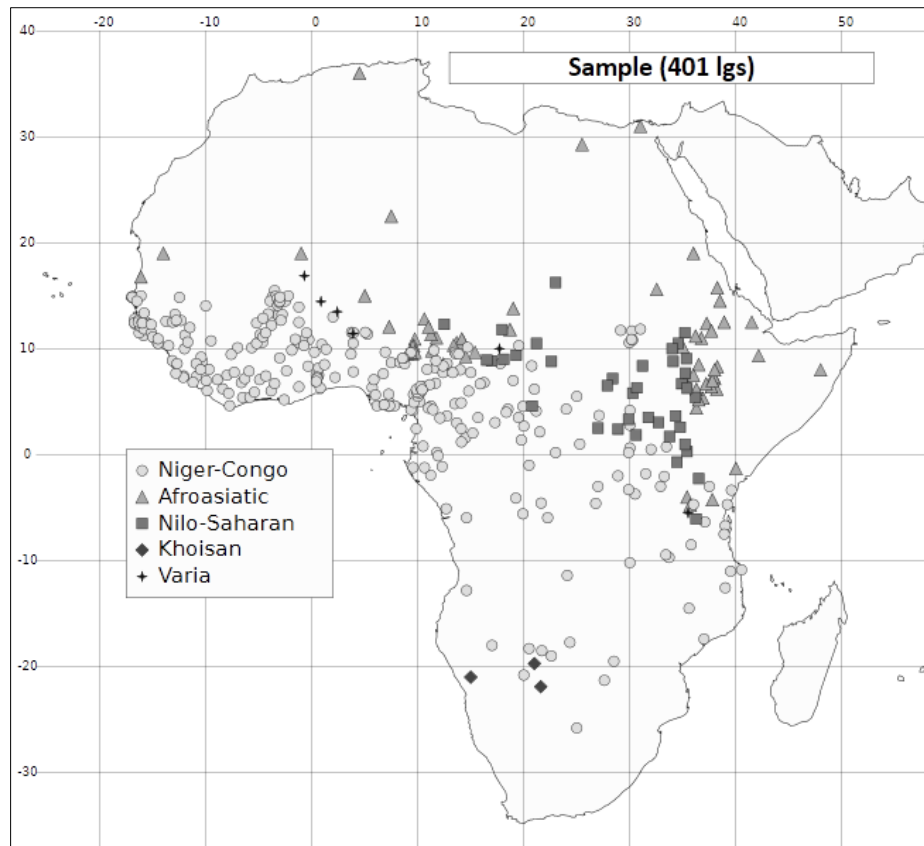


Figure 1: The language sample

RefLex (Segerer & Flavier 2011-2019) is an online cross-linguistic lexical database of African languages (that thus far excludes Berber and Semitic groups) with integrated access to the original sources (which makes the database as (un)reliable as the sources themselves and allows easy checking).⁶ Even though it was initially built for historical comparative purposes, not for the specific study of polysemy or colexification patterns, *RefLex* contains a set of integrated tools for the automatic retrieval of polysemy and homophony patterns across languages, with an association between words and concepts *via* a “unified translation”, making it a rich source for typological studies. A map construction tool is also associated with the *RefLex* database, making it easy to visualize genetic and areal patterns. As of July 2019, the database contains 1,228,194 lexical items covering 836 languages from 1,381 sources. The estimated number of African languages is roughly 2,400. The total amount of entries per language ranges from a few dozens for a handful of them to 10,000 to 23,000 for the largest dictionaries.

We largely relied on a bottom-up approach, i.e. we searched the language sample for each of the eleven colour terms and also recorded all the other meanings provided under each colour term entry. Colour terms translated by two or more different colour terms in a European language were recorded for each translated meaning, and counted as many times as needed. By systematically comparing the colexifications, we came across recurrent patterns of polysemy and lexico-constructions, only a few of which turned out to be interesting from an areal point of view.

Using a very large balanced sample made it relatively easy to distinguish contact-induced similarities from genetic-induced similarities. The online map tool which represents the various

⁶ The database is freely accessible online at www.reflex.cnrs.fr.

genetic stocks to which each individual language belongs by means of different colours is also a beneficial feature.⁷ The first author manually extended the map tool to all the languages of the sample that are not yet included in the *RefLex* database.

A word of warning before turning to the case studies: the reader should keep in mind that working on dictionaries is not the same as working with first hand language data. Therefore, in this paper, statements such as “in language X, there is such and such colour term” must be understood as a shortcut for “in the lexical sources consulted about language X, there is a translation in a European language of such and such colour term”.

3. ‘Yellow’ and ‘locust-bean tree’

In many languages of the world, botanical terms are the source of colour terms (Mollard-Desfour 2008), and the languages of Africa are no exception.⁸ By far the most frequent source from this semantic domain in our language sample is the locust-bean tree (*Parkia biglobosa*, *nééré* in West African French), which colexifies with ‘yellow’ (see Table 1). This metaphorical extension may be explained by the importance of the tree’s fruit for nutrition and traditional medicine in Western and, to a lesser extent, Central Africa. Of course, such a colexification occurs only in the geographical zones where the tree grows (see Hall *et al.* 1997 for its habitat). Such is the case for 74 languages in our sample,⁹ which belong to various language groups and genetic phyla: 5 Chadic (AA), 8 Atlantic-Bak (NC), 7 Atlantic-North (NC), 2 Benue Congo (NC), 11 Dogon (NC), 13 Gur (NC), 1 Kwa (NC), 22 Mande (NC), 1 Mel (NC), 1 Unclassified (NC), 2 Central Sudanic (NS), 1 Songhay (Varia).

The terms used for the colour denomination come from the name of the fruit, its dry or fresh pod, the yellow powder contained in the pod, or may be related, via a noun class prefix, to its leaves. The colexification may be a strict one, i.e. using the same lexeme for either meaning, or a loose one, involving a noun phrase or a different class prefix. Table 2 provides some examples, which show the variety of constructions with their exact semantic sources.

<i>Language</i>	<i>Glottocode</i>	<i>Country</i>	<i>Form</i>
Shanga (Mande, NC; McCallum Jones 2017)	shan1282	Nigeria	<i>kpàà’i</i> (<i>kpàà</i> ‘locust bean’, <i>i</i> ‘water’) ¹⁰
Kabiye (Gur, NC; CLNK 1999)	kabi1261	Togo	<i>sotv lm</i> (<i>sotv</i> ‘locust bean’, <i>lm</i> ‘water’)
Zodi (Chadic, AA; Caron 2002)	dass1243	Nigeria	<i>fà bèlki</i> (<i>fà</i> ‘water’, <i>bèlki</i> ‘locust-bean tree’)
Southern Samo (Mande, NC; SIL Burkina 2003)	sout2844	Burkina Faso	<i>kùsi</i> (‘dry locust bean pods’)

⁷ Reproduced in black and white in the present paper.

⁸ This is not only because plants are used in dyeing technology, as one of the anonymous reviewer suggests following a remark in Malt & Majid (2013: 586). Plants can be well known for other salient cultural practices, such as cooking and traditional medicine, as is the case for the locust bean. But one should note that botanical terms unused in human technologies, or not culturally salient for whatever reason, can also be extended to colour naming, as for instance ‘violet’, a flower of no particular importance in France or Great Britain, which nevertheless is the source of a colour term referring to the purple colour.

⁹ In one language (Bedik, bedi1235, Atlantic North, NC) spoken in Senegal, the term is translated by ‘orange colour’, and in another, Chumburung (chum1261, Kwa, NC) spoken in Ghana, ‘brown’ is expressed by ‘water of locust-bean flour’ while ‘yellow’ is expressed by ‘locust-bean flour for dawa-dawa’ (dawa-dawa is a food flavouring made from the seeds and the pulp of locust beans).

¹⁰ For details about the constructions with ‘water’, see Section 4.

Yanda Dom (Dogon, NC; Heath 2013)	yand1257	Mali	<i>lòl-púřá</i> ‘yellow flour from fruit of néré tree bean’ (<i>lòl</i> ‘locust bean’, <i>púřá</i> ‘flower, powder’)
Sar (Central Sudanic, NS; Palayer 1992)	sarr1246	Chad	<i>ndùjā mātā</i> (<i>ndùjā</i> ‘flour’, <i>mātā</i> ‘locust bean’)
Konyagi (Atlantic, NC; Sachot (Santos) 1996)	wame1240	Senegal	<i>yà-yéy</i> (‘locust tree foliage’; <i>yà</i> is a noun class prefix, <i>yéy</i> is the stem of ‘locust bean’).

