1. A survey of African languages

Harald Hammarström

1.1. Introduction

The African continent harbors upwards of 2,000 spoken indigenous languages – more than a fourth of the world’s total. Using ISO 639-3 language/dialect divisions and including extinct languages for which evidence exists, the tally comes to 2,169. The main criterion for the ISO 639-3 language identification is mutual intelligibility, but these divisions are not infrequently conflated with sociopolitical criteria. This causes the tally to be higher than if the language/dialect division were to be based solely on intelligibility. Based solely on mutual intelligibility, the number would be approximately 85% of the said figure (Hammarström 2015: 733), thus around 1,850 mutually unintelligible languages in Africa. A lower count of 1,441 is obtained by treating dialect chains whose endpoints are not mutually intelligible as one and the same language (Maho 2004).

The amount of information available on the language situation varies across different areas of Africa, but the entire continent has been surveyed for spoken L1 languages on the surface at least once. However, so-called “hidden” languages that escaped earlier surveys continue to be discovered every year. These are all languages that are spoken by a (usually aging) fraction of a population who otherwise speak another (already known) language. The least surveyed areas of Africa include Northern Nigeria, Eastern Chad, South Sudan and various spots in the Republic of Congo, the Democratic Republic of Congo and Angola.

The situation is entirely different with respect to sign languages (cf. Padden 2010: 19). Almost no surveys of sign languages have been carried out, but some admittedly incomplete listings are available (Kamei 2004; Nyst 2010; Schmaling 2012). Sign languages in Africa include both indigenous rural sign languages such as Adamorobe Sign Language in Ghana (Nyst 2007) and varieties of the overseas sign languages American Sign Language, British Sign Language, Langue de Signes Française and occasionally other European sign languages, taught in deaf schools that are tied to nation states. There is reason to believe there are proportionately more sign languages in sub-Saharan Africa than in Western countries due to the prevalence of bacterial meningitis (Molesworth et al. 2002). Table 1 lists sign languages in Africa and the Arabian peninsula so far identified in the literature, though there is insufficient information to confirm or deny that each one is mutually intelligible with the others or its overseas progenitor.

The surveying of other kinds of languages, such as initiation languages (Moñino 1977; Ngonga-ke Mbembe 2009), ritual languages (Brindle et al. 2015), or secret languages (Leiris 1948; Dugast 1950; Ittman 1959; Leslau 1964), can
Table 1: Sign languages in Africa and the Arabian peninsula so far identified in the literature. There is insufficient information to confirm or deny that each one is mutually intelligible with the others or its overseas progenitor.

<table>
<thead>
<tr>
<th>Sign Language</th>
<th>ISO 639-3</th>
<th>Progenitor/ Sign Language Family</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt Sign Language</td>
<td>esl</td>
<td>Arab SL</td>
<td>Hendriks and Zeshan 2009</td>
</tr>
<tr>
<td>Kuwaiti Sign Language</td>
<td></td>
<td>Arab SL</td>
<td>Al-Fityani and Padden 2010</td>
</tr>
<tr>
<td>Libyan Sign Language</td>
<td>lbs</td>
<td>Arab SL</td>
<td>Al-Fityani and Padden 2010</td>
</tr>
<tr>
<td>Yemeni Sign Language</td>
<td></td>
<td>Arab SL</td>
<td>Hendriks and Zeshan 2009</td>
</tr>
<tr>
<td>Congo-Brazzaville Sign Language</td>
<td>–</td>
<td>American SL</td>
<td>Dalle 1997</td>
</tr>
<tr>
<td>Ethiopian Sign Language</td>
<td>eth</td>
<td>American SL</td>
<td>Tamene 2015</td>
</tr>
<tr>
<td>Gambian Sign Language</td>
<td></td>
<td>American SL</td>
<td>Nyst 2010</td>
</tr>
<tr>
<td>Guinean Sign Language</td>
<td>gus</td>
<td>American SL</td>
<td>Nyst 2010</td>
</tr>
<tr>
<td>Moroccan Sign Language</td>
<td>xms</td>
<td>American SL</td>
<td>Wismann and Walsh 1987</td>
</tr>
<tr>
<td>Nigerian Sign Language</td>
<td>nsi</td>
<td>American SL</td>
<td>Nyst 2010; Odusanya 2000</td>
</tr>
<tr>
<td>Sierra Leone Sign Language</td>
<td>sgx</td>
<td>American SL</td>
<td>Nyst 2010</td>
</tr>
<tr>
<td>Ugandan Sign Language</td>
<td>ugn</td>
<td>American SL</td>
<td>Lutalo Kiingi 2014</td>
</tr>
<tr>
<td>Namibian Sign Language</td>
<td>nbs</td>
<td>British SL</td>
<td>Madison 2005</td>
</tr>
<tr>
<td>Eritrean Sign Language</td>
<td></td>
<td>Finnish SL</td>
<td>Moges 2015</td>
</tr>
<tr>
<td>Algerian Sign Language</td>
<td>asp</td>
<td>LS Française</td>
<td>Delaporte 2008</td>
</tr>
<tr>
<td>Tunisian Sign Language</td>
<td>tse</td>
<td>Italian SL</td>
<td>Khayech 2014</td>
</tr>
<tr>
<td>Malagasy Sign Language</td>
<td>mzc</td>
<td>Norwegian SL</td>
<td>Minoura 2012</td>
</tr>
<tr>
<td>Ghardaia Sign Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guinea-Bissau Sign Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Sayyid Bedouin Sign Language</td>
<td>syy</td>
<td></td>
<td>Sandler et al. 2014</td>
</tr>
<tr>
<td>Saudi Sign Language</td>
<td>sdl</td>
<td></td>
<td>Sprenger and Mathur 2012</td>
</tr>
<tr>
<td>Lesotho Sign Language</td>
<td></td>
<td></td>
<td>Machobane et al. 2010</td>
</tr>
<tr>
<td>Mozambican Sign Language</td>
<td>mzy</td>
<td></td>
<td>Ngunga 2013</td>
</tr>
<tr>
<td>Mbour Sign Language</td>
<td></td>
<td></td>
<td>Nyst 2010</td>
</tr>
<tr>
<td>Langue de Sign Malienne / Bamako</td>
<td>bog</td>
<td></td>
<td>Nyst 2015</td>
</tr>
<tr>
<td>Burkina Faso Sign Language</td>
<td></td>
<td></td>
<td>Nyst 2010</td>
</tr>
<tr>
<td>Chadian Sign Language</td>
<td>cds</td>
<td></td>
<td>Lewis et al. 2015</td>
</tr>
<tr>
<td>Nanabin Sign Language</td>
<td></td>
<td></td>
<td>Nyst 2010</td>
</tr>
<tr>
<td>Tebul Sign Language</td>
<td>tsy</td>
<td></td>
<td>Nyst et al. 2012</td>
</tr>
<tr>
<td>Adamorobe Sign Language</td>
<td>ads</td>
<td></td>
<td>Nyst 2007</td>
</tr>
<tr>
<td>Bura Sign Language</td>
<td></td>
<td></td>
<td>Blench and Warren 2003</td>
</tr>
<tr>
<td>Yoruba Sign Language</td>
<td></td>
<td></td>
<td>Orie 2012</td>
</tr>
<tr>
<td>Hausa Sign Language</td>
<td>hsl</td>
<td></td>
<td>Schmaling 2000</td>
</tr>
<tr>
<td>Mofu-Gudur Sign Language</td>
<td></td>
<td></td>
<td>Sorin-Barreteau 1996</td>
</tr>
<tr>
<td>Rwandan Sign Language</td>
<td></td>
<td></td>
<td>Woolley 2009</td>
</tr>
<tr>
<td>Tanzanian Sign Language</td>
<td>tza</td>
<td></td>
<td>Muzale 2004</td>
</tr>
</tbody>
</table>
be expected to be incomplete. Similarly, it has not been systematically investigated which languages can be whistled and which can be drummed, but there are a number of examples (Betz 1891; Carrington 1949; Arom and Cloarec-Heiss 1976; Meyer 2015: 6, 11, 16, 23–24, 30).

