Language and Identity in a Multilingual, Migrating World

J. Stephen Quakenbush and Gary F. Simons, Editors

Pike Center for INTEGRATIVE SCHOLARSHIP

Language and Identity in a Multilingual, Migrating World

J. Stephen Quakenbush and Gary F. Simons

This book is for sale at http://leanpub.com/languageandidentity

This version was published on 2019-11-06

Pike Center for Integrative Scholarship is an initiative of SIL International that builds capacity for language development through scholarship. Our Agile Publishing program uses the Lean Publishing platform to incubate in-progress ebooks using lightweight automated tools. These tools are placed in the hands of our scholars so that they can iterate and act on reader feedback to improve their work. Many of our titles are works in progress that are not yet ready for conventional publishing, but nonetheless have content that is worth sharing in its current state.

Contents

1. *Ethnologue* as a Sourcebook for Mapping Multilingualism: The Case of Sango

by Kenneth S. Olson and Gary F. Simons

Abstract. The focus of *Ethnologue* has been on first language (L1) use. This is reflected in the maps currently included in the resource, which show locations and boundaries corresponding to the distribution of L1 speakers. The location of second languages (L2s) is only occasionally represented by maps. Recent restructuring of the *Ethnologue* database provides a pathway for the production of L2 maps. As a test case, we produce a map of the geographic distribution of Sango [sag] in Central African Republic. Using available census data, we include on the map an estimate of the percentage of the population of each language community that uses Sango as an L2. We also discuss how the map sheds light on the identities associated with the use of Sango as an L2.

1.1 The changing role of *Ethnologue* in a multilingual world[1]

When the first edition of *Ethnologue* appeared in 1951, it consisted of ten mimeographed pages describing only forty-six languages. It opened with these words:

> Elaborate descriptions and classifications of the world's fauna and flora have been published in countless books and articles. Similar undertakings for the languages spoken in the world today have lagged far behind—only a few relatively rare and inaccessible lists and descriptions being available

[1]We wish to thank William Samarin, Christoph Müller, Mark Karan, and Elke Karan for helpful comments and suggestions.

> for many language areas. For the Christian who feels burdened with the urgency of giving the Word of God to every man in his own tongue, this want is most distressing. (Pittman 1951:1)

At the outset, the fundamental research question for *Ethnologue* was: What are all the distinct languages being used in the world today? The frontiers of knowledge involved identifying and describing thousands of previously unknown languages. The fifth edition in 1958 listed 2,360 living languages; the sixth edition (1965) listed 3,164; the seventh edition (1969) listed 4,493; the ninth edition (1978) listed 5,103; the eleventh edition (1988) listed 6,140; and the threshold of 7,000 living languages was crossed in 2005 with the fifteenth edition.[2] Once a language was identified, the most basic questions for describing it were things like: What is it called? Where is it located? How many people speak it? What other languages is it related to? What are its major dialects? Does it have the Bible or any other literature?

After seven decades the situation has changed dramatically. When *Ethnologue* began, the local languages of the world were largely unknown, and their users were thought to be basically monolingual. Today the languages are basically known, and their user communities are known to be largely multilingual. The trend of adding about a thousand languages every ten years has ceased with the number of living languages leveling off at around 7,100 during the 2010s. Today the fundamental research question for *Ethnologue* has become: What is the global language ecology? Language ecology is "the study of interactions between any given language and its environment" (Haugen 1972:325). That environment includes other languages—not just neighboring languages in a peer relationship, but also regional and national languages that may be in a dominant relationship. The description of a language within its environment must therefore include information about language use and language policy relative to the L1 and all the other languages its speakers may use. Haugen (1972:337) closes his article by listing ten basic questions (quoted verbatim below, including his use of italics) that are essential to answer in describing the ecology of a language:

1. What is its *classification* in relation to other languages?

[2]The fifteenth edition (2005) reports the number of living languages as 6,912. However, with the introduction of EGIDS in the seventeenth edition (2014), 126 languages that were previously described as extinct were reclassified as 9 (Dormant) and thus newly counted as living. Therefore, in terms of the current method for counting living languages, the 7,000 threshold was actually surpassed in the 2005 edition.