Table 2: Examples of ‘locust bean’ > ‘yellow’ in the languages of Africa

The spatial extension of the ‘yellow’ meaning is shown in Figure 2 (which only displays the languages of our sample that are spoken in the area of the natural habitat of the locust tree).¹¹

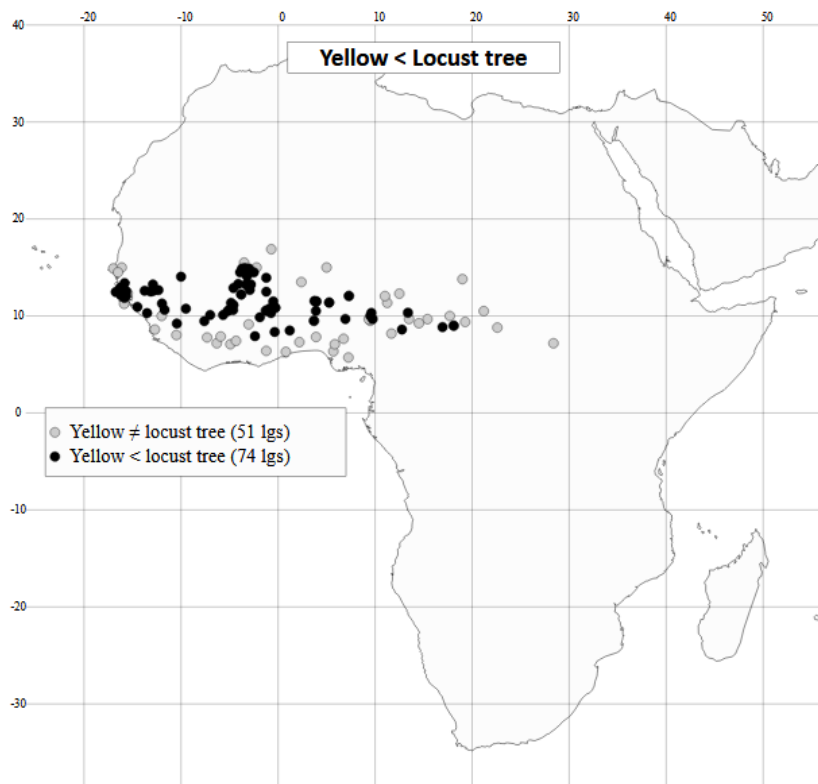


Figure 2: ‘Yellow’ and the locust tree in the languages of Africa

Interestingly, the 51 languages in which ‘yellow’ is not related to any parts of the locust-bean tree are situated at the margins on the map above (grey dots). Therefore, the black dots are located in what is the core area of the natural habitat of the locust-bean tree.¹² More information is needed to know whether the absence of colexification correlates with a lesser importance of this plant in the local diets or medicine.

When one looks at the actual words that denote the locust bean (fruit, pod, flour...) in the area (as exemplified in Table 2),¹³ what is striking is their formal diversity. This rules out the

¹¹ Note that Figure 2 concerns the number of LANGUAGES, not of sources (as in Table 1), hence the difference between the number of tokens in Table 1 (88) and the number of tokens in this map (74).

¹² The linguistic data are far from exhaustive as compared to the botanical data, but the general shape of the distribution is accurate enough.

¹³ The full list of languages used on this study is available at <http://reflex.cnrs.fr/Couleurs/>.

possibility of massive borrowings and lexical diffusion. It either may be that the metaphor developed independently (with possible sporadic borrowings, such as Pular [pula1262, Atlantic-North, NC, Guinea] *nete*, from Mandinka [mand1436, Mande, NC, Senegal] *nètè*), or that the semantic pattern alone, but not the matter, spread from an unknown centre.

The colexification of ‘yellow’ with a part of the locust-bean tree is thus a clear example of a connection between a shared physical environment, cultural practice, local diets and traditional medicine, and an areal pattern of colexification.

4. ‘Water of’ and ‘colour of’

The shared lexico-constructional pattern for colour naming in Africa we discuss in this Section is rarer than the colexification pattern discussed in Section 3 above. The “water” construction is based on noun phrases, either compounds, comitative, similitive, possessive or genitive structures, whose head is the word for ‘water’ (Seeger & Vanhove 2019), except in one language, Hausa, where ‘water’ is the dependent (see below). For a small minority of the languages, the lexicographers provide additional translations: ‘any liquid’, ‘juice’, and/or also ‘rain’, ‘river’, ‘lake’, ‘humidity’ (Bambara, Dii, Lingala, Mofu-Gudur, Sango), but we have no way to find out whether this could be the case for the other languages. For some languages, the dictionaries mention a different word for ‘juice’, but most often such a meaning (or any of the other meanings) is just unrecorded. It should be noted also that in the languages used in this study, the words for ‘water’ are not listed with the additional meaning of ‘paint’ or ‘dye’. However, two sources specify that the plant in the colour expression is used for dyeing purposes (Zaar *zà-k gáluirā* ‘blue’), or as a yellow paint (Illo-Busa *kpà’i* ‘yellow paint’), and one source that the ‘powder’ source term is used to dye sarongs and wash clothes (Fon *fě sin* ‘blue’). Even if the reviewers consider these as possible explanations for the rise of the constructions with ‘water’, we have no way to prove that these could be generalised to the majority, or all, of the languages. In Table 4 we only kept the ‘water’ translation in the literal translation column. This pattern concerns seven of our initial set of eleven colour terms, with the exception of ‘grey’ and the three colour terms attested for all languages in the database, ‘white’, ‘black’ and ‘red’ (the most “basic” ones cross-linguistically according to Berlin & Kaye; there is a shade of ‘red’, namely ‘dark red’, that use this pattern). The full set contains 36 occurrences in 28 languages of different phyla and groups (presented in Table 3), i.e. 7% of the language sample.

Language group	# languages	# colours
Chadic (AA)	8	13
Mande (NC)	8	9
Kwa (NC)	3	4
Benue-Congo (NC)	2	3
Adamawa (NC)	1	1
Atlantic (NC)	1	1
Gur (NC)	1	1
Ubangi (NC)	1	1
Central Sudanic (NS)	1	1
Semitic (AA)	1	1
Songhay (Varia)	1	1

Table 3: ‘Water of *X*’ > colour *X* per language group

Table 4 presents the data organised by language stocks and groups. It shows the clear predominance of the pattern in two unrelated groups, Chadic (AA) and Mande (NC), in this order, and the sporadic attestations in other neighbouring language groups. It also shows that ‘green’ and ‘yellow’ (11 tokens) are the most frequent colours or shades of colour expressed

with the “water” construction, followed by ‘brown’ (4 tokens), ‘orange’ (3 tokens), ‘blue’ (2 tokens), ‘dark red’ and ‘purple’ (1 token each). The other tokens are translated by at least two different colour terms: ‘yellow’ and ‘orange’ (2 tokens), ‘yellow’ and ‘brown’, ‘orange’ and ‘pink’, ‘purple’ and ‘pink’, and ‘orange’, purple’ and ‘pink’ (1 token each).

Stock	Group	Name	Glottolog index	Country	Form	Translation	Literally
AA	Chadic	Hausa	haus1257	Nigeria	<i>algashi mai-farin ruwaa</i>	light green	green with white of water
					<i>jaa mai-bak'in ruwaa</i>	dark red	red with black of water
					<i>zè màlkàṅ</i>	yellow	water of locust bean
		Geji	say1246	Nigeria	<i>zà kà wàldǎ</i>	green	water of leaves
					Kapsiki	psik1239	Cameroon, Nigeria
		Mofu-Gudur	mofu1248	Cameroon			
					Ngas	ngas1240	Nigeria
		Zaar	say1246	Nigeria			
						<i>zà-k nāállē</i>	dark brown
		Zodi	dass1243	Nigeria		<i>fà bàkàṅ</i>	green
Nigerian Arabic ¹⁴	diii1241				Nigeria	<i>fà bètki</i>	yellow
		Adamawa	Dii	Cameroon		<i>alme tamali</i>	yellow, orange
Atlantic-North Benue-Congo	Fulfulde (Adamawa) Lingala				adam1253 ling1263	Cameroon Democratic Republic of the Congo, Central African Republic	<i>mam gùù ~ sud gùù</i>
		Benue-Congo	Kol	kolc1235			Cameroon
					<i>mái ma malála</i>	orange, pink	water of orange
					<i>ñuûn mǎdibá má kwem</i>	green	water of cassava leaves

¹⁴ Nigerian Arabic has no Glottocode.

¹⁵ All the elements of this compound noun are loan words from Hausa.

Mande	Bambara	bamb1269	Mali	<i>wòròjì má</i>	orangish, purple, pinkish	kola nut.water.like
				<i>lánígùjì</i>	to have hints of green	factitive+leaves.water
Gur Kwa	Bobo Ilo-Busa	sout2840	Mali	<i>sàrā-zū</i>	green	bean leaves-water
		busa1253	Nigeria	<i>kpà'i</i> <i>lá'i</i>	yellow green	locust.bean.water leaves.water
	Jula Mandinka Maninka (Kankan)	woje1238	Ivory Coast, Mali	<i>wórójírámà</i>	reddish brown	kola nut.water.like
		mand1436	Senegal	<i>kùrùùjì</i>	brown	kola.water
		east2426	Guinea	<i>wòrojíláma</i> <i>wòrojílama</i> <i>jàlafin</i>	orange dark orange	kola nut.water.like kola nut.water.like dark
	Shanga Kyenga Kabiye Adele Chumburung	shan1282	Nigeria	<i>kpāa'i</i>	yellow	locust.bean.water
		kyen1242	Nigeria	<i>kɔ̃nɔ'i</i>	yellow	locust.bean.water
		kabi1261	Togo	<i>sɔtɔ lum</i>	yellow	locust bean-water
		adel1244	Ghana, Togo	<i>èkònlé ù bùtù</i>	yellow	tree.sp its water
		chum1261	Ghana	<i>lɔŋkɔnkyuronkyu</i> <i>pawurankyu</i> <i>nkyɔnɛ nkyu</i>	purple purple, pink brown	rabbit-water-in-water (?) powder.water locust bean flour water
Ubangi Central Sudanic NS Varia	Fon	fonn1241	Benin, Togo	<i>fé sìn</i>	blue	blue powder water
	Sango	sang1328	Centrafrican Republic	<i>ngú-ngùnzā</i>	green	water-cassava leaf
	Sar	sarr1246	Chad	<i>māñ yādā</i>	yellow	water-tree sp.
	Dendi	dend1243	Benin, Niger, Nigeria	<i>sùbù hářì</i>	green	grass water

Table 4: The colour construction “water (of/with/its) X” in African languages

Figure 3 visualises the distribution of the pattern in Africa and shows a clear cluster in West and Central Africa.

