Prioritizing African Languages: Challenges to macro-level planning for resourcing and capacity building

Tristan M. Purvis
Christopher R. Green
Gregory K. Iverson
University of Maryland Center for Advanced Study of Language

Abstract
This paper addresses key considerations and challenges involved in the process of prioritizing languages in an area of high linguistic diversity like Africa alongside other world regions. The paper identifies general considerations that must be taken into account in this process and reviews the placement of African languages on priority lists over the years and across different agencies and organizations. An outline of factors is presented that is used when organizing resources and planning research on African languages that categorizes major or critical languages within a framework that allows for broad coverage of the full linguistic diversity of the continent.

Keywords: language prioritization, African languages, capacity building, language diversity, language documentation

When building language capacity on an individual or localized level, the question of which languages matter most is relatively less complicated than it is for those planning and providing for language capabilities at the macro level. An American anthropology student working with Sierra Leonean refugees in Forecariah, Guinea, for example, will likely know how to address and balance needs for language skills in French, Susu, Krio, and a set of other languages such as Temne and Mandinka. An education official or activist in Mwanza, Tanzania, will be concerned primarily with English, Swahili, and Sukuma. An administrator of a grant program for Less Commonly Taught Languages, or LCTLs, or a newly appointed language authority for the United States Department of Education, Department of Commerce, or U.S. Africa Command (AFRICOM), on the other hand, may view the *Ethnologue* list of over 2,000 indigenous languages

spoken in Africa, wondering, in exasperation, "Where to begin?"[1] In order to try to answer this daunting question, we explore in this paper several key considerations and challenges involved in prioritizing language resources and training needs for highly linguistically diverse areas, specifically, the several regions of Africa. We also explore potential approaches that aim to enhance the utilization of the extensive multilingualism found in such areas as it affects education, literacy, diplomacy, and defense.

We begin by highlighting several aspects of the linguistic and geographic diversity of Africa and compare it to other traditionally defined world regions of focus and administration. Next, we review general considerations and issues involved in language priority setting, among them the factors that must be taken into account when assigning priority levels and the debate about which characteristics separate a language from a dialect. We then present the extent to which African languages have appeared in a select set of language priority lists made available from different United States and international administrative bodies. We close with a brief discussion of important points of departure for those embarking on an effort of language prioritization for their organization that will facilitate adaptation to varying levels of interest and engagement in these languages, including accommodations for broad coverage of linguistic diversity when organizing resources and planning research and training on African languages.

Linguistic diversity of Africa compared to other regions of the world

Presentations and scholarship focused on research and other work being carried out in Africa often highlight the challenges that have arisen in the face of the countless unique aspects of the vast geographic and demographic scope and complex linguistic diversity of the continent. While the fact that Africa is a large continent is perhaps self-apparent, the exact scope of its expanse is often overlooked

[1] *Ethnologue* is a widely used online and print reference encyclopedia of the world's languages (Lewis, 2009). *Ethnologue* aims to be the most up-to-date resource on the nearly 7,000 languages of the world spoken across all continents. In addition to the wide variety of language materials it contains, drawn largely from the worldwide network of SIL International, *Ethnologue* contains demographic, genetic, dialect, and bibliographic information for the world's languages searchable by their unique three character ISO (International Organization for Standardization) code.

by the fact that, for one reason or another, the entire continent is often anecdotally equated to a monolithic "country." Furthermore, traditional mapping conventions, such as the cylindrical Mercator projection, present a distorted view of the continent. In reality, Africa's land mass is roughly 11.7 million square miles, thus well over 3.5 times the size of the continental United States—or, as presented by Krause (see "Cartography," 2010), Africa covers an area roughly the same size as the United States, China, India, Japan, and all of Europe combined. The continent comprises 54 sovereign nations, as of July 9, 2011, with the inclusion of newly independent South Sudan, and its total population is approaching one billion and growing rapidly.

In terms of linguistic diversity, there are over 800 ethnic groups and approximately 2,100 unique languages spoken in Africa, taking into consideration that these most recent estimates often overlook important distinctions between dialects and other varieties of certain languages that may be only mutually intelligible to a marginal degree. Five major language families are represented by the indigenous languages of Africa, viz., Afroasiatic, Nilo-Saharan, Niger-Congo, Khoisan, and Austronesian, the last among these represented by Malagasy, spoken on the island nation of Madagascar. There are also numerous language isolates, i.e., languages yet unclassified or thus far deemed to be unrelated to any of these major language families. Furthermore, there is substantial representation by Indo-European languages owing to the lasting legacy of colonial languages, chiefly English, French, Portuguese, as well as closely-related emerging language varieties like pidgins, creoles, and mixed languages that have been born out of long-term contact between Indo-European and indigenous African languages or between different indigenous languages (Mufwene, 2008; Thomason, 1997). A striking example of this linguistic diversity is exemplified in Figure 1, which displays the 2,099 language ISO codes documented in *Ethnologue* (Lewis, 2009), color-coded by language family. Even more striking, as seen in the zoomed-in portion of the map, is that in a particular African country, Nigeria (albeit the most extreme example), we encounter an equally daunting subset of languages. Nigeria alone is home to over 500 indigenous languages from 3 major language families.

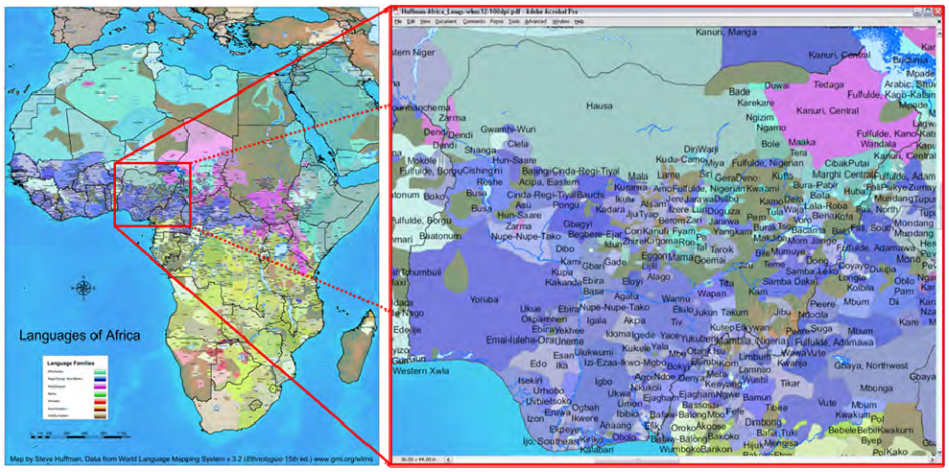

Figure 1. Linguistic diversity of Africa (source: GMI, n.d.).