2. Who are its *users*?
3. What are its *domains* of use?
4. What *concurrent languages* are employed by its users?
5. What *internal varieties* does the language show?
6. What is the nature of its *written traditions*?
7. To what degree has its written form been *standardized*, i.e. unified and codified?
8. What kinds of *institutional support* has it won, either in government, education, or private organizations?
9. What are the *attitudes* of its users towards the language?
10. Finally, we may wish to sum up its status in a *typology* of *ecological* classification, which will tell us something about where the language stands and where it is going in comparison with the other languages of the world.

The *Ethnologue* database has fields for addressing all of these questions; it is the editorial objective of the overall effort to complete this level of basic description for every known language.

1.2 Mapping the range of L2 use

The *Ethnologue* description of a language seeks to document its use both as a primary language (L1) and as a secondary language (L2). For given individuals, L1 is the language they are most comfortable using in the domains of everyday life. It is typically the first language they learn, but that is not necessarily the case. And a person typically has one L1, though in the case of someone who has grown up using more than one language from infancy, there is the possibility of more than one L1 (see Harris, this volume, and Kim et al. 1997:174). In *Ethnologue* usage, an L2 is then any other language that a person might use.

Ethnologue has become known for its color map plates, which number 236 in the most recent edition (Eberhard, Simons, and Fennig 2019). These maps show the range of territory in which languages are used as an L1. The absence of maps showing the range of use as an L2 can lead to misinterpretation of the resource. We recently encountered a scholar from another academic discipline who was researching how language use in Africa interacted with his field of study. His impression from *Ethnologue* was that Africans could choose between only two languages: (1) the L1 or (2) the national language (i.e. French, English, Portuguese, etc.). He was unaware

of the existence of lingua francas and the important role that they play in the life of Africans across the continent. In dialoguing with him, it became clear that the maps were foundational in forming his understanding of the language situation in Africa, so that the absence of L2 maps in *Ethnologue* is what led to his inaccurate impression. Although information about L2 use is present in the text, introducing maps showing the range of L2 use could be a more impactful way of conveying such information—not just to the public at large, but especially to educators, development workers, government workers, church workers, and others who are serving in the midst of societal multilingualism.

Recent additions to the structure of the *Ethnologue* database provide a straightforward path for the generation of L2 maps. Specifically, the database has added a new intersection table to record the many-to-many relationship between language communities and L2s—that is, a single language community may use many L2s, and many language communities may use the same L2. In *Ethnologue* products,[3] this information appears in the "Language Use" section of a language entry. The use of another language as an L2 by the language community in focus is indicated by the phrase, "Also use *Name of L2* [*ISO code*]." If they use multiple L2s, a list of second languages is given. If such use is limited to a particular domain or age group or gender, a brief remark about this may be added.

This information can be used to create a range map for a particular L2 by coloring in the regions corresponding to all the language communities that use it as an L2.[4] The map[5] in figure 1, first published in Olson and Lewis (2018), is an example of this. Lingala[6] [lin] and Bangala [bxg] are used as L2s across the northern and western regions of the Democratic Republic of the Congo (DRC), as well as the northern part of the Republic of the Congo (RC). Although it arose as a lingua franca in the late 19th century in the mid-river region—about halfway between Kinshasa and Kisangani (Meeuwis 2010, 2013)—Lingala has more recently emerged as an L1 in and around Kinshasa.

[3] http://www.ethnologue.com/products.

[4] We are indebted to our colleague, Marcus Love, who used information we provided from the *Ethnologue* database to create the two maps in figures 1 and 2 using ArcGIS® software published by Esri®.

[5] We thank the following people for help in producing the Lingala/Bangala map: Douglas Boone, Annette Harrison, Wendy Atkins, Maryanne Augustin, William Gardner, Bettina Gottschlich-Modibale, Constance Kutsch Lojenga, Rob McKee, David Morgan, Salikoko Mufwene, Cami Robbins, Larry Robbins, Paul Thomas, Angela Williams-Ngumbu, and Doug Wright.

[6] A succinct overview of Lingala is Meeuwis 2013.