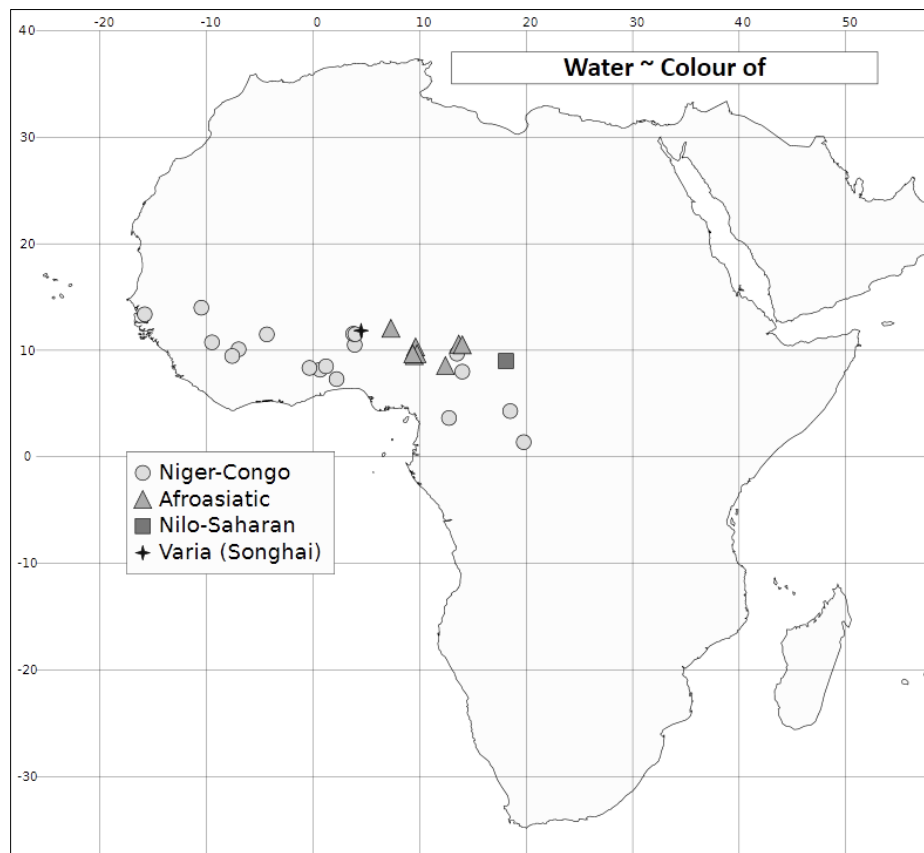


Figure 3: Distribution of the construction “water (of/with/its) X” per language phyla

In addition to the data in Table 4, the construction used in Hausa (haus1257, Bargery 1934) is noteworthy. As mentioned above, this language is distinctive in using a “water” construction in which ‘water’ is not the head, but the dependent of the noun phrase. Hausa is also the sole language in which the colour terms ‘white’ or ‘black’ are the heads. The constructions function as attenuative and intensive devices: e.g. *algashi mai-farin ruwaa* ‘light green’, lit. ‘green with white of water’, *jaa mai-bak’in ruwaa* ‘dark red’, lit. ‘red with black of water’. *ruwaa* ‘water’ is also used as a generic term for ‘colour’ and ‘shade of colour’ (Bernard Caron, p.c.; this meaning is not recorded in Bargery [1934] who only mentions the Arabic loan *launi* ‘colour’). Among the Arabic varieties of our language sample, only the one spoken in Nigeria (Kaye 1986: 7) shows the “water” construction, without borrowing of matter, pointing to a borrowing of a structural pattern from (an) unidentified surrounding Chadic language(s). A slightly different scenario occurred for Fulfulde, which borrowed the matter from Hausa, but followed the pattern of other NC and AA languages of the area, not that of Hausa.

Except for the two vehicular languages, Sango (Bouquiaux *et al.* 1978) and Lingala (Kawata 2004), and all but one Mande language, the languages are or have been in contact with each other (or with one other language of this set at least), so that a convergence phenomenon could explain this shared lexico-constructional pattern. A possible initial input could be Chadic (AA), in particular Hausa, which is a very widespread, dominant and expanding vehicular language spoken across West and parts of Central Africa. Dmitry Idiatov (p.c.) hypothesizes that such is the case for Illo-Busa, a Mande language of Nigeria, where Hausa is spoken as *lingua franca*. But the fact that in Hausa the syntactic pattern deviates from the common “water-as-head” pattern weakens the hypothesis. Moreover, the pattern is so pervasive in the area that even in

the case of direct borrowings from Hausa, as in Fulfulde, a language spoken by nomadic herders, the Hausa construction is not retained. Further research is needed about the precise sociolinguistic settings that led to the creation of this areal feature.

Independent parallel development is an alternative explanation that does not preclude the preceding one, especially for the other languages with no recorded contact with any other language sharing this feature. It remains striking that the pattern only occurs in a very limited area in Africa.

This shared constructional pattern may thus reflect both a phenomenon of convergence and a universal tendency, depending on the historical and sociolinguistic contexts.

Each language of this cluster developed (in addition to other colour terms) its own colour term inventory with the “water” construction”, not necessarily having exactly the same morphosyntax, nor covering the same part of the spectrum. For instance, three Mande languages, Bambara (Dumestre 2011), Jula (Braconnier 1999) and Maninka of Kankan (Diané 2012), add to the construction the Mande similitive suffix *-lama* / *-rama* / *-ma* (tones depend on the language) (Segerer & Vanhove 2019). On the other hand, Zaar (Chadic, AA; Caron 2005) is the language which has the highest number of colour terms denoted with the “water” construction (‘yellow’, ‘green’, ‘blue’ and ‘dark brown’), and they are the sole denominations for these colours recorded in the dictionary. Most of the other languages only have one colour term of this type, a few have two, and only one language (Chumburung, Kwa, NC; Hansford 1989) has three.

As is often the case cross-linguistically (see e.g. Wierzbicka 2008), most of the languages in Table 4 lack a term for the concept of ‘colour’, and only two languages in our sample have colexified ‘water’ and ‘colour’, Illo-Busa and Hausa.¹⁶

5. ‘Green’ and ‘unripe’ vs ‘red’ and ‘ripe’

In our language sample 401 languages have a term for ‘red’, and only 281 for ‘green’.

This section deals with two parallel colexification patterns (strict and loose) – ‘green/unripe’ and ‘red/ripe’. While the first one is well known from the European languages, the other one seems to be absent from them. African languages, on the contrary, manifest both colexification patterns, although the two differ in their distribution.

The prevalence of ‘green/unripe’ colexifications in European languages creates methodological problems for searches in the database. Translations such as ‘green, unripe’ are very frequent in the original sources, so it is not always clear whether the words translated as ‘green, unripe’ may apply to any object with the colour meaning only. Therefore, in order to avoid inconsistencies, we adopted the following practice: if a source contains a word translated as ‘green, unripe’ without further details, and contains another word with the meaning ‘green’, the former is considered as having ‘unripe’ as its primary meaning, the meaning ‘green’ being a side effect of the translation into a European language. The following are some examples: Bokobaru (boko1267, Mande, NC; Jones 2004) *ísi* ‘unripe; green’, *lá’i* ‘green’; Nzema (nzim1238, Kwa, NC; Aboagye 1992) *amunli* ‘not ripe; green’; *bɔnyema* ‘green’. In the other

¹⁶ One of the anonymous reviewers asks, “why do the constructions WATER + OBJECT OF A CHARACTERISTIC COLOUR yield terms for that colour at all? What is the relevant cognitive and cultural underpinning?” Unfortunately, we don’t know of any study that could help us understand the relevant cognitive and cultural underpinning of these expressions for any of the languages concerned. But the fact that two lexicographers specify technical uses for dyeing and painting in Zaar (Chadic, AA), and Illo-Busa (Mande, NC), is probably a strong cultural underpinning for the relevant colour terms in each language. Note, however, that these two languages have other colour terms based on this construction, for which no cultural usage is mentioned.

cases, we considered that both meanings are valid. Of course, this strategy may lead to some mistakes, but it is a consistent one.

In some sources, the expressions for ‘unripe’ and ‘green’ are related. In these cases ‘green’ is usually derived from ‘unripe’: Sheko (shek1245, Omotic, AA; Hellenthal 2010) *xhir* ‘be fresh, unripe, wet’ > *xhirnsh* > ‘green, unripe’; Grebo (sout2826, Kru, NC; Innes 1967) *kyãð* ‘state of being uncooked, uncooked condition, unripe condition’ > *ɲwa kyãð* ‘fresh herbs, greenness, green’ (lit. ‘green bush’); Kim (krim1238, Mel, NC; Childs 2012) *hòn* ‘unripe’ > *hòntòlò* ‘green’; Bangime (bang1363, isolate, NC; Hantgan 2012) *káráá* ‘unripe (grain)’ > *gúzè káráá* ‘green’.