The colonial languages – predominantly French, English and Portuguese – are strong in their respective nation states. Most of the colonial languages have developed a local variety in Africa, e.g., French in Burkina Faso (Carriere-Prignitz 1999) or Burundi (Bastin 1979), Italian in Eritrea-Ethiopia (Habte-Mariam 1976) and not least Afrikaans in South Africa and Namibia (Ponelis 1993). Languages from the Indian subcontinent are also found with a specifically African flavor on the eastern coast (Neale 1971; Lodhi 2005) and in South Africa (Mesthrie 1988).

Creoles, i.e., languages that (are presumed to) have expanded from pidgins, are spoken in West Africa. Kabuverdianu, Upper Guinea Crioulo, Annobonese (aka Fa d’Ambu), Angolar, Säotomense and Principense are Portuguese-lexified (Hagemeijer 2011) while Krio, Pichi (aka Equatorial Guinean Pidgin), Ghanaian Pidgin, Cameroon Pidgin, Nigerian Pidgin are derived from English (Huber 1999: 75–134). The Arabic expansion exhibits a spectrum of contact varieties, including rapidly nativised lects (Owens 1997; Luffin 2011; and below, in section 2.1). Simplified high-contact languages of indigenous lexical stock developed after contact with Europeans along the major rivers, namely, Lingala (from the Bantu language Bangi; Meeuwis 2013), Kituba (in two forms, from Koongo; Mufwene 1997) and Sango (from Ngbandi; Samarin 1982).

Several African capitals have seen the development of urban youth languages (Kießling and Mous 2004), i.e., a local vehicular language more or less consciously altered to assert the speakers’ separate identity. Table 2 lists the urban youth languages so far identified in Africa. The rampant development of urban youth languages in Africa has yet to be explained in sociodemographic terms, but
can plausibly be linked to the demographic profile of many African urban areas, where adolescents make up a large fraction of the population in streets, markets and transportation hubs\(^1\) (cf. Hollington and Nassenstein 2015).

With Africa’s population of over a billion and just over 2,000 languages, the average population speaking a given African language is half a million speakers. But the number of speakers is not evenly distributed. Among the largest languages we find Egyptian Arabic (Egypt), Nigerian Pidgin (Nigeria), Maghreb Arabic (Morocco/Tunisia), Hausa (Nigeria), Yoruba (Nigeria), Swahili (Tanzania/Kenya), Amharic (Ethiopia) and Oromo (Ethiopia) with over 20 million speakers each and a large number of L2 speakers in addition. Reliable figures for speaker numbers for comparison across Africa are lacking, not least because of rapidly fluctuating language alliances and demographic change. Unless otherwise noted, speaker numbers will be cited from Lewis et al. (2015), which has the most extensive coverage. The median number of speakers is 26,400, thus over 1,000 languages have fewer speakers than this. The distribution of languages is relatively sparse in the North African countries and dense all over the sub-Saharan area. Spots of even higher language density are the mountainous Nigeria–Cameroon border area, the Nuba mountains and the lowlands of southwestern Ethiopia. Large countries predictably have hundreds of languages, with the largest, Nigeria, home to over 500 languages. Most sub-Saharan countries encompass dozens of languages and a only a few small countries, e.  g., Burundi, are linguistically uniform. Although hard figures are lacking, bilingualism and multilingualism appear to have been the norm for African traditional societies. Cases where entire communities speak five or more languages are known from sub-Saharan Africa (e.  g., Lionnet 2010: 2; Lüpke 2013).

Chapter 2 in the present volume discusses the evidence regarding genealogical classification in Africa, covering every (sub-)family. A complete listing of individual African languages can be found via Ethnologue (Lewis et al. 2015) and Glottolog (Hammarström et al. 2015). Lewis et al. (2015) has information on speaker numbers, the geographic extent and division of languages/dialects, endangerment and other relevant metadata. Hammarström et al. (2015) encompasses a comprehensive bibliography with over 60,000 entries deriving from the Electronic Bibliography of African Languages and Linguistics by Jouni Filip Maho, and a genealogical classification that includes every individual language.