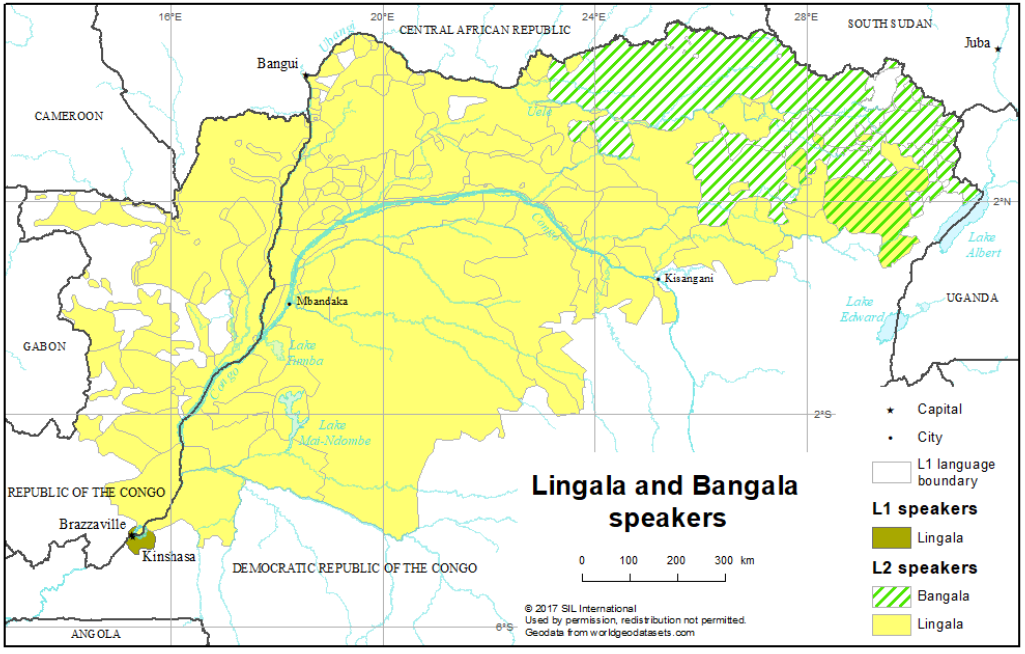

Figure 1. Geographic distribution of Lingala and Bangala (Olson and Lewis 2018). (White areas within the Lingala/Bangala region are sparsely populated or uninhabited.)

In figure 1, the boundaries of Lingala/Bangala L2 use are shown to coincide with the boundaries of the local language areas. This should be construed as an approximation. For example, along DRC's northern border with CAR, Lingala is used on *both* sides of the Ubangi and Mbomou Rivers, though its use tapers off quickly as one travels north.

In some cases it is necessary to distinguish the L1 and L2 boundaries. For example, for the Komo [kmw] community, Lingala is the primary L2 used west of the Lualaba River (the region directly south of Kisangani on the map), while Congo Swahili [swc] is used east of the river. As a result, we have included only the part of the Komo region west of the river in the Lingala L2 area in figure 1.

1.3 Mapping the degree of L2 use

Ethnologue further refines the "Also use" phrasing by the use of a set of quantifiers that provide rough estimates of the degree to which an L2 is used, as shown in table 1

below. As an example of the use of quantifiers, the "Language Use" field for Buduma [bdm] of Chad lists three L2s—two using a quantifier to estimate the degree of use, while the third provides no estimate of the degree to which the language is used:

> Most also use Kanembu [kbl], especially those living near Bol. Many also use Yerwa Kanuri [knc]. Also use Chadian Spoken Arabic [shu].

Quantifier	Criteria
All	At least 95% of the ethnic population use the reported language as L2
Most	At least 65% but less than 95% of the ethnic population use the reported language as L2
Many	At least 35% but less than 65% of the ethnic population use the reported language as L2
Some	At least 5% but less than 35% of the ethnic population use the reported language as L2
Few	Less than 5% of the ethnic population use the reported language as L2

Table 1. Quantifiers describing the extent of L2 use in a language community (Eberhard, Simons, and Fennig 2019)

Using this picklist of quantifiers to estimate the degree of L2 use is a recent innovation in the *Ethnologue* database introduced for the nineteenth edition (2016). Prior to this, the database required the amount of L2 use to be reported as the estimated number of bilingual speakers, but it had proven virtually impossible for the contributors to provide information of that precision. The editors thus introduced the picklist of quantifiers in hopes of making it feasible for contributors who have firsthand familiarity with a situation to provide an impressionistic estimate of the extent of bilingualism without needing to do research to establish an exact number.