The same applies for ‘red’ and ‘ripe’, but the derivation of ‘red’ from ‘ripe’, when it was possible to figure it out with enough confidence, only concerns six Kwa (NC) languages, e.g. Adele (adel1244, Kwa, NC, Rongier 1988) *sìnà* ‘it is ripe’, *sínán* ‘red’.¹⁷

In line with the definition of colexification, and more precisely here loose colexification, we coded these terms as EQUAL.

In some rare cases, one is led to consider that a source has distinct words even if there is a word with both translations. For instance Sembla (seek1238, Mande, NC; Prost 1971) has *mwòn* ‘be ripe, be cooked’; *ɲá* ‘red, ripe, be red, ripe’; *sò* ‘ripen’ and *sèn* ‘red, be red’. The word *ɲá* is translated as both ‘red’ and ‘ripe’ but commented as only applying to fruits that become red when they ripen. Since other words are more specific, we counted these sources as DISTINCT. While it may be relatively easy to state that a language (in fact a lexical source) has a word with the meanings ‘green’ and ‘unripe’ (or ‘red’ and ‘ripe’), it is often the case that a source lacks one of the two meanings, either because the source itself is not rich enough, or because of a lack of accuracy. Therefore, it is sometimes impossible to decide whether the language in question illustrates the colexifications or not. Such cases (mentioned as N/A in Figures 6 and 7 below) were left out of consideration.

To sum up, the following decisions were made:

- If a source has a word translated as ‘red’ and not ‘ripe’ (resp. ‘green’ and not ‘unripe’) AND a word translated as ‘ripe’ and not ‘red’ (resp. ‘unripe’ and not ‘green’), then it is coded as DISTINCT for the pair in question, even if there is another word for which both translations are given.
- If, on the contrary, a source does not show distinct words as defined above, then it is counted as EQUAL.
- Sources that show two morphologically related forms for the two meanings are coded as EQUAL.
- Sources that have a word for ‘(un)ripe’ and another one for ‘red’ or ‘green’ for which the translation mentions ‘(un)ripe (for certain fruits)’ are coded as DISTINCT.
- Sources where one meaning is absent are discarded (N/A).

In our sample of 401 languages, 350 could be tagged for either ‘green/unripe’ or ‘red/ripe’ or both, meaning that we lack information about potential colexifications in 51 languages. Languages tagged as N/A are of two kinds: those that either lack a term for ‘green’ or for ‘unripe’ are N/A for the pair ‘green’/‘unripe’ (207 cases); those that lack a term for ‘ripe’ are N/A for the pair ‘red/ripe’ (54 cases). So, since ‘green’ as a colour term has a limited distribution compared to ‘red’, it comes as no surprise that only 194 languages could be tagged for ‘green’/‘unripe’, as against 347 for ‘red’/‘ripe’. These metrics are detailed in Table 5, which

¹⁷ The exact grammatical categories of ‘red’ cannot be deduced from the sources, but unlike the case of ‘ripe’, their Kwa equivalent is never translated with a stative verb.

also shows that only a handful of languages (7) have both colexifications, but that 138 share the absence of colexification for both colour terms.

	'green' = 'unripe'	'green' ≠ 'unripe'	N/A for 'green'	TOTAL
'red' = 'ripe'	7	21	31	59
'red' ≠ 'ripe'	25	138	125	288
N/A for 'red'	0	3	51	54
TOTAL	32	162	207	401

Table 5: Colexification tagging of 'green/unripe' and 'red/ripe'

The two maps in Figure 6 and 7 show the distributions of 'green/unripe' and 'red/ripe', respectively. White dots show languages where no tagging could be made because of insufficient information. Grey dots show languages where there are distinct words for the two meanings, and black dots show languages with one word for the two meanings.

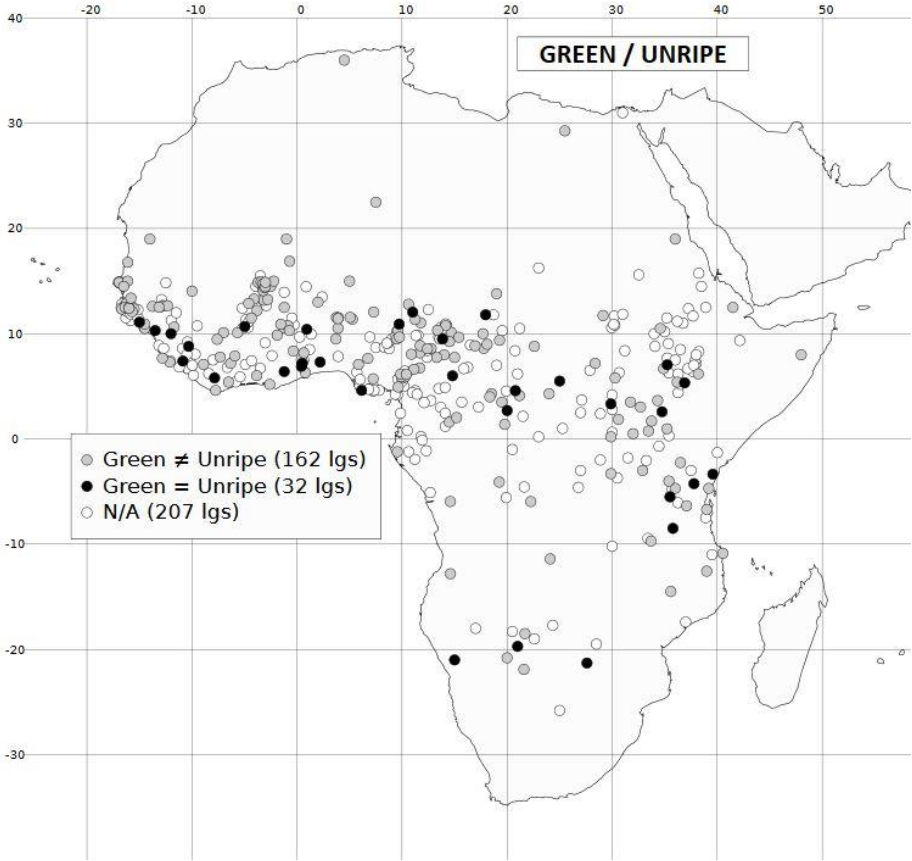


Figure 6: Colexification of 'green' and 'unripe' in the languages of Africa

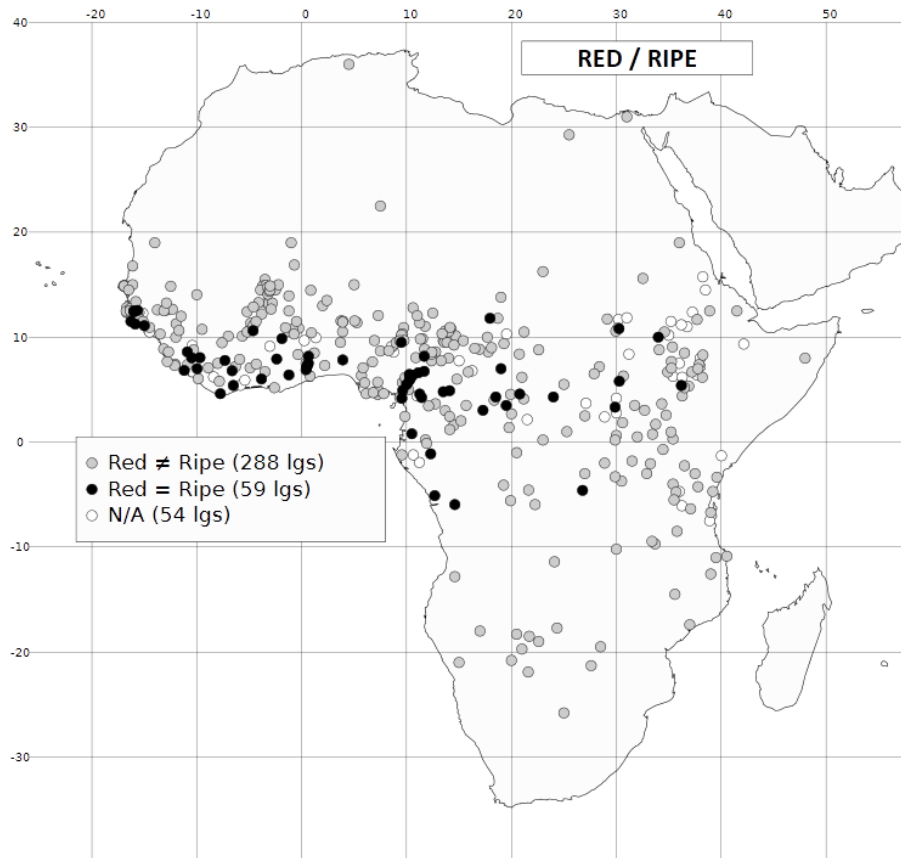


Figure 7: Colexification of 'red' and 'ripe' in the languages of Africa

The first finding is that languages that exhibit colexifications are relatively rare: 16% for 'green/unripe' and 17% for 'red/ripe' (based on the number of relevant languages). This is an unexpected finding even for a lexicologist who is accustomed to scrutinizing loads of dictionaries for decades, as is the first author of this article.

The second finding pertains to the geographical distributions. While the less frequent 'green/unripe' colexification is attested all over the continent, the more frequent 'red/ripe' colexification seems to be restricted to what is now called the Macro-Sudan Belt (Güldemann 2007),¹⁸ making it a feature to add to the bundle of features which serves to characterise this linguistic area in Africa. So far, it is the sole lexico-semantic feature that would contribute to characterise this area.