Describing the more than 2,000 African languages is an enormous task. Regarding the present state, fully systematic figures on description level are available for grammatical description only (as opposed to textual, audiovisual or lexical documentation). The figures relating to grammatical description (Hammarström et al. 2015) allow for some surface comparisons as follows. For every language we consider all the descriptive publications relating to it and count its most extensive description, i.  e., the highest level attained by at least one publication, according to

\(^1\) I owe this suggestion to Jeffrey Heath (p.  c. 2015).
Table 2: Urban youth languages so far identified in the literature. Some named urban youth languages can be formed from any of a number of local base languages. As such they do not represent a modification of a specific language, but the same kind of modification applicable to various base languages. These are indicated with a plural (s) in the listing.

<table>
<thead>
<tr>
<th>Urban Youth Language</th>
<th>Base Language</th>
<th>Urban Center</th>
<th>Country</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camfranglais</td>
<td>French</td>
<td>Douala/Yaoundé</td>
<td>Cameroon</td>
<td>Kießling 2004</td>
</tr>
<tr>
<td>Sango Godobé</td>
<td>Sango</td>
<td>Bangui</td>
<td>CAR</td>
<td>Landi and Pasch 2015</td>
</tr>
<tr>
<td>Nouchi</td>
<td>French</td>
<td>Abidjan etc</td>
<td>Côte d’Ivoire</td>
<td>Kube 2005</td>
</tr>
<tr>
<td>Indoubil Inverted</td>
<td>Congo Swahili</td>
<td>Bukavu</td>
<td>DR Congo</td>
<td>Goyvaerts 1988</td>
</tr>
<tr>
<td>Kindoubil Kindubile</td>
<td>Lingala</td>
<td>Kisangani</td>
<td>DR Congo</td>
<td>Wilson 2015</td>
</tr>
<tr>
<td>Kindoubil Langila</td>
<td>Lingala</td>
<td>Kisangani</td>
<td>DR Congo</td>
<td>Wilson 2015</td>
</tr>
<tr>
<td>Yanké</td>
<td>Lingala</td>
<td>Kinshasa</td>
<td>DR Congo</td>
<td>Nassenstein 2015a</td>
</tr>
<tr>
<td>Yabâcrane</td>
<td>Congo Swahili</td>
<td>Goma</td>
<td>DR Congo</td>
<td>Nassenstein 2016</td>
</tr>
<tr>
<td>Egyptian Arabic</td>
<td>Egyptian Arabic</td>
<td>Cairo</td>
<td>Egypt</td>
<td>Manfredi and Pereira 2013</td>
</tr>
<tr>
<td>Yarada</td>
<td>Amharic</td>
<td>Addis Ababa</td>
<td>Ethiopia</td>
<td>Hollington 2015</td>
</tr>
<tr>
<td>K’wank’wa Sheng</td>
<td>Swahili</td>
<td>Nairobi</td>
<td>Kenya</td>
<td>Rudd 2008</td>
</tr>
<tr>
<td>Sheng ya Kijaka</td>
<td>Dholuo</td>
<td>Kisumu</td>
<td>Kenya</td>
<td>Rüsch 2016</td>
</tr>
<tr>
<td>Chibrazi(s)</td>
<td>Chewa / Tonga /</td>
<td>various</td>
<td>Malawi</td>
<td>Kamanga 2014</td>
</tr>
<tr>
<td></td>
<td>Tumbuka</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maghreb Arabic</td>
<td>Maghrab</td>
<td>various</td>
<td>Morocco</td>
<td>Manfredi and Pereira 2013</td>
</tr>
<tr>
<td>youth language(s)</td>
<td>Arabic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randuk</td>
<td>Sudanese Arabic</td>
<td>Khartoum</td>
<td>North Sudan</td>
<td>Manfredi 2008; Mugaddam 2015</td>
</tr>
<tr>
<td>Imvugo y’Umuhanda</td>
<td>Kinyarwanda</td>
<td>Kigali</td>
<td>Rwanda</td>
<td>Nassenstein 2015b</td>
</tr>
<tr>
<td>Tsotsitaal(s)</td>
<td>Any major South</td>
<td>various</td>
<td>South Africa</td>
<td>Hurst 2015</td>
</tr>
<tr>
<td></td>
<td>African language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rendók ta Juba</td>
<td>Juba Arabic</td>
<td>Juba</td>
<td>South Sudan</td>
<td>Miller 2004</td>
</tr>
<tr>
<td>Lugha za Mitaaani(s)</td>
<td>Swahili</td>
<td>various</td>
<td>Tanzania</td>
<td>Reuster-Jahn and Kießling 2006</td>
</tr>
<tr>
<td>Leb pa Bulu</td>
<td>Acholi</td>
<td>Gulu</td>
<td>Uganda</td>
<td>Rüsch and Nassenstein 2015</td>
</tr>
<tr>
<td>Luyaayi</td>
<td>Ganda</td>
<td>Kampala, etc.</td>
<td>Uganda</td>
<td>Naluwooza 1995</td>
</tr>
</tbody>
</table>
the hierarchy in Table 3. As the figures in Table 3 show, grammatical information is lacking for almost half of Africa’s languages. The geographical distribution of described versus undescribed languages is fairly uniform on the geographical distribution of languages itself (Figure 1), making Nigeria the major country of undescribed languages. The average level of description measured numerically from 0 to 5 (as per the numerical score of Table 3) is 2.67, leaving Africa the second least known continent (after Oceania, dominated by the New Guinea area). Figure 2 shows how the average description level has increased yearly since 1500. Despite one and a half centuries of intensive investigation, especially since Koelle (1854), as much as has been done remains to be done.