When all of the "Also use" statements for a particular L2 within an area include a quantifier, it is possible to use the information in the *Ethnologue* database to create a more nuanced range map that uses shading to represent the degree of L2 use. As a test case, we developed such a map for the use of Sango[7] [sag] as an L2 in Central African Republic (CAR).[8] The result is shown in figure 2.

[7]A succinct overview of Sango is Samarin 2013.

[8]Due to lack of data, we did not attempt to map the distribution of Sango in adjacent countries.

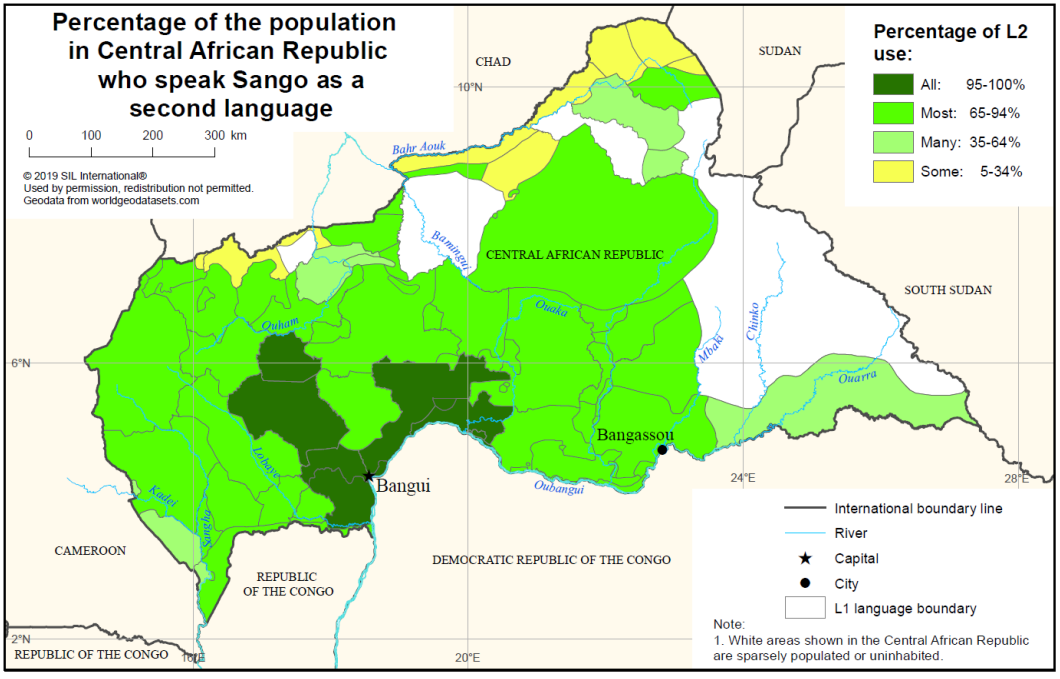

Figure 2. L2 use of Sango in Central African Republic.

We made use of data from the 1988 government census in CAR (Karan 2001:89–90) to select the appropriate quantifier for the use of Sango as an L2 for each language community in the country. One of the questions asked during the census was whether or not each person spoke Sango (Karan 2001:89). The census compiled the data for each region and reported a percentage of Sango use for each region.

For most language communities, the conversion of the data from political region to language region was relatively straightforward, with a single quantifier value being assigned to each language. However, it was necessary to assign different quantifiers to some geographic subregions of three languages: West Central Banda [bbp], Banda-Banda [bpd], and Manza [mzv]. Further modifications may be necessary to the *Ethnologue* database to accommodate such cases.