These distributions, because they are very different, reinforce each other in terms of reliability: since the possible biases in language tagging are the same for both pairs, there can be no good reason why only one of the distributions would be triggered by these biases more than the other. On the contrary, if one would have only looked at 'red/ripe' for instance, the limited extension of the colexification could be said to be the consequence of repeated errors in language tagging.

¹⁸ The Macro-Sudan Belt is "a broad belt south of the Sahara ... sandwiched between the Atlantic Ocean and the Congo Basin in the south and the Sahara and Sahel in the north, and spans the continent from the Atlantic Ocean in the west to the escarpment of the Ethiopian Plateau in the east. ... The core of the identified area is formed by the following language families: Mande, Kru, Gur, Kwa, Benue-Congo (excluding Narrow Bantu), Adamawa-Ubangi, Bongo-Bagirmi, and Moru-Mangbetu. The two easternmost families of Niger-Congo, Benue-Congo and Adamawa-Ubangi, as well as the two Central Sudanic families, Bongo-Bagirmi and Moru-Mangbetu" (Güldemann 2007: 151-152) In this area, some unrelated linguistic features, in genetically unrelated languages, are more frequent than in other parts of Africa. These features include e.g. logophoricity and labio-velar consonants. See Güldemann (2007) for details.

Moreover, since the number of languages having a word for ‘green’ (281) is noticeably lower than the number of languages with a word for ‘red’ (401), the distribution of the ‘green/unripe’ polysemy is more likely to be geographically restricted, which is not the case, and cannot thus be attributed to a bias of our language sample.¹⁹

The geographical difference between the two colexifications reveals that they are different in nature. The ‘green/unripe’ colexification is randomly distributed, whereas the ‘red/ripe’ one is areal, being limited to the Macro-Sudan Belt. This contrast points to the fact that these two pairs of meanings, which were chosen because of a perceived symmetry, are in fact clearly distinct. For example, a fruit can become ripe, but it cannot become unripe; you can ripen a fruit (by putting it in the appropriate conditions), but you cannot unripen a fruit. Therefore, it comes as no surprise that ‘red’ is much more often lexicalised as a verb than ‘green’ as shown in Table 6.

	# items	# items with part-of-speech spec.	# verbs
red	1681	809	334 (41%)
green	716	442	83 (19%)

Table 6: % of verbs for ‘red’ and ‘green’

Actually, the choice of these two pairs of meanings reflects a European bias, where (i) the category ‘colour’ is commonplace, (ii) the colexification ‘unripe/green’ is frequent, and (iii) ‘unripe’ is often (always?) derived from ‘ripe’. This bias favors an impression of symmetry (actually not precisely reflected in European languages for which we are not aware of a colexification ‘red/ripe’).²⁰ None of the three criteria are applicable to African languages. Still, the geographical distributions of the colexifications are interestingly divergent.

These colexifications were compared with the results obtained via the CLICS (List *et al.* eds. 2018) database.²¹ A strikingly similar pattern arises in two other geographical areas. The colexification of ‘red’ and ‘ripe’ is only recorded in 28 languages, seventeen in different stocks in Amazonia, and in eleven Papuan languages (ten Nuclear Trans New Guinea, and one Timor-Alor-Pantar). The colexification of ‘green’ and ‘unripe’ is much more widespread across the world (84 languages), but it occurs in the same Amazonian languages as ‘red’ and ‘ripe’ and not at all in the Papuan languages of the CLICS database. Taking into account the fact that CLICS contains far less languages and data than our language sample, we can tentatively propose that the African ‘red/ripe’ – ‘green/unripe’ asymmetry may be reflected worldwide (although there are proportionally more attestations of ‘green/unripe’ in CLICS, probably because of the linguistic coverage bias). The reasons for this asymmetry cannot be deduced from the database and further research is needed.

¹⁹ One of the anonymous reviewers was not sure we were right to ascribe the ‘red’/‘ripe’ polysemy to the Macro-Sudan Belt because the number of sampled languages to the south of Africa is smaller, and the languages that show ‘green’ = ‘unripe’ are still fewer. But as the number of languages spoken to the south of the Macro-Sudan Belt is far lower than in the Macro-Sudan Belt, a possible bias cannot be ascribed to our language sample.

²⁰ The colour of ripe fruits are as diverse in Africa as they are in Europe.

²¹ The Database of Cross-Linguistic Colexifications (CLICS) (List *et al.* eds. 2018), freely accessible online, is “a computer-assisted framework for the interactive representation of crosslinguistic colexification patterns. In its current form, it has proven to be a useful tool for various kinds of investigation into cross-linguistic semantic associations, ranging from studies on semantic change, patterns of conceptualization, and linguistic paleontology.” See also Rzymiski *et al.* 2019. The CLICS database is limited to strict colexifications and contains a much smaller sample of languages (and very few languages of Africa) and far less lexical items than RefLex.

In African languages, further investigations could examine other colexification patterns related to ripeness, especially those related to the notions ‘raw’, ‘young’, ‘fresh’, ‘humid’ (for ‘unripe’) and ‘cooked’, ‘ready’, ‘hard’ (for ‘ripe’), some of which are also recorded in CLICS, but this is well beyond the purpose of this article.

6. Colour ideophones

Colour ideophones are mostly intensifiers; they are used to emphasise the hue of a colour. They can be used exclusively with one colour term in order to add emphasis. Syntactically, they often show up as modifiers of a colour term, as in Keeraak (Atlantic-Bak, NC; Segerer, pers. field notes):

- (1) *kajilɔɔnak* *kɔkɔ* *kɔmɔ* *kɔhutɔm* *far*
 rooster this COPULA white IDEO
 ‘This rooster is really white!’

In some languages, the colour ideophone may be a predicate, as in Dii (Adamawa, NC; Bohnhoff 2014: 264):

- (2) *zɔ́m* *ɔn* *mbàà* *vìm*
 empty land burn COPULA IDEO
 ‘Burnt grass is black’

In a few other languages, the ideophone modifies a stative verb whose primary meaning is one of lightness only, not including colour, as in Tima (Kordofanian, NC; Schneider-Blum 2013) - *héh* (classified as an adjective by the author) ‘light’: *àhéh k̀r̀r̀nd̀i* ‘light yellow (like millet)’, *àhéh ẁèl̀ẁèl* ‘light blue’, *àhéh k̀á̀ỳi* ‘grass green’.

In our 401 language-sample, there are 156 languages from all language phyla for which colour ideophones are recorded in the sources (Figure 4). A total of 1148 ideophones were found for the eleven colour terms under study. Most of them (1068, i.e. 93%) concern ‘black’ (365), ‘white’ (335) and ‘red’ (368).

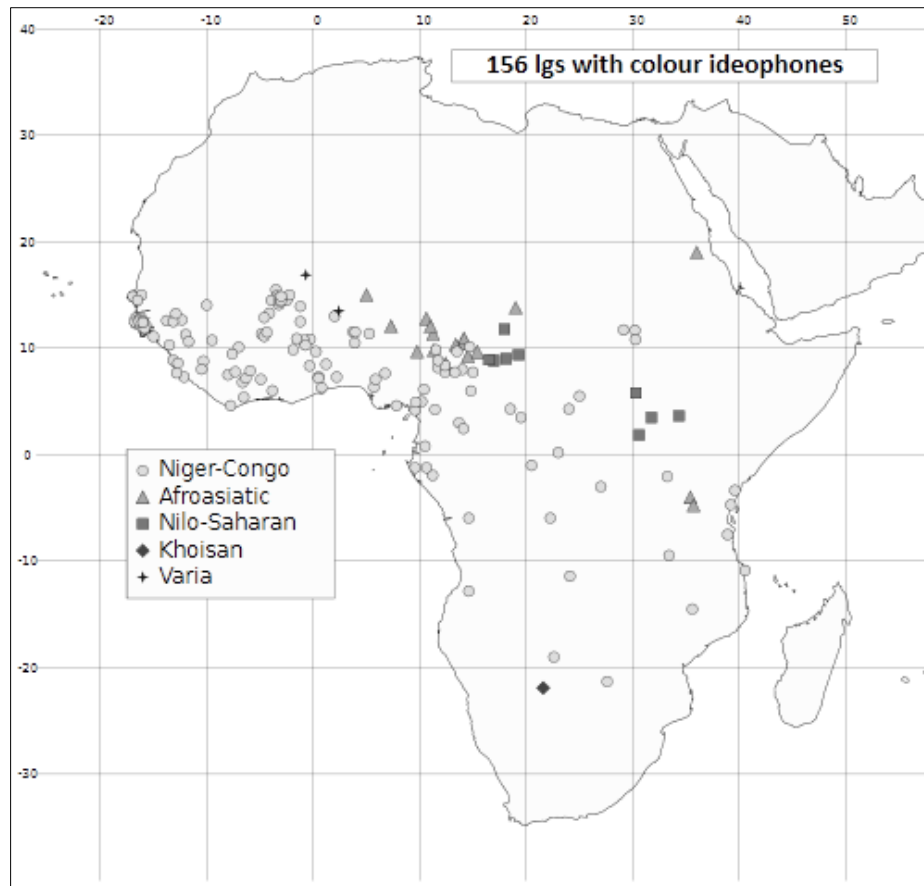


Figure 4: The distribution of colour ideophones in the languages of Africa

Three ideophones intensifying these three colours are recurrent in one region in West Africa, in four branches of Niger-Congo (Atlantic-North, Mande, Dogon and Adamawa), two branches of Afroasiatic (Chad and Semitic), and one Songhay variety (Varia). The data are presented in Table 7, and Figure 5 shows their geographical distribution.