Although in the current paper we have chosen to highlight Africa, we point out that many of these same issues and indeed the solutions proposed here readily apply to other major world regions. For the ease of comparison, we will employ the United States State Department regional divisions in our discussion. In light of the U.S. State Department's division of regional bureaus, in the case of Africa, we are looking specifically at sub-Saharan Africa. That is to say, North Africa is included as part of a separate region, namely the Near East. This comparison of U.S. State Department regional divisions serves as a preface to many of the challenges involved in language priority determination (as discussed in later sections), such as how to divide the world into regions, what counts as a language (as opposed to dialects of a single language), how to estimate speaker populations most reliably, and how to account for L2 (second language) vs. L1 (mother-tongue) speakers.

As a means for comparison between the many languages spoken in different world regions, we limit ourselves and our discussion below to those languages with over three million speakers. Also, for the sake of consistency in this comparison, we base this comparison only upon L1 speaker population (which therefore overlooks major African trade languages and lingua francas with low L1 populations, e.g. Swahili and Lingala). We draw upon population figures

found in *Ethnologue* (Lewis, 2009) which, by their nature, come from disparate sources.

To begin our discussion, let us consider Table 1 containing major languages with more than three million speakers as they are divided between six world regions, namely Sub-Saharan Africa, Western Hemisphere, Near East, Europe and Eurasia, East Asia and Pacific, and South and Central Asia. For each region, the number of countries, size in millions of square miles, total number of languages spoken, and number of languages with more than three million speakers are given. These languages are listed by region and subregion where appropriate. Bracketed languages within a single region are considered to be closely related.

Table 1. Linguistic, demographic, and geographic comparison of U.S. State Department regions. (Bracketed languages in a cell indicate closely related languages based upon *Ethnologue*'s classification.)

Sub-Saharan Africa Countries: 48 Size: 9.5 mil. sq. m Total languages: 2033 Total over 3 million: 42	Afrikaans, Akan, Amharic, [Arabic-Sudanese, Hassaniyya], Bemba, Dhuluo, Ewe, [Fula (Pulaar)], Ganda, Gikuyu, Hausa, Igbo, Kamba, Kanuri-Central, Kimbundu, Kituba, Koongo, Luba-Kasai, Makhuwa, Malagasy-Plateau, Mandingo, Moore, Nyanja, [Oromo-Boran-Arsi-Guji, Oromo-Eastern, Oromo-West Central], [Rundi, Rwanda], Shona, Somali, [Sotho-Northern, Sotho-Southern, Tswana], Sukuma, Tigrigna, Tsonga, Umbundu, Wolof, [Xhosa, Zulu], Yoruba **Major trade languages**: Swahili, Lingala
Western Hemisphere Countries: 37 Area: 16.3 mil. sq. mi. Total languages: 794 Total over 3 million: 5	Haitian, Jamaican, Guarani-Paraguayan, Hunsrik, Quechua-South Bolivian **Major European languages**: English, French, Portuguese, Spanish
Near East Countries: 19 Area: 2.1 mil. sq. mi. Total languages: 139 Total over 3 mil-	**Middle East** (17): [Arabic-Gulf, Arabic-Hijazi, Arabic-Mesopotamian, Arabic-Najdi, Arabic-North Levantine, Arabic-North Mesopotanian, Arabic-Sanaari, Arabic-South Levantine, Arabic-Taizzi-Adeni], Azerbaijani-South, Domari, Farsi-Western, [Gilaki, Mazanderani], Hebrew, [Kurdish-Central, Kurdish-Southern] **North Africa** (9): [Arabic-Algerian, Arabic-Egyptian, Arabic-

lion: 26	Libyan, Arabic-Moroccan, Arabic-Saidi, Arabic-Tunisian], [Kabyle, Tamazight-Central Atlas, Tachelhit]
Europe and Eur Countries: 50 Area: 10.9 mil. sq. mi. Total languages: 280 Total over 3 million: 43	**West** (28): Catalan-Valencian-Balear, Danish, [Dutch, Vlaams], English, Finnish, French, [German-Standard, German-Swiss, Bavarian, Mainfränkish], Greek, Hungarian, [Italian, Napoleta-no-Calabrese, Sicilian], [Lombard, Piemontese, Venetian], Norwegian, Polish, [Portuguese, Galician], Romanian, Spanish, Swedish, Turkish, Kurdish-Northern **East** (15): [Albanian-Gheg, Albanian-Tosk], Armenian, Azer-baijani-North, Bulgarian, [Croatian, Serbian], Czech, Slovak, Georgian, Lithuanian, [Russian, Belarusan, Ukrainian], Tatar
East Asia and P Countries: 31 Size: 9.6 mil. sq. n Total languages: 2633 Total over 3 million: 43	**East Asia** (27): Burmese, [Chinese-Gan, Chinese-Hakka, Chi-nese-Huizhou, Chinese-Jinyu, Chinese-Mandarin, Chinese-Min Bei, Chinese-Min Dong, Chinese-Min Nan, Chinese-Min Zhong, Chinese-Wu, Chinese-Xiang, Chinese-Yue], Hmong, Japanese, Khmer-Central, Korean, Monogolian (Peripheral), Shan, [Thai, Thai-Northeastern, Thai-Northern, Thai-Southern, Lao], Uyghur, Vietnamese, Zhuang **Pacific** (17): Aceh, Bali, Banjar, Bikol, Bugis, Cebuano, Filipi-no, Hiligaynan, Ilocano, Indonesian, Javanese, Madura, Malay, Minangkabau, Musi, Sunda, Tagalog
South and Centr Asia Countries: 13 Size: 2 mil. sq. mi. Total languages: 603 Total over 3 million: 53	**South Asia** (49): [Assamese, Bengali, Chittagonian], [Awadhi, Bagheli, Chhattisgarhi], Balochi-Southern, [Bhojpuri, Magahi, Maithili], Deccan, Dogri, [Farsi-Eastern (Dari)], Gujarati, [Ha-dothi, Malvi], Haryanvi, Hindi, Urdu, Kanauji, Kannada, Kashmiri, [Konkani, Konkani-Goan], Lambadi, Malayalam, Marathi, [Marwari, Merwari, Dhundari, Shekhawati, Godwari], Mewati, Mina, Nepali, Oriya, [Panjabi-Eastern, Panjabi-Western, Seraiki], [Pashto-Central, Pashto-Northern], [Rangpu-ri, Sylheti,] Santali, Sindhi, Sinhala, Tamil, Telugu, Varhadi-Nagpuri **Central Asia** (4): Kazakh, Tajiki, Turkmen, Uzbek-Northern

The lists in Table 1 immediate highlight the fact that two world regions, in particular, boast the most linguistic diversity, namely Sub-Saharan Africa and the East Asia/Pacific regions, having 48 countries/2,033 languages and 31 countries, 2,633 languages, respectively. South and Central Asia, on the other hand, has far fewer countries, but contains the highest number of languages with at least three million speakers.