Before discussing the results, a few comments about the methodology are in order. On the positive side, the census question that was asked directly corresponds to the information encoded in the quantifiers—the percentage of a population that uses Sango. Also, every household was interviewed, so the data are quite complete.

At the same time, the results need to be qualified. The census interview is an

indirect measure of language use, rather than a direct measure of language use itself. Mark Karan (2001:89) suggests that several factors may have led to the over-reporting of Sango use: the question was asked by a government employee, the head of the household provided the information on behalf of the entire family, and use of Sango is considered prestigious in CAR (cf. Samarin 1955:262–263). Sango's status as a symbol of national identity (see below) was likely also a significant factor (Mark Karan, personal communication). That being said, Samarin (2007:352–353) reports that in a 1962 survey of Sango use in the rural Gbeya [gbp] area, he found that the average percentage of Gbeya-Sango bilinguals was 66%, which was close to the number reported in the 1988 census for that region.

In much of the country, either all or most people report speaking Sango. The most concentrated areas are in and around the capital Bangui, north and east of Bangui along the Ubangi River, and inland to the northwest and northeast of Bangui. The places where use of Sango is more limited are in the eastern, northern, and southwestern edges of the country. The regions in white on the map are generally uninhabited.

1.4 Identity and the spread of Sango

Sango is the main lingua franca in CAR. The generally accepted view is that it developed on the Ubangi River in the late 19th century as a pidginized[9] form of one of the speech varieties from the Ngbandi group[10] (Lekens 1951, cited by Samarin 1955:256; Samarin 1986:382; Boyeldieu and Diki-Kidiri 1988:31; Karan 2001:6). The source variety is often identified as "Sango Riverain," an appellation that corresponds to one of the extant varieties [snj]. However, the term is sometimes used in a broader sense to refer to all the riparian varieties of the Ngbandi group. Hence some uncertainty remains about the exact source variety.

[9]Diki-Kidiri (1986) and Morrill (1997) hold a different view, arguing instead that Sango did not undergo pidginization as it emerged from its source language.

[10]Eberhard, Simons, and Fennig (2019) consider the Ngbandi group to consist of several distinct languages. However, Boyeldieu (1982) finds minimal linguistic variation between the varieties in CAR: Sango Riverain, Yakoma [yky], and Dendi [deq]. Buchanan (2007:11) reports, "Les Dendis comprennent bien le yakoma, le ngbandi et le sango, c'est-à-dire qu'il existe une bonne intercompréhension entre ces parlers" [The Dendis understand Yakoma, Ngbandi, and Sango well, that is to say there is a good intercomprehension between these dialects], but earlier (p. 8) he notes that Dendi speakers have some difficulty understanding Ngbandi [ngb] from DRC.

Sango has spread extensively, particularly to the north of the river, so that it is now used throughout most of the country. It is also spoken to some degree across the borders in northern DRC, northern RC, eastern Cameroon, and southern Chad (Samarin 2007:347, Meeuwis 2013). It has also emerged as an L1 (Karan 2001:5), with roughly 500,000 L1 speakers (Eberhard, Simons, and Fennig 2019). Jacquot (1961:163) describes the creolization of Sango in Bangui among children from various language groups using Sango as a common language in school (cf. Samarin 2001). In 1964 it was named the national language by constitutional law no. 64/37—the same law that made French the official language. Then in 1991 it became an official language of CAR alongside French by constitutional law no. 91/001 (Koyt 1994:503).

Karan (2001:92–93) offers two key factors in explaining the areas of high Sango use. First, these regions correspond to major transportation arteries—the Ubangi River (navigable up to Bangassou), the Lobaye River, and the Ouaka River. Considering Sango's history as a lingua franca, this is not unexpected.

Second, the region also generally corresponds to the location of the two major Protestant church communities before independence—the churches connected to the Baptist Mid-Mission (BMM) and the Grace Brethren International Mission (GBIM, now called Encompass World Partners). Both missions began work in CAR in the early 1920s, and both made administrative decisions early on to use Sango exclusively. The BMM adopted Sango in 1923, and the GBIM somewhat later following World War II, after initially focusing on the local languages (Samarin 2007:351). This was mainly for the pragmatic reason of facilitating mission work across each respective region. The entire Sango Bible was published in 1966—a collaboration between the two communities. The Roman Catholic Church initially used the local languages in its work but later switched to Sango, likely for the same reason (Karan 2006:239).