Stock	Group		Language	Glottocode	Country	Black	White	Red
NC	Atlantic-North	Fula cluster	Fulfulde (Adamawa)	adam1253	Cameroon, Nigeria	<i>kurum</i>	<i>tal</i>	<i>coy</i>
			Fulfulde (Maasina)	maas1239	Mali	<i>kirim ~ kurum</i>	<i>tal</i>	<i>coy</i>
			Fulfulde (Nigeria)	nige1253	Nigeria	<i>kurum</i>	<i>tar ~ tal</i>	<i>coy</i>
			Fulfulde (Burkina Faso)	west2454	Burkina Faso	<i>kirim ~ kurum</i>	<i>tarr ~ tal</i>	<i>coy ~ coroy</i>
			Jaad	bady1239	Senegal, Guinea	<i>kirim</i>	<i>tal</i>	
			Sereer	sere1260	Senegal			<i>coy</i>
			Noon	noon1242	Senegal			<i>cɔy</i>
			Wolof	wolo1247	Senegal		<i>táll</i>	<i>coyy</i>
	Adamawa		Dii	diii1241	Cameroon		<i>tál</i>	<i>sóy ~ sóy</i>
			Mbum	mbum1254	Cameroon	<i>kúrúm</i>	<i>tál</i>	
			Peere	peer1241	Cameroon		<i>tál</i>	<i>sóí</i>
			Samba leko	samb1305	Cameroon	<i>kúró</i>	<i>tál</i>	
	Bantoid		Chamba	samb1311	Nigeria		<i>tál</i>	<i>sóy</i>
	Dogon		Bankan-Tey	bank1259	Mali	<i>kirim-kirim</i>		<i>cóy-cóy</i>
			Najamba	bond1248	Mali	<i>kirim-kirim</i>		
			Nanga	nang1261	Mali	<i>kúrúŋ-kúrúŋ</i>		
			Perge Tegu	jams1239	Mali	<i>kúrúm-kúrúm</i>		<i>cóy-cóy</i>
			Tommo-So	tomm1242	Mali	<i>kirim-kirim</i>		
			Toro Tegu	toro1253	Mali	<i>kirim-kirim</i>		<i>cóy-cóy</i>
	Mande		Bambara	bamb1269	Mali	<i>kirikiri</i>		<i>cóyí</i>
			Jula	woje1238	Ivory Coast, Mali	<i>kiri</i>		<i>shue</i>
			Maninka (Kankan)	east2426	Guinea	<i>kidikidi</i>		<i>cóee</i>
			Maninka (Niokolo)	west2500	Senegal			<i>cúyí</i>
			Susu	susu1250	Guinea			<i>soi</i>
	Ubangi		Gbaya (Yaayuwee)	nort2775	Cameroon		<i>tál-tál</i>	<i>sóé-sóé ~ sóí-sóí</i>
AA	Chadic		Gude	gude1246	Cameroon, Nigeria		<i>tálál</i>	

		Hausa	haus1257	Niger, Nigeria	<i>k'irín</i>	<i>tal</i>	
		Hausa (Ader)	haus1257	Niger	<i>kìrín</i>	<i>tás</i>	
		Mofu-Gudur	mofu1248	Cameroon		<i>tál tál</i>	
		Tangale	tang1378	Nigeria		<i>twal</i>	
	Semitic	Nigerian Arabic ²²		Nigeria	<i>kurum</i>	<i>tall</i>	<i>ču</i>
		Chadian Arabic	chad1249	Chad	<i>kurum</i>	<i>tal</i>	<i>ču</i>
Varia	Songhay	Zarma	zarm1239	Niger		<i>tás</i>	
		Djenné Chiini	djen1243	Mali	<i>kiriŋkirĩ, kurun</i>		<i>čoy</i>

Table 7: Borrowed colour ideophones in the languages of Africa

²² Nigerian Arabic has no Glottocode or ISO code.

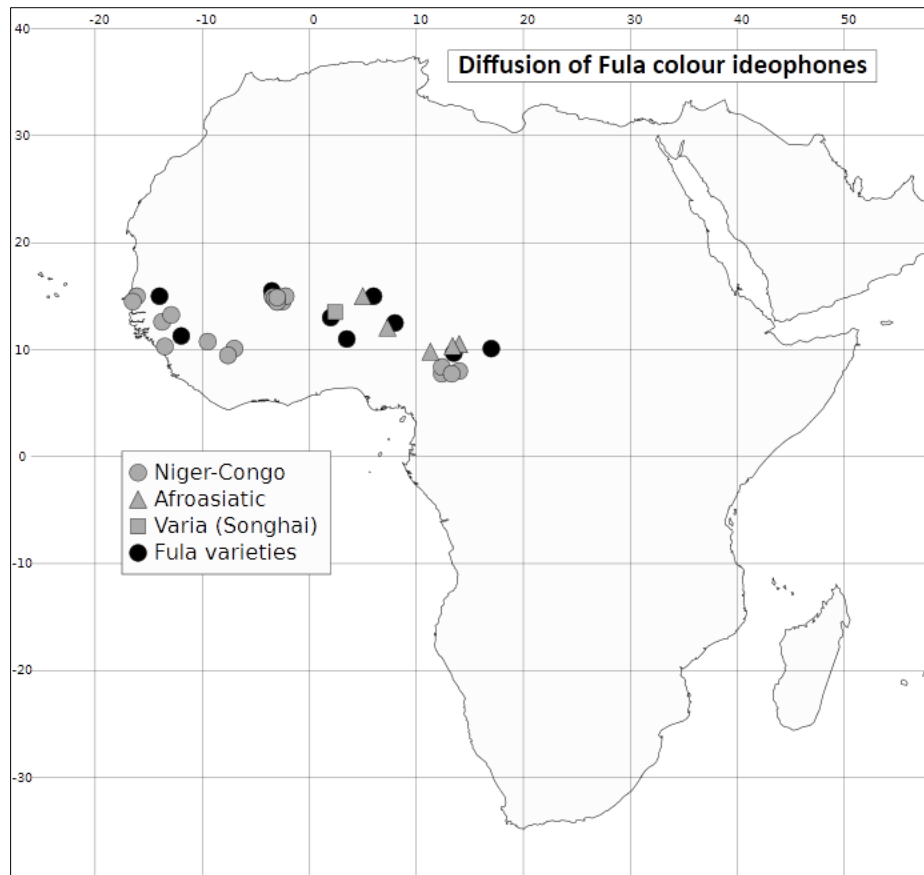


Figure 5: Diffusion of Fula ideophones *kirim* ~ *kurum* ‘black’, *tal* ‘white’, *coy* ‘red’

It is widely acknowledged that ideophones are difficult to reconstruct (see e.g. Blench 2010), and so far we have very little evidence for reconstructed ideophones in African languages, particularly in Niger-Congo. As far as we know, the existing reconstructions are only for Bantu languages (Fivaz 1963), and they are not cognates with the ones found in our language sample. Thus the sole fact that the three colour ideophones are found in different branches of Niger-Congo may be sufficient to advocate an areal hypothesis of diffusion (see also Gast and Koptjevskaja-Tamm [2018] for the colexification of ‘feather’ and ‘hair’ in Meso-America). This is reinforced by their occurrence in the genetically unrelated Chadic and Semitic languages spoken in the same region, and supported by the particular sociolinguistic setting, nomadism, a common, and by no means unusual, situation in Africa until recently. Moreover, it is possible to hypothesize on good grounds the origin of the diffusion: in each case, only Fula (Atlantic-North, NC), a cluster of languages spoken by traditionally nomadic cattle herders travelling across vast areas, are or have been in contact with all the others. Therefore, it should be considered as the source of the diffusion. Note that Nigerian Arabic (Owens 1993: 201-204), and Chadian Arabic (Jullien de Pommerol 1999b: 123), which are also partly spoken by nomads, are the sole Arabic varieties that have a category of ideophones (including colour ideophones), the result of a contact influence.

Only the languages of the Fula cluster, the source of the diffusion, and Nigerian and Chadian Arabic, in close contact with the languages of this language cluster, have all three ideophones. The other 25 languages borrowed one (10 languages) or two (15 languages) ideophones, not necessarily the same ones in the latter case (only those for ‘black’ and ‘red’ in Mande and Dogon languages, only those for ‘black’ and ‘white’ in Chadic languages, and there is more diversity in other language groups).

The loans themselves have often been adapted to the phonology and morphosyntactic patterns of each language. Reduplications in the Dogon languages, in Mofu-Gudur (Chadic, AA; Barreteau 1983) and in Bambara (Mande, NC; Dumestre 2011) are easily explained by the fact that these languages have adapted the loanwords to the (favoured) morphological pattern of their numerous local ideophones. Mande languages have also adapted the loans to each language's structure, namely their preference for the CV syllables, which led to the loss of the final consonant of *kirim*. Ideophones exist in the target languages, even for the same referents as the borrowed colour ideophones, but they may apply to different, or to a small subset of, referents. For instance, *cóee* in Maninka of Kankan (Mande, NC; Diané 2012) applies only to blood and red fruits. Our language sample did not reveal other areas of diffusion, but further research is needed to determine if this type of matter diffusion is restricted to this West African area or not.