Figure 1: The geographical distribution of language description in Africa. Legend: Red (minimal/overview/wordlist), orange red (phonology/text/dictionary/specific feature), orange (grammar sketch), green (grammar), dark green (long grammar).
Table 3: Statistics on the type of most extensive description for the languages of Africa computed from the bibliographical database of Hammarström et al. (2015), largely based on the *Electronic Bibliography of African Languages and Linguistics* by Jouni Filip Maho.

<table>
<thead>
<tr>
<th>Num. Score</th>
<th>Most Extensive Grammatical Description Type</th>
<th># languages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>extensive description of most elements of the grammar ≈300+ pages</td>
<td>411 (18.9%)</td>
</tr>
<tr>
<td>4</td>
<td>a description of most elements of the grammar ≈150 pages</td>
<td>243 (11.1%)</td>
</tr>
<tr>
<td>3</td>
<td>a less extensive description of many elements of the grammar ≈50 pages</td>
<td>562 (25.9%)</td>
</tr>
<tr>
<td>2</td>
<td>description of some element of grammar (i.e., noun class system, verb morphology, etc.)</td>
<td>157 (7.2%)</td>
</tr>
<tr>
<td>2</td>
<td>a description of the sound inventory utilizing minimal pairs</td>
<td>82 (3.7%)</td>
</tr>
<tr>
<td>2</td>
<td>≈75+ pages</td>
<td>53 (2.4%)</td>
</tr>
<tr>
<td>2</td>
<td>text material</td>
<td>13 (0.5%)</td>
</tr>
<tr>
<td>1</td>
<td>≈100–200 words</td>
<td>476 (21.9%)</td>
</tr>
<tr>
<td>0</td>
<td>a small number of morphemes</td>
<td>124 (5.7%)</td>
</tr>
<tr>
<td>0</td>
<td>document with meta-information about the language (i.e., where spoken, non-intelligibility to other languages, etc.)</td>
<td>48 (2.2%)</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>2,169</strong></td>
</tr>
</tbody>
</table>

![Graph showing the fraction of African languages for different types of descriptions](image)
Figure 2: The average description level of African languages over time. The average description level (y-axis) is the average numerical score of the most extensive description for each African language available at the year indicated by the x-axis.

1.2. Survey by region

The people speaking indigenous African languages inhabit a variety of geoclimatic zones and display a range of cultural expressions. The survey of Baumann (1975) divided precolonial Africa into 26 zones based primarily on subsistence type, but also in consideration of other cultural factors (Figure 3). In the present survey we follow these divisions, grouping them into nine larger regions as per linguistic contingencies that extend across the divisions (Figure 4). Table 4 tabulates a concordance between the two divisions.
Figure 3: A culture-based division of Africa into 26 regions, adapted from Baumann (1975: 378).
Figure 4: Divisions used in the present chapter.
### Table 4: Concordance between the regional divisions used in the present chapter and those of Baumann (1975: 378).

<table>
<thead>
<tr>
<th>Region in the present chapter</th>
<th>Baumann (1975: 378)'s classification according to <em>Kulturprovinz</em> (‘culture area’)</th>
</tr>
</thead>
</table>
| North Africa and the Arabian Peninsula | 24. NW-Afrika (u. Sahara-Oasen)  
25. Nil-Oase und Nubien |
| The Greater Saharan Area | 23. Sahara und Sahel |
| Ethiopic Sphere | 12. Osthorn-Provinz (ostkuschitische Hirten)  
13. West-Äthiopien  
14. Hoch-Äthiopien |
| Sudanic West Africa | 19. Ostatlantische Provinz  
20. Ober-Volta-Provinz  
21. Westatlantische Provinz  
22. Ober-Niger-Provinz |
| Central Sudan and Cameroon Grassland | 15. NO-Sudan-Provinz  
16. SO-Sudan-Provinz  
17. Kamerun-Nigeria-Pufferzone  
18. Zentral-Sudan-Provinz |
| Equatorial Rain Forest | 7. Südkongo-Provinz  
8. Nordkongo-Gabun-Provinz  
9. Wald-Wildbeuter (Pygmäen) |
| Eastern Savanna | 10. Ost-Bantu-Provinz  
11. Niloten |
| Southern Tip | 1. Die Steppenwildbeuter SW-Afrikas  
2. Khoikhoi (Hottentotten)  
3. SW-Bantu-Provinz  
4. SO-Bantu-Provinz  
5. “Zwischenfluß”-Provinz (“Simabwe-Provinz”)  
6. Sambesi-Angola-Provinz |
| Madagascar | 26. Madagaskar |

1.2.1. North Africa and the Arabian Peninsula

North Africa was an integral part of the Mediterranean political scene in antiquity. As part of the Roman empire, Christianity made significant inroads in the centuries after Christ. The traditional economy was based on grain agriculture, olive plantations and livestock farming. The domesticated dromedary, with its numerous adaptations to a desert climate, arrived from the east and was established only in late antiquity.
While at present the area comprising North Africa and the Arabian Peninsula harbors relatively few languages, it presumably hosted a larger array of languages in the past few millennia. Written records only allow the identification of a dozen or so, before the spectacular takeover of Arabic, in the wake of the Islamic expansion (Retsö 2003, see below). A few non-indigenous languages from the Mediterranean are amply attested on African soil in the past, such as Punic (Kerr 2007) and Greek (Torallas Tovar 2010) but are no longer spoken there. Peripatetic people with roots in India (“gypsies”) are found in the Maghreb, Egypt and Sudan (Ḥannā 1982; Streck 1989; Weber 1989) but no case of an Indo-Aryan language retained as a mother tongue is attested.