Identity issues come into play in explaining the spread of Sango. First, it is not coincidental that the spread of the language reached—and was generally attenuated by—the international borders. Sango was not used by the French colonial government (Le Page 1997:60), but Central African recruits in the French Army promoted Sango use for interethnic communication (Karan 2006:239, cf. Samarin 1955:256–257). Subsequently, the language became associated with the independence movement (Le Page 1997:60). When William Samarin conducted a sociolinguistic survey in a Gbeya [gbp] village, the residents considered Sango to be "the language of the Central African Republic" (Samarin 1986:379). With factors like these in the background, the language gradually became a symbol of national identity in CAR (Karan 2001:12, 18;

Koyt 1994).[11]

Identity issues also came into play in the adoption of a new official Sango orthography in 1984. This orthography was seen as reinforcing an independent national identity for two reasons: (1) It was created by Central African linguists, and (2) it intentionally distanced itself from the previous orthography, which was based extensively on the French orthography (Karan 2006:269–270, 307).

Second, while the use of Sango in the major church communities started for pragmatic reasons, it has now taken on a function of building church unity (Karan 2001:108). Samarin (1955:256) states, "Some Africans feel that the native languages tend to divide the people whereas Sango unites them all into one Christian body. Moreover, there is some evidence to indicate that Sango has come to be identified, at least among the Protestants, as the Christian language." The churches encourage the spread of Sango because they see it as a way to more efficiently carry out their ecclesiastical functions and also unite the various ethnic groups.

1.5 Conclusion

In summary, the inclusion of L2 maps in *Ethnologue* would help it to reflect better the realities of our multilingual world. This in turn would benefit users of this resource by giving them a better understanding of the linguistic realities on the ground. Such maps would not simply portray the geographic range of use for languages of wider communication, but also of the identities their users could share.

1.6 References

Boyeldieu, Pascal. 1982. Structures sociales et particularismes linguistiques en pays de langue 'Ngbandi': Éléments pour une étude [Social structures and distinctive linguistic identities in the Ngbandi language region: Data for a study]. In Pascal Boyeldieu and Marcel Diki-Kidiri (eds.), *Le domaine ngbandi*, 13–80. Bibliothèque de la SELAF 93. Paris: SELAF.

Boyeldieu, Pascal, and Marcel Diki-Kidiri. 1988. Le group ngbandi-sango-kpatiri [The Ngbandi-Sango-Kpatiri group]. In Yves Moñino (ed.), *Lexique comparatif des*

[11]In additional to a national identity, Samarin (1955:259–260) suggests that additional factors may have contributed to the attenuation of Sango at the borders. These include competition with other lingua francas (e.g. Lingala in DRC, Chadian Spoken Arabic in Chad) and economic isolation from its neighbors.

langues oubanguiennes, 31–33. Paris: Geuthner.

Buchanan, Michael. 2007. Rapport d'enquête sociolinguistique: Première évaluation parmi les Dendis [Sociolinguistic survey report: Initial assessment among the Dendis]. *SIL Electronic Survey Reports* 2007-001. https://www.sil.org/resources/publications/entry/9152.

Diki-Kidiri, Marcel. 1986. Le sango dans la formation de la nation centrafricaine [Sango in the creation of the Central African Republic]. *Politique Africaine* 23:83–99. Paris: Karthala. http://www.politique-africaine.com/numeros/023_SOM.HTM.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2019. *Ethnologue: Languages of the world.* Twenty-second edition. Dallas, TX: SIL International. http://www.ethnologue.com

Harris, Kyle. 2018. L1 and L2 comprehension and emotional impact among early proficient bilinguals. In J. Stephen Quakenbush and Gary F. Simons (eds), *Language and identity in a multilingual, migrating world.* Dallas, TX: SIL International.

Haugen, Einar. 1972. The ecology of language. In Einar Haugen, *The ecology of language: Essays by Einar Haugen*, 325–339. Language Science and National Development. Stanford: Stanford University Press.