On the other hand, ideophones, as a word category, are usually considered an areal *functional* feature of sub-Saharan Africa, although they are *formally* highly local, and most studies on ideophones claim that they are difficult to borrow if the target language already has ideophones (e.g. Childs 1994). Dingemanse (2017: 379), points to the few studies that claim that ideophones can be borrowed, namely Dravidian and Indo-Aryan languages (Emeneau 1969), Atlantic creoles (Bartens 2000), and Khoisan languages (Nakagawa 2011).²³ In spite of the poor evidence found so far, Dingemanse (2017: 379-380) nevertheless considers that the “low degree of morphosyntactic integration [of ideophones ...], should increase their borrowability” since “[o]ne of the factors influencing the borrowability of linguistic items is the degree of system integration” (referring to Matras [2007], and Enfield [2008]), and that “[t]he areal distribution and borrowability of ideophones is likely to be a fruitful locus for further research”. Another argument that could be added in favour of their borrowability is their high degree of expressiveness.²⁴ Even if we were not *a priori* expecting anything in this domain, our large language sample proved to be fruitful in this respect, hence our decision to include this topic in the present study.²⁵

7. Conclusion

Working with a method which does not have recourse to spatial auto-correlations in the data and to statistics as in Gast & Koptjevskaja-Tamm (2018), but which is in line with their overarching principle of searching large databases (not necessarily built for a typological

²³ Regarding languages of sub-Saharan Africa, the borrowings are not as obvious as Bartens and Nakagawa are claiming. Bartens's study has been rightly criticised by Childs (2002). As for Nakagawa, if he indeed claims in his introduction that “ideophones similar to Glui ideophones can be attested also in !Xóð” (Nakagawa 2011: 280), a closer look at the article shows that it is only the light verb used with ideophones that has been borrowed, not the ideophones themselves (Nakagawa 2011: 285).

It is for specialists of Dravidian and Indo-Aryan languages (which we are not) to confirm or disprove the validity of Emeneau's comparisons, but it is striking that among the 44 comparative sets, only fifteen do not concern the auditory sense, none concerns colours, and only one concerns brightness.

²⁴ One of the anonymous reviewers mentions that local varieties of French and English also use ideophones, a further proof of their borrowability. We only recently came across a publication by Souag (2013) who mentions ideophone loans from Hausa and Kanuri into Ghadames Berber in Libya (not included in our language sample). Two of them are intensifiers for ‘white’ (*fərr*) and ‘black’ (*çlək*), but they are unrelated to the three ideophones discussed here.

²⁵ One of the anonymous reviewers suggests a comparison of colour ideophones with the discussion in O'Meara, Kung & Majid (2019) on taste and smell ideophones in the Tepehua-Totonac languages, and the higher longevity and stability of taste ideophones. It would indeed be interesting to know how colour fits in this comparative modality perspective, and how our data fit into the broader picture of perceptual lexicons, but this would be the topic of additional research and another article, far beyond the scope and limits of the present paper.

purpose), we have demonstrated that it is possible to uncover areal patterns of colexifications and shared lexico-constructional patterns that were previously unacknowledged, or underestimated. The case studies illustrate five types of lexico-semantic phenomena showing parallels across languages:

- (i) similarities due to shared environments and cultural practices (the colexification of ‘locust bean’ parts and ‘yellow’);
- (ii) a shared colexification as one semantic feature of a Sprachbund (‘ripe’ and ‘red’);
- (iii) a randomly distributed colexification (‘unripe’ and ‘green’);
- (iv) shared lexico-construction patterns (‘water X’ and ‘colour X’); and
- (v) in addition to our initial aim, lexical diffusion (colour ideophones).

The case of “water” constructions is a good example of the limits and advantages in using large lexical databases for lexico-semantic typology. The limits are of course due to possible gaps. The advantages lie in the size of the data sample, which helps in identifying potential areas for shared lexico-constructional patterns, and prompts further research in order to fill the gaps and look for an explanation for areal features.

Our initial bottom-up systematic search in the *RefLex* lexical database of African languages for colour terms allowed a very large coverage of the languages of Africa in a well-balanced sample and made it much easier to distinguish the genetic inheritance factors from the areal factors of diffusion. It was complemented by extra-linguistic knowledge, and smaller scale studies (relying on our Africanist background), in order to account for and explain the similarities.

In one case this first approach was complemented by a hypothesis-driven method, which turned out to be biased by our Western view of possible polysemies (Section 5 on ‘red’ and ‘green’), but nevertheless provided a very interesting counter-intuitive result.

The borrowing of ideophones further showed that it is also possible to “stumble” upon unexpected patterns of lexical diffusion when searching a database.

Thus, combined with complementary information and methods of investigation to provide a proper explanation to a potential areal feature, this study has provided avenues of research that large lexical databases offer for lexical typology.

Having limited ourselves to the languages of Africa, it remains to be seen whether some of the patterns studied are not more widespread (or “natural”) than our sample would suggest. In this respect, our first comparative attempt in Section 5 is too limited to be conclusive. It only can be considered as an indication for further cross-linguistic studies beyond Africa. Moreover, in the instances where we could provide an explanation to the areal diffusion of a semantic feature, this is mainly based on our personal knowledge of the languages and areas in question, i.e. on micro-scale studies. The approaches, in our view, are complementary.

Of course, like any lexical source, *RefLex* hardly offers a solution to the tricky question of the influence of the source language (language of elicitation) on the target language.

As a final note, we should mention that each of the patterns studied in this paper concerns just one feature, with various geographical extensions depending on the feature. They do not form the bundle of features that are required in order to claim the existence of a Sprachbund. We want to make it clear that nothing as such can be deduced from the semantic domains that we illustrated. Nevertheless we did find a (and so far the only) semantic feature of the Macro-Sudan Belt with the colexification of ‘red’ and ‘ripe’.

Abbreviations

AA Afroasiatic; IDEO ideophone; K Khoisan; NC Niger-Congo; NS Nilo-Saharan; p.c. personal communication; pers. personal.

Acknowledgements

We gratefully acknowledge the financial support of our research unit, the LLACAN, the ANR project *RefLex* (<https://sites.google.com/site/referencelexicon/home/presentation-en>), and the CNRS *Fédération Typologie et Universaux Linguistiques* (<http://www.typologie.cnrs.fr/>). This work was also partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the program “Investissements d’Avenir” (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris - ANR-18-IDEX-0001. We are also indebted to our colleagues who provided us with data on their language(s) of expertise: Odette Ambouroué, Pascal Boyeldieu, Mark Dingemans, Dmitry Idiatov, Rozenn Guérois, Konstantin Pozdniakov, Nicolas Quint, Serge Sagna, Yvonne Treis, Mark Van de Velde, Alexandra Vydrina, and to Abbie Hantgan-Sonko for having checked our English. Our special thanks also go to the two anonymous reviewers, and to the editors of this special issue of *Linguistic Typology* for their helpful comments. Of course, as usual, all remaining errors are ours.

References

The full list of sources and languages together with their ISO code and their (under construction) Glottocode is available on line at <http://reflex.cnrs.fr/Couleurs/>.

- Aboagye, P. A. Kwesi. 1968. *Nzema-English, English-Nzema dictionary*. Accra: Ghana Publishing Corporation.
- Bargery, George Percival. 1951 [1934]. *A Hausa-English dictionary, and English-Hausa vocabulary*. London: Humphrey Milford for Oxford University Press.
- Barreteau, Daniel. 1983. *Description du mofu-gudur (langue de la famille tchadique parlée au Cameroun)* (2 vols.). Paris: Université de la Sorbonne Nouvelle (Paris 3) (PhD Thesis).
- Bartens, Angela. 2000. *Ideophones and Sound Symbolism in Atlantic Creoles*. Helsinki: Annales Academiae Scientiarum Fennicae.
- Berlin, Brent & Paul Kay. 1969. *Basic color terms: their universality and evolution*. University of California Press. Berkeley and Los Angeles. (First Paperback Edition: 1991).
- Blench, Roger M. 2010. The Sensory World: Ideophones in Africa and Elsewhere. In Storch, Anne (ed.), *Perception of the Invisible. Religion, Historical Semantics and the Role of Perceptive Verbs*, pp. 271-292. Köln: Rüdiger Köppe.
- Bohnhoff, Lee Edward. 2014. *Dictionnaire de la langue dii (duru)*. <https://www.silcam.org/download.php?sstid=030401&folder=document&file=DiiDuruDictionary2014.pdf>. Accessed 18 September 2018.
- Bouquiaux, Luc, Jean-Marie Kobozo, Marcel Diki-Kidiri, Jacqueline Vallet & Anne Behaghel (eds.). 1978. *Dictionnaire sango-français et lexique français-sango*. Paris: SELAF.
- Braconnier, Cassian. 1999. *Dictionnaire du dioula d’Odienné* (2 vols). Paris: Université Paris 7 Denis Diderot.
- Caron, Bernard. 2002. Dott, aka ZoDi: Grammatical notes, vocabulary, text. *Afrika und Übersee* 85: 161-248.
- Caron, Bernard. 2005. *Za:r (Dictionary, grammar, texts)*. Ibadan (Nigeria): IFRA.
- Childs, Tucker G. 1994. Expressiveness in contact situations: The fate of African ideophones. *Journal of Pidgin and Creole Languages* 9(2): 257-282.
- Childs, Tucker G. 2002. Review of *Ideophones and Sound Symbolism in Atlantic Creoles*, by Angela Bartens. Helsinki: Annales Academiae Scientiarum Fennicae, 2000. *Journal of Pidgin and Creole Languages*: 284–289.
- Childs, Tucker G. 2012. *Kim data* (personal fieldwork). Available on line at www.reflex.cnrs.fr (accessed 18 September 2018).
- CLICS, see List et.al. (eds.). 2018.