From the perspective of addressing educational, administrative, development, management, national security, and other needs, professionals working in the United States may already have in mind ready comparisons between regions like the Americas and the Near East, or perhaps Europe. Most foreign language interactions conducted in the America's, particularly at the level of administration and education, can be handled in English, Spanish, Portuguese, French, and perhaps a small number of related creole languages like Haitian and Jamaican. While there are some 800 or so different languages spoken in the Americas, most of them do not boast the vibrant numbers of speakers found for other LCTLs spoken elsewhere in the world. Unfortunately, most are facing endangerment, i.e. dying off as their speakers (or descendants) and associated ethnic groups adopt other major languages as their primary means of communication.

It should be clear from the presentation of Near East languages in Table 1 that language planning in this region, although surely requiring more in the way of high level language training and support for language resource development, is somewhat less daunting. Building capacity in a handful of languages, primary among them being Arabic, Farsi, and Hebrew (as well as English, French, and perhaps major Berber languages), has the potential to yield significant rewards in the approximately 19 countries found in that region. While expertise in numerous varieties of Arabic is clearly needed, depending on where one is working, it is nonetheless relatively easy to apply language skills gained for one variety of Arabic to learn other related varieties. The equivalent task of attaining high levels of proficiency in geographically proximal but genetically distant languages, such as is the case in many parts of sub-Saharan Africa, does not come with this benefit of linguistic transfer to the same extent as that found for closely related varieties of Arabic.

The Europe and Eurasia region poses more of a challenge than one might think, given that this region covers roughly the same geographic area, has nearly the same number of countries, and approximately the same number of languages with at least 3 million L1 speakers as Africa.[2] Of course, a notable difference here is that both

[2] Note that what is actually presented here is better considered to be Europe and New Independent States, given that Central Eurasian countries are included in the South and Central Asia region.

material and human resources for European languages are far better established and better studied compared to African languages. Furthermore, as with the Near East region, the number of languages with over 3 million L1 speakers diminishes when one takes into consideration related groups of languages, for example dialects of German or Italian, and the interesting case of Serbian and Croatian, which, linguistically speaking, constitute one mutually intelligible language with different 'varieties' defined by national boundaries.

More closely paralleling the challenges for language capacity building in Africa are the regions of South and Central Asia and East Asia and Pacific. Although covering a much smaller geographic area and comprised of few countries, the densely populated South and Central Asia region has numerous language varieties boasting speech communities of well over three million L1 speakers. The East Asia and Pacific region, likewise, has a comparably large number of languages with over three million speakers located in a region that covers roughly the same geographic area and number of countries as Africa. It is possible, however, to compare the case of Chinese to that of the Arabic dialects discussed briefly above in the Near East region. Building capacity in a major variety of Chinese like Mandarin is likely to serve as a useful springboard into learning other related Chinese languages. Furthermore, the acquisition of reading and writing skills using Chinese characters (Hanzi) can often be applied to multiple literate language varieties in the East Asia region, for example Hanja characters in Korean and Kanji characters in Japanese.

Having thus provided this brief introduction to the diversity, breadth, and expanse of the world's linguistic architecture, we next turn more specifically to the task of language prioritization in Africa. As discussed below, a multitude of factors must enter into one's consideration in taking this issue to task, not least among them being the longstanding question, "what is a language?"

General considerations and dilemmas for language prioritiztion

How are priorities set?

In the preceding section, we offered a brief comparison of Africa to other world regions in order to frame its linguistic landscape against similar problems and potential solutions to be found

for languages across the globe. While we have alluded to the types of considerations and challenges that go into setting priorities for language capacity building, in this section, we shall focus more specifically on these challenges as they pertain to African language prioritization and steps taken thus far to address these issues.

In 1979, Michigan State University hosted a conference titled *African Language Instruction in the U.S.: Directions and Priorities for the 1980's*. Three broad factors were ultimately agreed upon and defined that aimed to assist a diverse audience in the selection of the highest priority languages from among the hundreds of languages of Africa (Dwyer, 1986; Wiley & Dwyer, 1980; Wiley, 2004). In a later update to African language prioritization stemming from this conference, Wiley (2004) added a fourth criterion. The four criteria are as follows: (i) Number of speakers, (ii) Political, cultural, and social importance (i.e. its status as an official, national language, or lingua franca, importance for education, literacy, etc., (iii) Importance for U.S. national interests, and (iv) Importance for scholars working in archives (Wiley, 2004). We explore the first three of these criteria in detail below.

In many cases, a high ranking for one of these criteria predictably goes hand in hand with a high ranking for one or more of the others; however, one can identify instances in which languages like Wolof and Tigrinya, having 4 million and 5.8 million speakers, respectively, typically rank higher on priority lists than a language like Kikuyu, having 7 million speakers, due to their status as a national language or other socio-political criteria. Similarly, Somali (13.8 million speakers) generally ranks somewhat higher than Igbo (18 million speakers) when taking into consideration geopolitics, security interests, and the status of Somali as one of three official languages of Somalia (along with English and Arabic). It should be clear that a complex methodology of cross comparison and the weighting of many factors must be brought to bear on the outcome of prioritization as a result of these competing concerns. There are also additional considerations and dilemmas behind each of these component criteria themselves, as addressed in the subsections that follow.

Number of speakers. One of the clearest challenges surrounding the justification to use speaker population as a factor in prioritization concerns the longstanding issue of ensuring that accurate,

reliable, and up-to-date estimates of speaker populations are available. *Ethnologue* (Lewis, 2009) serves as a convenient resource in terms of its comprehensiveness and accessibility, as it often contains demographic information for both L1 and L2 speaker populations. The issue that arises, however, is that these figures come from disparate sources ranging from the 1980s to present. Take, for example, the different types of data contained in the relatively current 2006 population estimates for Wolof, as shown in Table 2, compared to those given for Lingala (2000 and 1999), and for Yoruba (1993). Furthermore, the Wolof data, as well as that for the L1 speakers reported for Congo, ostensibly comes from SIL survey work as is typically implied by the unmarked source dates; whereas the figure for Nigeria comes from a separate missionary publication (Johnstone, 1993) and the L2 estimates for Lingala and Bangala is from the 1999 *World Almanac and Book of Facts* (Wiesenfeld, 1999).