Jacquot, André. 1961. Notes sur la situation du Sango à Bangui: Résultats d'un sondage [Notes on the status of Sango in Bangui: Results of a survey]. *Africa: Journal of the International African Institute* 31(2):158–166. DOI: 10.2307/1158103.

Karan, Elke. 2006. Writing system development and reform: A process. MA thesis. University of North Dakota, Grand Forks. http://arts-sciences.und.edu/summer-institute-of-linguistics/theses/2006-karan-elke.cfm.

Karan, Mark E. 2001. *The dynamics of Sango language spread.* SIL International Publications in Sociolinguistics 7. Dallas, TX.

Kim, Karl H. S., Norman R. Relkin, Kyoung-Min Lee, and Joy Hirsch. 1997. Distinct cortical areas associated with native and second languages. *Nature* 388(6638):171–174. DOI: 10.1038/40623.

Koyt, Michel M. 1994. Situation et politique linguistiques en République Centrafricaine [Language status and policy in the Central African Republic]. In Pierre Martel and Jacques Maurais (eds.), *Langues et sociétés en contact: Mélanges offerts à Jean-Claude Corbeil*, 503–516. Canadiana Romanica 8. Berlin: De Gruyter.

Le Page, Robert B. 1997. Political and economic aspects of vernacular literacy. In Andrée Tabouret-Keller, Robert B. Le Page, Penelope Gardner-Chloros, and Gabrielle Varro (eds.), *Vernacular literacy: A re-evaluation*, 23–81. Oxford Studies

in Anthropological Linguistics 13. Oxford: Clarendon Press.

Lekens, Benjamin. 1951. Nota over het ngbandi als voertaal in Ubangi [Note on Ngbandi as the language of wider communication in the Ubangi region]. *Kongo-Overzee* 17:162–164.

Meeuwis, Michael. 2010. *A grammatical overview of Lingála*. LINCOM Studies in African Linguistics 81. Munich: Lincom Europa.

Meeuwis, Michael. 2013. Lingala. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber (eds.), *The survey of pidgin and creole languages. Volume 3: Contact languages based on languages from Africa, Asia, Australia, and the Americas*. Oxford: Oxford University Press. http://apics-online.info/surveys/60.

Morrill, Charles Henry. 1997. Language, culture, and society in the Central African Republic: The emergence and development of Sango. PhD dissertation. Indiana University.

Olson, Kenneth S., and M. Paul Lewis. 2018. The *Ethnologue* and L2 mapping. In Ericka A. Albaugh and Kathryn M. de Luna (eds.), *Tracing language movement in Africa*, 45–65. New York: Oxford University Press. DOI: 10.1093/oso/9780190657543.003.0003.

Pittman, Richard S. 1951. *Missionary Ethnologue for intercessors, translators, missionaries, and mission councils* [Ethnologue. First edition]. Berwick, Australia: Wycliffe School of Linguistics. 10 pp., mimeo. First page reproduced at http://www.sil.org/history-event/first-edition-ethnologue.

Samarin, William J. 1955. Sango, an African lingua franca. *Word* 11:254–267. DOI: 10.1080/00437956.1955.11659562.

Samarin, William J. 1986. French and Sango in the Central African Republic. *Anthropological Linguistics* 28:379–387. https://www.jstor.org/stable/30027964.

Samarin, William J. 2001. Explaining shift to Sango in Bangui. In Robert Nicolaï (ed.), *Leçons d'Afrique: Filiations, rupture et reconstitutions de langues: Un hommage à Gabriel Manessy*, 351–391. Louvain-Paris: Peeters.

Samarin, William J. 2007. Review of *The dynamics of Sango language spread*, by Mark Karan. *Journal of Pidgin and Creole Languages* 22:347–366. DOI: 10.1075/jpcl.22.2.08sam.

Samarin, William J. 2013. Sango. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber (eds.), *The survey of pidgin and creole languages. Volume 3: Contact languages based on languages from Africa, Asia, Australia, and the Americas*, 13–24. Oxford: Oxford University Press.

http://apics-online.info/surveys/59.