The entire North African area is now dominated by varieties descending from Arabic spoken in the South Arabian peninsula in the seventh century. Conventional divisions of African Arabic varieties recognize a North African group (Pereira 2011), corresponding to an early expansion in the century following the birth of Islam. The North African group includes the Arabic varieties in Libya, Malta, Algeria, Tunisia, Morocco and Andalusia (now extinct) as well as the Hassaniyya dialect at the western end in Mauretania. Another expansion, many centuries later and much slower, is the line going southwest from Egypt. The Sudanese-Chadian Arabic subgroup stretches from North Sudan, through Darfur into Chad, northern Cameroon and northeastern Nigeria (Kaye 1976). The varieties so far mentioned do not show signs of regularization due to second language imperfect learning, and the westernmost Nigerian Arabic retains the unusual emphatic series of consonants (Owens 1993). But there were (and probably still are, cf. Bell 1975b) Arabic pidgins in the Arabic border area in the past. One variety documented in the early 20th century is dubbed Turku (‘soldiers’) because it was used by soldiers in the Sara-speaking area in Chad (Tosco and Owens 1993). Another is Bongor Arabic, spoken around the town of that name in Chad (Luffin 2013). Two other named high-contact Arabic varieties originating in southern Sudan went on to become mother tongues. In the late 20th century, an army from North Sudan became stranded in South Sudan. They recruited locally into the lower ranks and after a series of turns were incorporated into the Imperial British East Africa company. Subsequently demilitarized, communities are now found in Uganda and Kenya (Wellens 2003; Luffin 2005) under the name Nubi. Independently, in South Sudan the role of Arabic continued to play a major role as a lingua franca. Juba is the main urban center and hence the Arabic emerging here is termed Juba Arabic (aka Sudanese Creole Arabic). Rather than a single uniform Juba Arabic, there is a range of varieties depending on individual speaker background, including natively spoken Juba Arabic. Among the core defining features of Juba Arabic we find influences from the local (mostly Nilotic) languages and regularizations such that, e. g., inherited Arabic broken plurals are no longer productive (Smith and Ama 1985). Although not on an expansion frontier, the large influx of foreign workers from South Asia into the Gulf States of the Arabian peninsula has given rise to several varieties of pidginized Arabic (Almoaily 2014).
The harsh climate has kept the Arabian Peninsula relatively sparsely populated except for the mountainous and coastal south, where a succession of small kingdoms emerged. The languages of these kingdoms belong to the Sayhadic subgroup of Semitic and are attested from the southwest of the Arabian peninsula in the centuries 800 BC–500 AD (Nebes and Stein 2008). Sabaic is the best-known representative, for which there are enough inscriptions to compile a grammar sketch (Beeston 1962). The inscriptions are written in a family of South Semitic scripts that are ultimately akin to the Hebrew, Greek and various other well-known scripts. Inscriptions were made on rock as well as palm-leaf stalks and wooden sticks and included building inscriptions, legal texts, graffiti, letters, deeds, contracts and so on. There are also inscriptions from this time of a language difficult to interpret as Sayhadic, dubbed “unknown” by Beeston (1981) and associated with the tribal name Himyaritic by Stein (2008). The island of Dhofār holds another set of undeciphered inscriptions called the Dhofār dipinti inscriptions (Al-Jallad 2014: 13).

The insessional languages survived maximally until perhaps a millennium ago. A possible modern reflex is the variety of Arabic of Jebel Rāzih in Yemen, famously unintelligible to other modern Arabic varieties in the peninsula. Arabic of Jebel Rāzih contains morphological features otherwise only found in Sayhadic. These must have either been taken over at some stage in the Arabic incursion, or Jebel Rāzih is a relexified Sayhadic continuation (Watson et al. 2006).

A different subgroup of West Semitic languages are the half a dozen Modern South Arabian languages, endangered but surviving in Yemen and Oman (Simeone-Senelle 2011). They are not descendants of either Arabic or the Sayhadic (also known as Ancient South Arabian) languages. The best-known language is Mehri (Rubin 2010). Further witnesses to other pre-Arabic languages of the Arabian peninsula are absent. Though a non-pastoral lifestyle remains – the Solubba (Pieper 1923; Dostal 1956) – if they represent a continuation of a pre-Arabic ethnic group, nothing has survived of their language.

Oman harbors one Indo-European language, Kumzari, originating as a Southwestern Iranian variety but subsequently influenced by Shihhi Arabic to the extent that synchronically it consists equally of elements from either source (van der Wal Anonby 2015).

1.2.1.1. The greater Nile area

Moving west to Ancient Egypt, hieroglyphic (on stone) and hieratic (on papyrus) records attest the Ancient Egyptian (3000–2000 BC), Middle Egyptian (2000–1300 BC) and Late Egyptian (1300–700 BC) stages. Later stages of this language are Demotic (700 BC–500 AD) written in a script derived from hieratic (Loprieno 1995) and the final stage is known as Coptic, written in a Greek-derived script. The ability to read hieroglyphs was lost in antiquity, and famously deciphered by Jean-François Champollion using trilingual texts and preceding work by Thomas
Harald Hammarström

Young (Allen 1960). Coptic, on the other hand, was spoken until the 17th century and is in fact the first ever African language to be described by a grammar sketch, by an Arabic scholar in the 1300s (Bauer 1972 [1300]). Coptic was also studied by early European scholars (e.g., Tuki 1778), and it continues to be used for liturgical purposes today. The Egyptian-Coptic language represents 4,000 years of recorded history, longer than any other known language.

The Egyptians made no organized effort to describe the languages or language situation further up the Nile, but the various ethnicities depicted or mentioned throughout the millennia must have represented various languages. One such language that can be safely identified is the language of the Meroe civilization (300 BC–400 AD) in what is now North Sudan. This language was written in a script of its own, discovered in the early 20th century. Thanks to the existence of name equations to hieroglyphic Egyptian, the script was deciphered shortly after discovery (Griffith 1909), and is alphabetic, making it one of the earliest known alphabetic scripts. However, beyond personal names, the meaning of the individual words and sentences recorded in this script still remains elusive. No more than a couple of dozen have a convincing interpretation. With the assumption that Meroitic forms resembling forms in languages found later in the area are related, the meaning of a few dozen more can be posited (Rilly 2009). One more indigenous language of the antiquities in the Nile region is amply attested in texts from 800–1500 AD in a Greek-derived script, namely Old Nubian in what is now North Sudan, the ancestor of Nobiin (300,000 speakers) still spoken there today (Browne 2002). The modern Nubian languages are spoken not only on the Nile but also in the west and south of Khartoum (North Sudan). The westernmost Birked Nubian language in Darfur is nearly extinct (Bell 2006) but the southernmost Hill Nubian languages are still spoken in the Nuba mountains. The Harâza Nubian variety provides the missing link geographically between the Nile and the Nuba mountains but succumbed to Arabic before it could be adequately recorded (Mohamed 1974; Bell 1975a). The distribution of Nubian language has thus prompted the question whether they originated in the Nile region and then migrated south and west, or vice versa (Bechhaus-Gerst 1985).