Table 2. Examples of population estimates in *Ethnologue* (Lewis, 2009)

Lingala: Population: 2,040,000 (2000) in Democratic Republic of the Congo
L2 speakers together with Bangala: 7,000,000 (1999 WA) in Democratic Republic of the Congo
Population total all countries: 2,141,300

Wolof: Population: 3,930,000 (2006) in Senegal
Population total all countries: 3,975,500

Yoruba: Population: 18,900,000 (Johnstone 1993) in Nigeria
L2 speakers: 2,000,000
Population total all countries: 19,380,800

An alternative resource like the *World Factbook* is perhaps more consistent, as overall population figures, based on U.S. Census Bureau estimations and projections, are updated regularly. An additional complicating factor, however, is the fact that, for the majority of African countries with diverse ethnic populations, only major ethnic groups and languages are reported, and even then, only estimates for L1 speakers are provided at best. Furthermore, these percentages do not appear to be updated regularly, and the sources for these figures are not clearly indicated. Consider, for example, the case of Senegal in Table 3, where only the top 6 of 37 ethnic groups are documented in the *World Factbook*. Specific language populations are not documented. In general, one must question if it is possible and, moreover, if it is advisable to base estimates for speaker population upon estimates of ethnic population? Doing so would present us with a puzzling situation where the *World Factbook* reports 5.3 million eth-

nic Wolof in Senegal in 2010 (or 43.3 percent of the total population), while *Ethnologue* reports approximately 4 million speakers of the Wolof language in 2006.[3]

Table 3. Examples of population estimates in the *World Factbook*
Senegal: Last updated July 05, 2011

Population	12,643,799 (July 2011 est.)
Ethnic groups	Wolof 43.3%, Pular 23.8%, Serer 14.7%, Jola 3.7%, Mandinka 3%, Soninke 1.1%, European and Lebanese 1%, other 9.4%
Languages	French (official), Wolof, Pulaar, Jola, Mandinka

This issue is a complicated one. Can one assume that all ethnic Wolof report speaking the Wolof language? How many of the 4 million speakers of Wolof reported by *Ethnologue* are L2 speakers? Moreover, are L2 speakers accounted for in some other way that is not reported? It is equally plausible, given the sprawl of French in Senegalese urban areas, that an ethnic Wolof speaks French as his or her L1 and perhaps Wolof only as an L2. Indeed, this situation arises in many African urban centers where, for example, an ethnic Fulani in Mali who was born and raised in Bamako may be a native speaker of Bambara, an L2 speaker of French, but have only elementary capabilities in the Fulani language. These reports, thus, require a high degree of scrutiny and investigation when used for prioritization.

Political, cultural, and social importance. Although information on political, cultural, and social criteria are not as difficult to collect as demographic information on speaker population, these data are nonetheless prone to change and can be questionable in terms of their practical relevance for measuring language prioritization. One must take into consideration shifts in official or national

[3] For a more chronologically comparable figure, the *2006 World Factbook* estimate was 5.2 million, or 43.3% of 11,987,121 (11 July 2006 update). Interestingly, the *World Factbook* population estimate for Senegal had risen as high as 14,086,103 for their initial July 2010 estimate (according to www archives). The dramatic drop back to 12,323,252 for a revised "July 2010" estimate in the first half of 2011 was possibly influenced by the U.N. World Population Division's 2010 revision (12,434,000), barely up from the 2006 revision figure of 12,379,000. The U.N. 2010 revision more than likely reflects an opinion that the previous figures had been overestimated, as Senegal did not experience an increased mortality rate or dramatic emigration events that would lead to such as drastic drop in population. Whatever the reason for the adjustment, this just serves to underline the challenges of assuring and ascertaining the reliability or accuracy of population estimates. As of July 2011, Wikipedia still cites the *World Factbook*'s population estimate of 13,711,597, and this figure has most likely been perpetuated in other sources as the "current" population of Senegal.

status, for example in Mauritania, where the official language of state has flipped between French and Arabic. This raises the question of how to account for the de facto "national" language(s) in such cases and how to document such criteria accurately and consistently in instances where formal documentation on the subject does not exist.

Similarly, in terms of criteria such as language use in education, literacy development, and the extent of a literate tradition, one must be cognizant of the fact that official records do not necessarily reflect actual practices. A notable example can be taken from the case of Oromo in Ethiopia. In the 1980s, the Derg government tried to produce written materials for Oromo using the Ge'ez or Ethiopic script; however, these failed to be adopted by the Oromo people. Among the reason for this failure was the practical concern that the script was not the most suitable vehicle for this language. Furthermore, the Oromo considered the Ge'ez script to be a symbol of Amharic dominance and thus chose not to support its use (Bulcha, 1994). In later years, as part of grassroots movement for regional independence, Oromo literacy using a Latin-based script was more widely accepted and adopted by the Oromo population. Issues such as this can be found for practically any language or nation where the complexities of politics, tribal, ethnic, and religious affiliations, and unfortunately corruption play a role in language in both official and national capacities.

Importance for U.S. national interests. For those institutions who have a vested interest in language prioritization for the purposes of national security, diplomacy, and intelligence, it is a difficult task to maintain priority lists that are up-to-date with the most current trends. This is particularly true when taking into consideration the need to forecast new and/or waning priorities and to ensure that resources are available or in development to fit these changing needs. In the current world environment, one expects that most eyes are turned to newsworthy areas like Somalia where instability of government, piracy, and the threat of terrorism are worldwide concerns. One must also watch the climate in Sudan closely, where the world awaits the impending success or failure of the newly formed Republic of South Sudan, not to mention the ongoing crisis in the Darfur region. Nations like Guinea-Conakry, Côte d'Ivoire, and Nigeria have

long been in the international consciousness given the propensity for civil and/or governmental strife to flare up at a moment's notice. The so-called Arab Spring beginning in 2011 also raised the level of uncertainty for eventual peace on the continent.

The factors discussed above highlight the wide range of concerns that must be entertained in language prioritization. While the three factors mentioned thus far focus upon the very practical concerns of demographics, the role of language in the sociocultural climate of a nation or region, and the sociopolitical and strategic importance of a given language or group of languages for U.S. interests, there are even more fundamental questions that must be addressed. One of the most important among these is the question of defining what constitutes a unique language or a world region. These concerns are discussed in the sections below.

What is a language?

The deceptively simple task of defining the basic characteristics of a language becomes incredibly complicated when developing a methodology for language prioritization. The unfortunate reality is that there is no single set of criteria that can be applied to the gamut of situations in order to determine whether or not two or more languages are similar enough to be considered varieties of a single language or if they are best considered separately. Every situation is unique. As the oft-cited maxim goes: a language is "a dialect with an army and a navy" (Max Weinreich)—or as Browne (2002) later put it: with "an airline ..., a seat in the United Nations, and a soccer team with the national colors." Well-known examples can be found wherein two or more languages sharing a high degree of mutual intelligibility have come to be considered separate languages owing solely to the presence of political boundaries or different religious affiliations shared by the majority of individuals in a country. At the opposite end of the spectrum, cases can be found where languages that share little intelligibility but co-exist peacefully are sometimes mislabeled or misperceived as being a single language by some circles. Let us explore how this situation has played out in an applied way in language training.