- CLNK. 1999. *Tòm kpou kabiyè-fransu - Dictionnaire kabiyè-français (avec lexique français-kabiyè et esquisse de grammaire kabiyè)*. Kara, Lomé (Togo): Comité de Langue Nationale Kabiyè (CLNK) & SIL.
- Dedrick, Don. 1998. *Naming the rainbow*. Dordrecht: Kluwer Academic Publishers.
- Diané, Mamadi. 2012. L'expression des couleurs en maninka de Kanakan. *Mandenkan* 48: 21–38.
- Dimmendaal, Gerrit. J. 1995. Studying lexical-semantic fields in languages: Nature versus nurture, or where does culture come in these days. *Frankfurter Afrikanistische Blätter* 7: 1–28.
- Dimmendaal, Gerrit. J. 2015. *The Leopards Spots. Essays on Language, Cognition and Culture*. Leiden, Boston: Brill.
- Dingemans, Mark. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *STUF* 70(2): 363–384.
- Dumestre, Gérard. 2011. *Dictionnaire bambara-français - suivi d'un index abrégé français-bambara*. Paris: Karthala.
- Emeneau, Murray B. 1969. Onomatopoeics in the Indian linguistic area. *Language* 45(2): 274–299.
- Enfield, Nick J. 2008. Transmission biases in linguistic epidemiology. *Journal of Language Contact* 2: 299–310.
- Evans, Nicholas. 2012. Semantic typology. In Jae Jung Son (ed.), *The Oxford handbook of linguistic typology*, Oxford: Oxford University Press. Oxford Handbooks Online.
- Fivaz, Derek. 1963. *Some aspects of the ideophone in Zulu*. Hartford, Conn.: Hartford Seminary Foundation.
- Foley, William A. 1997. *Anthropological Linguistics: An Introduction*. Oxford: Blackwell.
- François, Alexandre. 2008. Semantic maps and the typology of colexification: intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From polysemy to semantic change*, pp. 163–215. Amsterdam, Philadelphia: Benjamins.
- Gast, Volker & Maria Koptjevskaja-Tamm. 2018. The areal factor in lexical typology. In Van Olmen, Daniël, Tanja Mortelmans and Frank Brisard (eds.), 2018. *Aspects of Linguistic Variation*, pp. 43–81. Berlin: De Gruyter Mouton.
- Güldemann, Tom. 2007. The Macro-Sudan belt: Towards identifying a linguistic area in northern sub-Saharan Africa. In B. Heine & D. Nurse (Eds.), *A Linguistic Geography of Africa*, pp. 151–185. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486272.006.
- Güldemann, Tom (ed.). 2018. *The languages and linguistics of Africa*. The World of Linguistics 11. Berlin: Mouton de Gruyter.
- Hall J.B., Thomlinson H.F., Oni P.I., Buchy M., & Aebischer D.P. 1997. A monograph of *Parkia biglobosa*. *School of Agricultural and Forest Sciences Publication* 9 (Bangor, University of Wales, 107 pp.)
- Hansford, Gillian. 1989. *Chumburung-English, English-Chumburung dictionary: a bilingual dictionary for the Chumburung language of Northern and Volta Regions, Ghana*. Tamale (Ghana): Ghana Institute of Linguistics, Literacy and Bible Translation (GILLBT).
- Hantgan, Abbie. 2012. *Bangime lexicon*. Available on line at www.reflex.cnrs.fr. Accessed 18 September 2018.
- Heath, Jeffrey. 2013. *Yanda Dom lexicon (Yanda)*. Available on line at www.reflex.cnrs.fr (accessed 18 September 2018).
- Hellenthal, Anneke Christine. 2010. *A grammar of Sheko*. Utrecht: LOT - Netherlands Graduate School of Linguistics.

- Innes, Gordon. 1967. *A Grebo-English dictionary*. Cambridge: University Press in association with the West African Languages Survey and the Institute of African Studies, University of Ibadan.
- Jullien de Pommerol, Patrice. 1999. *Grammaire pratique de l'arabe tchadien*. Paris: Karthala.
- Jones, Ross. 2004. *Bokobaru dictionary*. München: Lincom Europa.
- Kawata, Ashem Tem. 2004. *Bagó - Dictionnaire Lingala/Falansé - Français/Lingala*. Paris: Editions Karthala.
- Kaye, Alan. 1986. *Nigerian Arabic-English dictionary*. Malibu: Udena Publications.
- Koptjevskaja-Tamm, Maria & Henrik Liljegren. 2017. Semantic patterns from an areal perspective. In Hickey, Raymond (ed.), *The Cambridge handbook of areal linguistics Aspects of Linguistic Variation*, pp. 204-236. Cambridge: Cambridge University Press.
- List, Johann-Mattis, Simon Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi & Robert Forkel (eds.). 2018. *Database of Cross-Linguistic Colexifications*. Jena: Max Planck Institute for the Science of Human History. Available online at <http://clics.clld.org> (accessed 18 September 2018).
- List, Johann-Mattis, Simon J. Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi, & Robert Forkel. 2018. CLICS2: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology* 22(2): 277–306
- Malt, Barbara C. & Asifa Majid. 2013. How thought is mapped into words. *WIREs Cogn Sci* 4:583–597. doi: 10.1002/wcs.1251 (accessed 10 July 2019).
- Matras, Yaron. 2007. The borrowability of structural categories. In Yaron Matras & Jeannette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, pp. 31–73. Berlin, New York: Mouton de Gruyter.
- McCallum Jones, Ross. 2017. *Illo-Busa Dictionary*. München: Lincom Europa.
- Mollard-Desfour, Annie. 2008. Les mots de couleur: des passages entre langues et cultures. *Synergies Italie* 4: 23-32.
- Moñino, Yves. 2004. Une autre conception des lumières: Sur les noms des couleurs en gbara. In Elisabeth Motte-Florac & Gladys Guarisma (eds.), *Du terrain au cognitif: Linguistique, ethnolinguistique, ethnosciences. A Jacqueline M.C. Thomas*, pp. 241-265. Leuven, Paris: Peeters.
- Nakagawa, Hiroshi. 2011. A first report on G'ui ideophones. In Osamu Hieda, Christa König & Hiroshi Nakagawa (eds.), *Geographical typology and linguistic areas: With special reference to Africa*, pp. 279–286. Amsterdam, Philadelphia: John Benjamins.
- O'Meara, Carolyn, Susan Smythe Kung & Asifa Majid. 2019. The challenge of olfactory ideophones: Reconsidering ineffability from the Totonac-Tepehua perspective. *International Journal of American Linguistics* 85(2): 173-212.
- Owens, Jonathan. 1993. *A grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz.
- Palayer, Pierre. 1992. *Dictionnaire sar-français: Tchad*. Paris: Librairie Orientaliste Paul Geuthner.
- Prost, [Révérend] [Père] André. 1971. Eléments de sembla. *Afrique et langage* 8: 12-156.
- Rongier, Jacques. 1998. *Dictionnaire français-adele*. Abidjan: Institut de Linguistique Appliquée (ILA), Université Nationale de Côte d'Ivoire.
- Rzyski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast, Timotheus A. Bodt, Abbie Hantgan, Gereon A. Kaiping, Sophie Chang, Yunfan Lai, Natalia Morozova, Heini Arjava, Natalia Hübler, Ezequiel Koile, Steve Pepper, Mariann Proos, Briana Van Epps, Ingrid Blanco, Carolin Hundt, Sergei Monakhov, Kristina Pianykh, Sallona Ramesh, Russell D. Gray, Robert Forkel & Johann-Mattis List. 2020. The Database of Cross-Linguistic

- Colexifications, reproducible analysis of cross-linguistic polysemies. *Nature*. DOI: 10.1038/s41597-019-0341-x. Accessed 5 February 2020.
- Sachot (Santos), Rosine. 1996. *Le mey: langue ouest-atlantique de Guinée*. Paris: Université Paris 3.
- Schneider-Blum, Gertrud. 2013. *A Tima-English Dictionary: An Illustrated Lexicon of a Niger-Congo Language Spoken in the Nuba Mountains (Sudan)*. Köln: Rüdiger Köppe.
- Segerer, Guillaume & Sébastien Flavier. 2011-2019. *RefLex: Reference Lexicon of Africa*, Version 1.1. Paris, Lyon. www.reflex.cnrs.fr. Accessed 20 July 2019.
- Segerer, Guillaume & Martine Vanhove. 2019. Color naming in Africa. In Ida Raffaelli, Daniela Katunar & Barbara Kerovec (eds.), *Lexicalization Patterns in Colour Naming: A Cross-Linguistic Perspective*, pp. 287-330. Amsterdam, Philadelphia: Benjamins
- SIL Burkina. 2003. *Boo nen séwe san-fransi, fransi-san* [Lexique san–français, français–san]. SIL Burkina Faso.
- Souag, Lameen. 2013. Sub-Saharan lexical influence in North African Arabic and Berber. In Mena Lafkioui (ed.), *African Arabic: Approaches to Dialectology*, pp. 211-236. Berlin: Mouton de Gruyter.
- Wierzbicka, Anna. 2008. Why there are no ‘colour universals’ in language and thought. *Journal of the Royal Anthropological Institute* 14: 407-425.
- World Color Survey*. <http://www1.icsi.berkeley.edu/wcs/>. Accessed 08 July 2019.