1.2.1.2. Northwest Africa

West of the Egyptian language area, languages of the Berber subfamily are found in the habitable mountains, oases, and coastal areas all the way to the Atlantic coast. Some have given way to Arabic already, while others remain with sizable speaker communities, although not without influence from Arabic. In some cases, the Arabic contact has penetrated far into the Berber basic vocabulary, while the local Arabic acquired Berber phonology (Kossmann 2013), to the effect that, anecdotally, one cannot tell if Berber or Arabic is the language spoken in a market conversation. Some Berber varieties have become famous for allowing vowel-less
utterances, such as the tongue-twisting *tfktstt tfktstt* in Tachelhit of Taroudant, meaning ‘You sprayed it (fem.) and then you gave it (fem.)’ (Dell and Elmedlaoui 1988).

Despite the great distances from the Siwa Oasis in the east to Mauritania in the west, the Berber languages are closely related, roughly comparable to the Germanic subgroup of Indo-European. Yet the Berber languages show little in the way of neatly separable subbranches. The Tuareg languages, spoken by nomads of sub-Saharan ethnic descent, show unique innovations (Kossmann 1999: 31) as do Zenati and the newly discovered Tetserret forming the Western Berber branch (Lux 2011) and Libyan-Egyptian Oases Berber in the east (Souag 2014: 17–26). Yet the remaining core varieties show chains of innovations characteristic of a very long dialect chain breakup or of long-term language contact, though see, e.g., Souag (2014: 17–26), for some further suggested subgroups. A curious case is the Guanche language spoken in the Canary Islands up until the 18th century. The body of data consists of a few short amateur wordlists and another batch of individual words that can be extracted from travelers’ records. The data show major similarities and major differences from known Berber languages on the mainland. Though there is a compilation of all that is known about Guanche (Wölfel 1965), it remains to be investigated with a modern understanding of Berber subgrouping whether the Guanche language is a separate Berber branch, belongs to a specific Berber branch, is a non-Berber language with a Berber stratum of some specific Berber language, or is a non-Berber language with a miscellaneous Berber stratum. Another enigmatic language associated with Berber is the language of the Numidians in present-day Libya, mentioned in both Roman and Egyptian sources. Only personal names and a distinctive script have been passed down to us (Rössler 1979). The script is either a highly altered continuation of Phoenician-Punic or a separate branch of the Semitic family of scripts (Pichler 2007). While variants of this script were used along the Mediterranean coast in antiquity, the only modern survival is Tifinagh, sometimes used to write Tuareg languages, chiefly in Algeria. The value of Tifinagh as a source of ethnic pride is greater than its everyday use, and has led to the adoption of it in modern times as an alphabetic script, “Neo-Tifinag”, to write other Berber languages. The oldest substantial written records that can be securely identified as being of Berber provenance date back to the medieval period; they are written in an Arabic script (Brugnatelli 2011).

1.2.2. The greater Saharan area

1.2.2.1. Northeast Bilād as-Sūdān

Trans-Saharan trade networks were in full swing by the time of the Arabic expansion southwards. One of the several short-lived kingdoms formed around hub market towns was the Funj empire between 1504 and 1821 in Sennār, in north-
eastern Sudan (O’Fahey and Spaulding 1974). Curiously, it is not known what language(s) was current in the empire. However, a listing of 10 numerals and a poem written in Arabic script are attributed to the Funj language by the 17th-century Turkish traveler Evliya Çelebi (Spaulding [1672] 1973). The numerals are undoubtedly Kanuri, a major language of the Lake Chad area that could easily have been encountered at this time from itinerant individuals in northeastern Africa, but it could hardly have been the language of the Funj empire. The poem, on the other hand, cannot be read as coherent Kanuri (p.c. Dimitry Bondarev 2012), and its identity has yet to be ascertained. While the poem may hold the answer to the question of the language of the Funj, there are also good reasons to doubt the accuracy of the data. Evliya Çelebi probably never made it personally as far as the Funj empire (Peacock 2012), and the Turkish translation for the poem may not, or only loosely, correspond to the text in the unknown language, since there are recurrent morphemes in the source text that seemingly have no counterparts in the translation.

The slave trade and associated slave raiding in the centuries before European colonization must have severely affected the indigenous populations of the North Sudan–Ethiopia borderland. But the languages that survived the turbulent times bear witness to considerable linguistic diversity in this region. The Eastern Jebel group, the Berta dialect cluster, the Koman group, the Gumuz dialect cluster, the Mao group, the Nara language and the Kunama language are still spoken while the Gule language gave way to Arabic earlier this century (Delmet 1980: 7–8). These languages have been relatively neglected in both description and appreciation, though this is changing thanks to recent descriptive work. They are all typically subsumed under the Nilo-Saharan umbrella in one constellation or the other (except the Mao group, which is typically filed as Omotic since Bender 1975, see Gülde-mann, this volume, Chapter 2). Even if some version of this hypothesis should ultimately prove to be correct, the grammar and vocabularies of each lineage show large differences, so that any common ancestor must have been very far back in the past. The peoples of the Ethio-Sudan borderland traditionally had agro-pastoral economies supplemented by hunting and gathering (Cerulli 1956: 17–19) and formed small political units, rarely larger than the village (Cerulli 1956: 25–26).