In certain sectors of the government and military, personnel are eligible to receive handsome salary bonuses for maintaining profi-

ciency in one or more foreign languages. How might the situation be evaluated in an organization employing an Arabist who maintains proficiency in multiple varieties of Arabic that are by no means mutually intelligible, e.g. Modern Standard Arabic, Iraqi Arabic, and Sudanese Arabic? Linguistically speaking, these varieties of Arabic are not considered by most to constitute a single language when judging by contemporary standards of language classification. Indeed, the latest *Ethnologue* identifies 35 varieties of Arabic that have received unique ISO codes (assigned for unique identification of all known languages). They are, however, all called 'Arabic.' How would this situation compare, then, to that of a colleague who is trained in Persian (Iranian Farsi) where it would require minimal effort to test at an equivalent level of proficiency for the mutually intelligible languages like Dari (spoken in Afghanistan) and Tajiki (spoken in Tajikistan)? There are the other well-known cases of Hindi vs. Urdu and Serbian vs. Croatian, where the pairs of languages may well be considered to be dialects of the same language, but are merely divided by sociopolitical boundaries, with their affinities being further masked by the use of different orthographic conventions. The ambiguity of the word "Arabic" has the potential to result in an unfair disadvantage to the Arabist.

These issues also abound in the African context. An individual who is proficient in multiple Berber languages, most of which are not mutually intelligible, or one proficient in Congolese Swahili, Zanzibari Swahili, and Kenyan Swahili, (grouped together under the heading of Swahili but mutually intelligible only to some degree), will only be credited with skills in one language. An individual proficient in both Xhosa and Zulu, however, would typically be credited with proficiency in two distinct critical languages that some linguists classify as mutually intelligible varieties of a single language. The data in (1) exemplify the linguistic similarity of Xhosa, Zulu, and two other closely related Nguni languages, namely Ndebele and Siswati.

(1) Exemplification of the linguistic affinity of Nguni languages (source: "Nguni," n.d.)

Xhosa:	Ndi-qonda	ka-ncinci	nje	isi-Ngesi
Zulu:	Ngi-qonda	ka-ncane	nje	isi-Ngisi
Ndebele:	Ngi-zwisisa	ka-ncani	nje	isi-Ngisi
Siswati:	Ngi-condza	ka-ncane	nje	si-Ngisi
English gloss:	I speak	a little	of	English

'I speak a little English.'

Many of these same questions that center upon language prioritization as it relates to language proficiency and training can be extended to concerns about resource allocation for the development of literacy and educational materials, as well as for the development of assessment tests. A key component compounding these concerns is the question of "prioritization for whom"? It is to this subject that we turn next.

African languages priorities in other resources: U.S. federal agencies lists

In an effort to address the varying needs for language capacity building in different United States government entities, all U.S. federal agencies are tasked to document worldwide priority languages for their interests and operations. It should be clear from the components of Table 4 that only the United States Department of Education's list of critical languages (as used in part to facilitate evaluation for support by Title VI funding) contains a fairly comprehensive set of African languages that are considered to be high priority. Indeed, 16 of the 78 languages on the list (including Arabic) are indigenous to Africa.

Table 4. Priority of African languages by U.S. Government agency (including Arabic)

Department of Education: Akan, Amharic, Arabic (all dialects), Bamanakan (Bamana, Bambara, Mandikan, Mandingo, Maninka, Dyula), Berber (all languages), Dinka, Hausa, Igbo, Oromo, Somali, Swahili, Tigrigna, Wolof, Xhosa, Yoruba, Zulu

Department of Defense: Arabic, Hausa, Igbo, Swahili, Somali, and Yoruba

Department of State: Super critical – Arabic (MSA, Egyptian, Iraqi); Critical – Arabic (others)
Department of Agriculture: Arabic
Department of Commerce: Arabic (North Africa/Middle East)
Department of Health and Human Services: Arabic
Department of Housing and Urban Development: no recommendations
Department of Interior: Arabic (North Africa/Middle East), French (sub-Saharan Africa)
Department of Labor: Arabic
Department of Treasury: Arabic (North Africa/Middle East), Chinese, Spanish
Department of Veterans Affairs: no recommendations

Besides the Department of Education, the Department of Defense has a relatively impressive list for high priority African languages, with 6 out of the 12 languages on their list (including Arabic) being spoken on the African continent. What might be surprising to some readers is that, for the remaining agencies, the only reference to Africa (if any) is in association with French and Arabic. This state of affairs is startling if one considers the increasingly important role that Africa has come to play in global concerns on many levels. In terms of geopolitical strategizing, potential markets for goods, public health concerns, terrorism, and global population growth, Africa should be on everyone's radar. The continent experiences population growth in the ballpark of five percent annually, is home to numerous humanitarian crises, widespread HIV/AIDS concerns, political corruption and unrest, and growing threats of piracy and terrorist activity by Al-Shabaab and Al-Qaeda in the Islamic Maghreb (AQIM).

What is a region?

A question intimately related to "priority for whom" concerns the definition of the relative priority status for a language in a given world region. A complicating factor is that this definition may depend in large part on how a particular unit or organization defines their identified region or regions on interest. Consider the four displays below in Figures 2 through 5 that map the world into distinct

regions according to the United States State Department (2), United States Department of Defense Unified Combatant Commands (3), the United Nations Geoscheme (4), and the World Health Organization (5).

Figure 2. United States Department of State regions (Source: "Maps," n.d.).

In Table 1, above, we briefly introduced the world region divisions offered by the United States Department of State. The visual representation of these regions presented here in Figure 2 illustrates a somewhat prototypical division common to many other organizations. Important to our concerns in this paper, the African continent is divided into two portions, as is typical, between North Africa and Sub-Saharan Africa. Perhaps less familiar to the reader is the combatant command regions of the United States Department of Defense shown in Figure 3, which, by and large, follow a similar division as the Department of State. The exception to this generalization is the continent of Africa, which falls almost entirely under the purview of the U.S. Africa Command (AFRICOM). Although North African nations are predominantly Muslim and Arabic-speaking, Egypt is the only one among them to be grouped with other Arabic-speaking nations of the Middle East.