It is no coincidence that languages in the area of the Saharan desert are found around mountains. With mountains there are rivers with water, which facilitate human subsistence conditions. A veritable mountain of tongues is found one day’s bus ride south of Khartoum. The Nuba mountains are home to some 50 languages from up to nine different lineages. Grammar sketches for one language from each group are found in Stevenson (1957) but for most languages only wordlists are available. The northernmost group is the Hill Nubian languages, which have already been mentioned above. Next, in the center, we have two large dialect clusters, Heiban and Talodi, spoken in a range of localities with populations on the order of 5,000. The Katla-Tima group consists of two clearly distinct but related
branches, Katla-Julud and Tima (Alamin Mubarak 2012). In the Tegali-Tagoi group, two dialect clusters with a total of about 50,000 speakers are spoken in the northeast, in the Rashad hills. Ama, with 40,000 speakers in the mid-West Nuba Mountains is related to Afitti (some 4,000 speakers), spoken in a more northern location far away enough that the two are not aware of each other. The Temein dialect cluster is also spoken in the midwest Nuba mountains and totals perhaps 15,000 speakers (Blench 2013). The southwest, around the urban center Kadugli, is dominated by the Kadu dialect cluster, whose main language Krongo (20,000 speakers) is described with a full-length grammar (Reh 1985). West of the Kadu group we find two incoming languages of the Daju family (see below). In the far south, the language Lafofa is spoken in two distinct varieties (Amira and Eliri) by perhaps 600 people. The Heiban, Talodi, Rashad and Lafofa groups have nominal class systems with alliterative concord, and have therefore (along with Katla-Tima, on some lexical resemblances) long been argued to be genealogically related to the geographically separated Niger-Congo languages (see Chapter 2 of the present volume).

The peoples of the Nuba mountains share a number of cultural features (Nadel 1947), such as having birth-order names, and many have pondered the reasons for the linguistic diversity. Classically, Thelwall and Schadeberg (1983) interpret the linguistic situations as layers of refuge, where the oldest layers are represented by deep genealogical relatives within the mountains and languages on the fringe, with relatives outside, are more recent layers. In contrast, based on data from Tima, Dimmendaal (2009a) speculates that the key to the diversity lies in a conscious ideology to keep and accentuate linguistic identity. Indeed, one curiosity that seems to challenge the received views on what deliberate language change can access is the language spoken in the village of Laro. Laro is 90% lexicostatistically cognate with the larger language Heiban but shares no form – meaning noun-class pairings with Heiban or the rest of the languages of its subfamily (Schadeberg 1981b: 147–149)! The Laro story is that the noun class prefixes were obfuscated deliberately to confuse their neighbors. As Schadeberg (1981a) concludes, there appears to be no convincing alternative to this story.

1.2.2.2. Sahara and Sahel

Moving west into Darfur we find the Fur language, which has around half a million speakers and is traditionally the lingua franca of the Fur area. Fur has a relative (Greenberg 1972) in the Amdang language, also known as Mimi or Jungraithmayr’s Mimi, spoken by some 40,000 people in the Biltine region across the border in Chad (Wolf 2010). Also straddling the Chad–Sudan border is the Tama dialect cluster with 100,000–150,000 speakers in total. The Tama verb undergoes extensive morphophonological changes, leading earlier researchers to remark that “no two verbs in Tama appear to be conjugated alike” (Dimmendaal 2009b: 315–
The Daju languages have a wider east–west extension stretching from Dar Sila in Chad to the Nuba mountains in Sudan. The total number of speakers of Daju languages is around 100,000 with two poorly known southern varieties in Bahr-el-Ghazal presumed extinct. The Daju family owes its spread to the Daju empire, mentioned by many Arab geographers in the 10th–13th centuries (Thelwall 1981: 162–174).

The handful of Maban languages dominate the border area more to the south, including the very inaccessible areas around the border of Chad Sudan–Central African Republic. The most important languages are Masalit (mainly Sudan) and Maba (mainly Chad), with around 300,000 speakers each. Data on two mysterious languages in the Maban region, both called Mimi, were collected about a century ago but neither has been heard of since and the languages are thus presumed extinct. Their genealogical relation(s), to each other and the Maban family, remains somewhat enigmatic (see Chapter 2).

Perhaps even more enigmatic is the Kujarge language first reported by Doornbos and Bender (1983: 59–60) with a 100-word list. Under challenging circumstances, Doornbos collected 200 words in total on two different occasions in 1981 from a father and a son, near the border town of Foro Boranga. The informants reported that the language is spoken in seven villages in Chad, near Jebel Mirra (11º 45' N, 22º 15' E) and scattered among the Fur and Sinyar in the lower Wadi Azum valley. In 1981, the Kujarge were a hunting and gathering people estimated to number 1,000. The uncertain classification of this language (see Güldemann Chapter 2), the inaccessibility of the area today, the rarity of a predominantly foraging subsistence pattern in the region and the fact that no other researcher is on record to have encountered the Kujarge ethnic group contributes to the mystery.

A couple of dozen East Chadic languages are found across central and eastern Chad. Speakers of East Chadic and Sara-Bongo-Bagirmi languages in the mountain massifs have a common identity as “Hadjerai” (cliff dwellers). The Hadjerai have an agropastoral subsistence pattern and a sociopolitical organization around clans and age groups (Fuchs 1979: 217–221).

Dominating the central Saharan area we find the aptly named Saharan family, consisting of a western branch and an eastern branch, as envisaged by Nachtigal.

---

2 Yet another Mimi is sometimes posited on the basis of listings ultimately emanating from van Bulck’s (1954) survey of southern Chad. No linguistic data (if any were collected) has surfaced but since van Bulck associates this Mimi with the Biltine region, it seems safe to assume it refers to the same Mimi as Jungraitmayr’s Mimi, also known as Amdang.

3 The name Kujarge is a local designation for ‘sorcerers’, and this name also occurs (MacMichael 1918: 45; Lebeuf 1959: 116) referring to other groups than the people whose language Doornbos recorded.