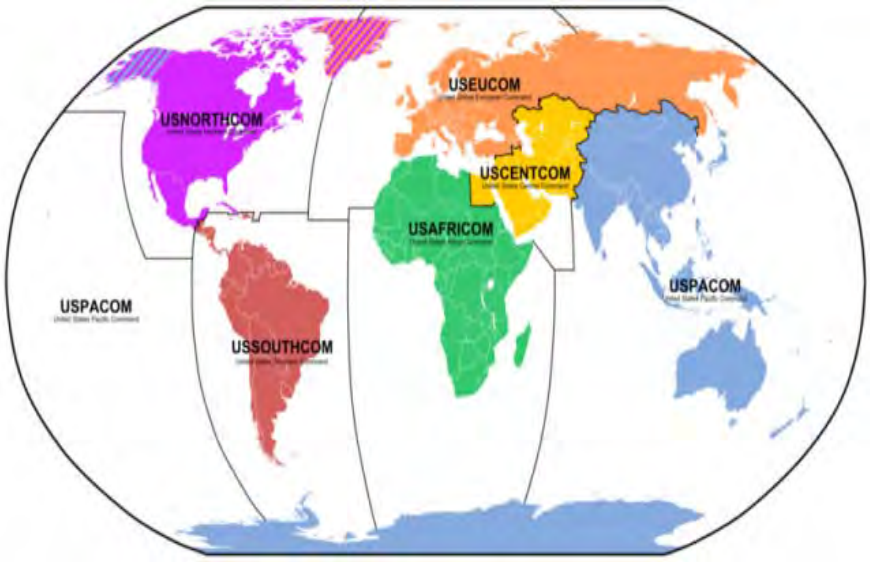

Figure 3. Unified Combatant Commands (U.S. Department of Defense) (source: "Unified," n.d.).

The UN classification of regions in Figure 4 is arguably a more ideal perspective. While many might presume that a pan-African ethos exists throughout most of sub-Saharan Africa, this is not entirely true, particularly in terms of political and economic ties and shared history. In this respect, it makes far more sense to avoid seeing "Africa" as a monolith and rather to focus on distinct sub-regions of Africa, if not individual countries. Unfortunately, it is not always possible to have the resources for such a fine-tuned focus of interests. Regardless of how one divides a given area, certain languages will naturally and undoubtedly take on relative prominence when it comes to prioritization for language capacity building. Importantly, a given organization must keep these inevitabilities and their implications and potential shortcomings in mind as they proceed in making their policy and training decisions.

Figure 4. United Nations geoscheme (source: "United Nations," n.d.).

The division of African regions according to the World Health Organization in Figure 5 is perhaps the most striking and unclear, presumably a consequence of bureaucratic gerrymandering within the administration of this organization. What is most unusual about this division is the split of the North Africa nations between the Middle East and the rest of sub-Saharan Africa. The question that must be asked concerns the exclusion of the large predominantly Muslim nations of Algeria and Mauritania from the Middle East region, while Morocco remains included in the group. Compounding this apparent conundrum is the fact that Somalia is included in the Middle East region, although it is almost always considered part of sub-Saharan Africa in other arenas.

Figure 5. World Health Organization regional offices (source: "WHO regional offices," n.d.)

In order to gain further perspective on the inherent difficulties of language prioritization in the Africa context, we turn next to a comparison of three different priority language lists developed in 1979, 2004, and 2010. These lists highlight the fact that language priorities are subject to change at the whims of current world events, policy changes, funding shortfalls (or surpluses), among other factors.

Comparison and changes in African language prioritization

Thus far, we have explored a variety of factors that come into play when devising a schema for language prioritization. These many factors are brought to the fore when one considers high priority language lists developed at three distinct points in time, namely 1979, 2004, and 2010. A comparison of these lists highlights issues related to the inconsistencies regarding "what is a language," as discussed earlier. These lists also illustrate changes in prioritization due to flare-ups and declines in sociocultural and geopolitical hot spots on the African continent.

In Table 5, the list from Dwyer (1986) resulted from the 1979 Michigan State University conference on *African Language Instruction in the U.S.: Directions and Priorities for the 1980's*. The second list from Wiley (2004) provides an update to the 1979 priority list. The third list,

containing the 2010 Title IV priorities from the United States Department of Education (DOE), is provided for comparison. Languages, as they are provided on these lists (in alphabetical order), are labeled here by letter for ease of comparison.

Table 5. Comparison of major priority lists for African languages

	Dwyer 1986	**Wiley 2004**	**U.S. DOE – Title VI**
a.	*(listed as priority B, combined with Luo and Lango)*	Acholi	
b.	*(listed as priority C)*	Afrikaans	
c.	Akan (Twi/Asante/Akuapem/Fante)	Akan/Twi	Akan (Twi-Fante)
d.	Amharic	Amharic	Amharic
e.	Arabic	Arabic	Arabic (all dialects)
f.	*(listed as priority B)*	*(listed as priority B)*	Berber (all varieties)
g.	Chewa/Nyanja	Nyanja/Chewa	
h.	*(listed as priority C)*	*(listed as priority C)*	Dinka
i.	Fulfulde (Fula/Peulh/Fulani)	Fufulde/Pulaar/Fula/Peul	
j.	Hausa	Hausa	Hausa
k.	Igbo	Igbo	Igbo
l.	*(listed as priority B)*	Kikuyu/Gikuyu	
m.	Kongo	Kongo	
n.	*(listed as priority C)*	Kpelle	
o.	*(not listed)*	Krio	
p.	*(listed as priority B)*	Luganda	
q.	Malagasy	Malagasy	
r.	Mandingo (Bambara/Mandinka/Dyula)	Bamana/Bambara/Mandingo/Mandekan/Maninka/Dyula	Bamanankan (Bamana, Bambara, Mandikan, Mandingo, Maninka, Dyula)
s	Ngala (Lingala)	Lingala	
t.	*(listed as priority B)*	Mende/Bandi/Loko	
u.	Oromo (Galla)	Oromo	Oromo
v.	Rwanda/Rundi (Kirwanda/Kirundi)	Rwanda/Rundi	
w.	Sango	*(not listed)*	

x.	Shona	Shona	
y.	Somali	Somali	Somali
z	Sotho/Tswana (Ndebele)	Sotho/ Tswana/Ndebele	
aa.	Swahili	Swahili	Swahili
bb.	*(listed as priority C)*	Temne	
cc.	Tigrinya	Tigrigna	Tigrigna
dd.	Umbundu	Umbundu	
ee.	*(not listed)*	West African Pidgin English	
ff.	Wolof	Wolof	Wolof
gg.	Xhosa/Zulu/Swazi	Xhosa, Zulu, Ndebele, Swati	Xhosa
hh.	Yoruba	Yoruba	Yoruba
ii.	*(grouped with Xhosa and Swazi)*	*(grouped with Xhosa, etc.)*	Zulu

It should be clear from this list that the prioritization decisions for African languages across time have changed in noticeable ways. The first thing to note when comparing these resources is that they raise the question of where to set the bar for establishing the highest priority languages. The original Dwyer list included a relatively small number of top priority African languages, 23 in total, but still more than what is found in the U.S. DOE's less comprehensive list of critical languages. The 2004 update from Wiley further broadened the category of "Priority A" languages. Only 11 languages are represented as top priority in all three lists, however even within some of these groupings, the languages or language varieties represented are quite different. Consider the case of Dwyer's 1979 Xhosa/Zulu/Swazi, compared to Wiley's 2004 list where Ndebele was added. Compare these, then, against the U.S. DOE list where only Xhosa and Zulu are represented, and in this case listed as separate languages. Languages like Sango were included in the 1979 list but are noticeably absent from the 2004 list at any level of priority, where nine additional language 'groups' were added, including pidgin and creole languages like West African Pidgin English and Krio. The U.S. DOE priority list from 2010 for Title VI contains still other lan-

guages, for example Berber and Dinka, that were listed at lower levels of priority in the earlier lists. Focus on these particular languages is a clear sign of the times, as the geopolitical importance of the Maghreb and Sudan continues to percolate. Still other languages that were represented in the 1979 and 2004 lists are absent from the U.S. DOE list, among them Malagasy, Lingala, Shona, and Umbundu.

These and other priority lists are often used by individuals as a starting point to identify and establish the highest priority languages in Africa from a broad perspective and from multiple viewpoints. The languages contained on these lists are largely representative of what can be considered "Priority A" languages. Beyond these languages, Priority B and C languages have also been identified; however, they have been suggested by both Wiley and Dwyer to be "open" classes that are subject to change with world events and geopolitical trends. The appearance of numerous languages from Sierra Leone and Liberia on the updated 2004 list is clearly a sign of "current events" and the criteria of "national security interests," i.e. they are remnants of headlines from roughly 2004 when the updated list was distributed. It is precisely these types of languages with perhaps fleeting importance that may be best considered for the Priority B/C lists. Similarly, the extraction of Acholi from the priority B grouping of Luo/Acholi/Lango and its subsequent placement on the Priority A list in 2004 was also a consequence of security interests and current events, though presumably the priority ranking of this language (or at least the Luo/Acholi/Lango group) will be more enduring than the relatively minor languages of Sierra Leone and Liberia. The language grouping of Luo/Acholi/Lango also exemplifies the question of whether one should set strict language priorities based on what we define as distinct languages or if we rather consider looking at clusters of related and largely mutually intelligible languages. It may ultimately be that this question is less consequential to actual language capacity building and more a question of documentation and perhaps relative ranking.

A final consideration to weigh in the comparison of these priority lists concerns the question of whether or not one should focus interest only upon indigenous languages of the region, as well as where one should set the peripheral boundaries of a given region. For example, should a priority list for Africa include Malagasy? This one

language spoken on the island nation of Madagascar is representative of an entire language family that is not technically indigenous to Africa and shares no genetic commonalities with other African languages. Afrikaans, which, to some extent, may be defined as a historical variant of Dutch, is also in question. The language is a practical working language spoken in South Africa and Namibia; however, its origins are clearly in Europe despite the incorporation and adaptation of many loanwords from other surrounding indigenous African languages. As a result of these decisions, how does one best address the many widely spoken pidgin, creole, and mixed languages that have grown out of necessity and contact between indigenous languages and colonial European languages? Such non-indigenous language varieties are deemed worthy and important by bodies like AFRICOM.

Summary and discussion: Methodology for prioritization

With these many concerns, complications, and challenges in mind, one can embark on the task of developing and implementing an informed and appropriate methodology for prioritization of the world's language in a framework that will accommodate varying levels of focus. In this section, we shall suggest components of this methodology that should be taken into consideration in planning and priority-setting for languages in the African context.

Assessment of current resources

Of key importance is availability of resources and accurate information drawn from reliable sources. Depending on the requirements, mission, or project at hand, it is important to weigh this information and its implications carefully so as maximize the outcome of the prioritization process. In light of the challenges for acquiring accurate, reliable, and current information, resources like *Ethnologue* and the *World Factbook* present a convenient means to begin evaluating languages in terms of level of priority. Such resources provide information on languages spoken in specific countries, cross-references to other countries in instances where a language group is split by national boundaries, and often, links to other closely related languages or easily confusable languages are included. Language maps are also provided for each country which can assist in familiarization of the proximity of languages to geographic areas of interest. This

information alongside what has been published in the lists cited above and that available from other organizations offers a strong starting point for prioritization planning. One must keep in mind, however, that these lists are limited in their scope in ways discussed in sections above.

Depending upon the level of language skill that will be required of individuals embarking on a particular project in the African context, it is important to bear in mind the type and availability of instructional and reference materials for a given language. While it is the case that many African languages have been documented to some degree, few comprehensive and readily available resources for these languages are commercially available. University libraries and even internet searches can often yield results containing contact information for African language experts, information on where major African languages are taught in universities or colleges, bibliographies containing valuable resources like grammars, dictionaries, and course books. Some of these materials for select languages may even be available online in open-source formats. Sample language data like word lists, grammar sketches, and even sound files for common phrases and greetings may also be available.

Efforts for establishing, updating, and servicing priorities for African languages would greatly benefit from a central resource and/or better cross-referencing between major resources. The *UCLA Language Materials Project* (http://www.lmp.ucla.edu/) provides both a convenient portal for finding learner resources for numerous less commonly taught languages and useful summary descriptions that can be used to evaluate this set of languages, which includes 26 major African languages, in terms of relative priority. Among the shortfalls of this resource, information is only provided for a select set of languages. For the most part, this reflects a comprehensive set of what most experts would judge as the top priority African languages; however, there are some notable exclusions and inconsistencies. For example, there is no coverage for the major languages of the Great Lakes region, Kinyarwanda and Rundi. This is a striking gap given that languages which have much fewer speakers and relatively less geopolitical status, such as Mende in Sierra Leone or Ewe in Ghana and Togo, are included. Perhaps a greater shortfall is the fact that much of the information is not updated regularly, and this re-

source may be on the decline like some other projects which otherwise held promise for facilitating language capacity building and planning.

The now defunct Michigan State University *Webbook of African Language Resources* (http://africa.isp.msu.edu/afrlang/hiermenu.htm) was another resource with potential for facilitating priority setting and capacity building for African languages. As a resource dedicated to African languages, the coverage of languages went much deeper than that of the *UCLA Language Materials Project*, drawing on the language lists established at the 1979 MSU conference discussed in previous sections. Unfortunately, the database backing this resource was not updated regularly and the *Webbook* was officially archived in 2010. Additional resources, such as *Web Resources for African Languages* (http://www.africanlanguages.org/) and the *Pan-African Localisation* project (PanAfriL10n, PanAfrLoc, http://www.bisharat.net/PanAfrLoc/), the latter having drawn a substantial amount of information from the MSU *Webbook*, are also in a state of flux. The *National African Language Resource Center* (http://nalrc.wisc.edu/) has potential to carry on the mission for facilitating teaching and learning of major African language, but they appear to have started from scratch and tend to highlight specific resources produced through their center.

Steps towards improved management and maintenance for data on African languages

Whether carried out by a new central resource or through the revival of existing resources with regular maintenance and better cross-referencing, additional enhancements for data management are recommendable based on the challenges of language prioritization discussed earlier. First of all, data should be keyed by the ISO (International Organization for Standardization) codes documented in and updated by *Ethnologue* to facilitate data maintenance, integration of data for a variety of user needs, and cross-referencing of information between resources.

Furthermore, efforts should be made to augment the set of standard codes for "macro languages"—that is, specialized codes that help group together related languages varieties such as [ara] for some 30 varieties of Arabic spoken throughout the world or [orm] for the three distinct varieties of Oromo, as documented by *Ethnologue*. The

number of codes established for this purpose is far fewer than what is warranted, currently limited to Arabic [ara], Dinka [din], Gbaya [gba], Grebo [grb], Fula [ful], Kanuri [kau], Kongo [kon], Kpelle [kpe], Luyia [luy], Malagasy [mlg], Mandingo [man], Swahili [swa], and Tamashek [tmh] in Africa. Among others, notable groupings which would help facilitate the varying levels of interest in African language prioritization as discussed in this article include the Nguni languages and a more comprehensive set of Berber languages. On the other end of the spectrum, where fine-tuned regional focus is called for, standards for distinguishing information on dialects of a particular language variety should be established. This step will come with substantial challenges, foremost being the accuracy and acceptability of distinguishable dialects; but a simple starting measure can be to tie language ISO codes with geographic locations (city names, country codes) as well as drawing on dialectal varieties documented in *Ethnologue*.

While the reliability of any population estimate may always be called into question (as was made apparent in the opening section of this article), a database of information servicing language prioritization needs should be as systematic as possible in terms of accurately and completely documenting the sources for population figures. Furthermore, greater efforts should be made to estimate or collect data on L2 (second language) speakers of African languages and to better document the scope of use of colonial languages and a better account of the distinctive pidgins and creoles spoken on the African continent. As means to obtain updates and verification of not only demographic information but also details of socio-cultural importance and political status, summary pages for individual languages and countries can be set up as wikis open to vetted experts or open to a wider community of contributors with the submissions verified by designated experts.

As much as Africa has been branded with the image of the "dark continent," there is actually a wealth of information about African languages and cultures that can be tapped into for purposes of setting language priorities and helping individuals gain training in and familiarity with these languages. We should endeavor to reverse the trend whereby resources that are established for such purposes end up fading out of use (whether due to lack upkeep or loss of fund-

ing in this era of economic difficulty), with data collected to support these resources becoming inaccessible. Improved database management methods such as those touched upon in this discussion will not only help preserve the longevity and utility of the resources served by these data, but also help provide a flexible system that can respond to varying requirements and levels of interest for the prioritization of African languages.

References

Browne, W. (2002). What is a standard language good for, and who gets to have one? (The Kenneth E. Naylor Memorial Lecture Series in South Slavic Linguistics.) Columbus, OH: The Ohio State University Department of Slavic and East European Languages and Literatures.

Bulcha, M. (1994). The language policies of Ethiopian regimes and the history of written Afaan Oromoo: 1844-1994. *The Journal of Oromo Studies, 1*, 91–115.

Cartography: The true size of Africa. (2010, November 10). *The Economist*.

Dwyer, D. J. (1986). A resource handbook for African languages. A listing of the institutional, human, and material resources for the teaching and learning of the eight-two highest priority African languages.

Global Mapping International. (n.d.). Steve Huffman language maps: An academic application of the World Language Mapping System. *World GeoDatasets*. Retrieved from July 7, 2011 from http://www.worldgeodatasets.com/language/huffman.

Johnstone, P. (1993). *Operation world*. Grand Rapids, MI: Zondervan Publishing House.

Lewis, M. P. (Ed.). (2009). *Ethnologue: Languages of the world* (16th ed.). Dallas, Tex.: SIL International. Retrieved from http://www.ethnologue.com.

Maps and country information. (n.d.). U.S. Department of State. Retrieved August 1, 2011 from http://future.state.gov/where/maps.html.

Mufwene, S. S. (2008). *Language evolution: Contact, competition and change*. London & New York: Continuum.

Nguni languages. (n.d.). In *Wikipedia*. Retrieved April 3, 2011 from http://en.wikipedia.org/wiki/Nguni_languages.

Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. (2010). *World Population Prospects: The 2010 Revision*. Retrieved August 10, 2011 from http://esa.un.org/unpd/wpp/index.htm.

Senegal. (2011). *The World Factbook* [online]. Washington, DC: U.S. Central Intelligence Agency. Retrieved January 11, 2011 from

https://www.cia.gov/library/publications/the-world-factbook/geos/sg.html.
Thomason, S. G. (Ed.). (1997). *Contact languages: A wider perspective*. Amsterdam: John Benjamins.
Unified Combatant Command. (n.d.). In *Wikipedia*. Retrieved April 3, 2011 from http://en.wikipedia.org/wiki/Unified_Combatant_Command.
Unified Nations geoscheme. (n.d.). In *Wikipedia*. Retrieved April 3, 2011 from http://en.wikipedia.org/wiki/United_Nations_Geoscheme.
WHO regional offices. (n.d.). World Health Organization. Retrieved April 3, 2011 from http://www.who.int/about/regions/en.
Wiesenfeld, L. P. (1999). *World almanac and book of facts*. Mahwah, NJ: World Almanac Books.
Wiley, D. (2004). Priorities for U.S. instruction in the African languages: New criteria and new choices. Unpublished manuscript. Retrieved from http://elctl.isp.msu.edu/archive/regions/africa.php.
Wiley, D., & Dwyer, D. (1980). African language instruction in the United States: Directions and priorities for the 1980s. East Lansing, Michigan: Michigan State University, The African Studies Center.
The World Factbook 2011. (2011). Washington, DC: Central Intelligence Agency. Accessed April 3, 2011 from https://www.cia.gov/library/publications/the-world-factbook/index.html.

Author note

We are grateful for helpful comments from numerous colleagues, especially Nikki Adams and Michelle Morrison.

Correspondence concerning this article should be addressed to Tristan M. Purvis, University of Maryland Center for Advanced Study of Language, Box 25, College Park, MD 20742. E-mail: tpurvis@umd.edu