4 Not to be confused with the more famous Jebel Marra, on the Sudan side.
already in 1881. The eastern branch consists of Zaghawa (over 150,000 speakers in Chad and Sudan; Jakobi and Crass 2004) and the extinct Berti (Petráček 1966). The western branch consists of the Teda-Daza branch (ca. 500,000 speakers) and a Kanuri-Kanembu branch, a dialect cluster totaling up to 4 million speakers covering a large area around Lake Chad. Kanuri was the language of the Bornu empire (14–19th centuries) and thus an important language politically. Kanuri was written (mainly in religious or legal contexts) using the Arabic script, in an adaptation that rendered Kanuri tones. Kanuri was recorded early with a wordlist from the 17th century in the works of the Turkish traveler Evliya Çelebi (Habraszewski 1967). A recent discovery to Western scholars is that Tarjumo, an old form of Kanuri-Kanembu not intelligible to any variety spoken today, has been used for religious analysis and passed on as a literary language until today (Bondarev 2005).

Many millennia earlier than the Arabic expansion, the Sahara was more fertile and likely hosted human populations in places where there are none today (Drake et al. 2011). With the gradual drying up of the Sahara, many populations must have moved or perished altogether. Rivers completely dried up, leaving wadis, the local Arabic term for an empty riverbed. The Wadi Howar represents a (once) major tributary to the Nile, west of the White Nile, which may have left consequences traceable in modern languages (Dimmendaal 2007). One may speculate about further remnants of a once-greener Sahara in ethnographically marginal groups. The Daouada are a little known ethnically distinct group in South Libya (Pauphilet 1953) but the few reported encounters with them have revealed no other language than the local variety of Arabic. Groups who subsist mainly on hunting and gathering in the Saharan area are restricted to the Nô-êy (until recently, Matthey 1966), Kujarge (see above), Haddad-Cherek (Tedaga-speaking) and Haddad-Nichab (Kanembu-speaking). The latter had a hunting strategy similar to those known from prehistoric rock engravings from the Sahara (Nicolaisen 2010). However, not even in the earliest descriptions of the Haddad (Nachtigal 1881: 258–264, 330–331) is there any trace of a separate language.

The language family Songhay owes its distribution to the Islamic empire centered at Gao in the beginning of the 11th century. It is hence a rather shallow language family often figuring in discussions of the potentials of language contact, in several ways. Firstly, there is the idea that the core lexical stock of Songhay is cognate with another African language family (“Nilosaharan” according to Ehret 2001, or an otherwise unattested branch of Afroasiatic according to Nicolaï 2003) and its grammar somehow taken over from neighboring Mande language(s). Less radical suggestions merely argue that the typological similarity between Songhay and Mande languages may be the result of long-term interaction between the two. Secondly, the Northern Songhay languages are so influenced by neighbor-

---

5 Disguised in sheepskins and masks, the hunters sneak up on the animals and kill them with bows and poisoned arrows (Nicolaisen 2010: 24).
ing Berber languages that entire Berber morphological subsystems operate within Northern Songhay (Christiansen-Bolli 2010; Souag 2010). Various stages of the breakup of the Songhay family can be dated thanks to loanwords from Arabic (Souag 2012).

The Dogon languages form a small family in the mountainous area around the city of Bandiagara in Mali. The Dogon were studied relatively intensively by anthropologists and much has been written on their religion and cultural traits, such as mask dances and sculptures. A secret language, Sigi, was used by a male section of society for ritual purposes (Leiris 1948). The languages remained relatively neglected until the 21st century when a systematic documentation project headed by Jeffrey Heath commenced. More than a dozen different Dogon languages have to be recognized on grounds of mutual (un)intelligibility. Furthermore, the Bangi Me people live in the same mountain massif and are culturally similar to the Dogon. They speak a language traditionally filed as a Dogon “dialect”, which, on closer inspection, turns out to have a core vocabulary completely different from Dogon as well as any other nearby language family (Blench 2007a). Unless there are similar cases still “lurking” (cf. Siamou, Chapter 2 of the present volume), Bangime represents the only fully surviving language isolate in West Africa.

1.2.3. Ethiopic sphere

The Ethiopian Highlands provide a large habitable area in the Horn region of Africa. The bulk of the area lies above 1,500 meters and is thus relatively cool and free from mosquitoes. A number of crops originate in the Ethiopian Highlands, such as coffee and t’ef. T’ef is a small-grained cereal that has been cultivated for more than 2,000 years and is a staple food in the region but unlike coffee, it is difficult to grow in other ecoclimatic zones. The Highlands are dominated by a set of languages forming the Ethiopic branch of the Semitic (sub-)family that must have entered from the Arabian peninsula sometime in the first millennium BC (Gragg 2008: 211). The earliest testimonies are pagan epigraphic inscriptions, but later materials include a significant body of Christian literature, written in Ge’ez, the language of the Aksum empire. In some ways the Ethiopian Highlands resemble European medieval states. Around 350 AD Christianity was made the state religion, which continues up until today as the Ethiopian Orthodox Tewahedo Church. The empire had a strategic location for trade between Egypt and the Indian Ocean, and had diplomatic contacts with the East Roman empire. After a “dark ages” the Ethiopian empire gradually lost its connections with the civilizations to the north. Beginning in 1270, there was a long dynasty of kings culminating in Haile Selassie’s dethronement in 1974. Ge’ez, extinct since a millennium as a spoken language, holds a position as a liturgical language and is used and studied in some ways similar to Latin in Europe. Ge’ez was also studied by European scholars as early as the 16th century. A famous anecdote recounts how Job Ludolf, a 17th-cen-
A survey of African languages

21

tury German religious scholar who had learned Ge’ez from books, met with Abba Gregorius, an Ethiopian convert to Catholicism, in Rome. Gregorius spoke Amharic as his everyday language\(^6\) and did not know Latin or any contemporary European language. Thus they were obliged to converse in Ge’ez, the only language in common, which neither party had used as a spoken language (Pankhurst 1965: 56–66). While Ge’ez is a verb-initial language and similar in typology to the other ancient Semitic languages (Lambdin 1978), the modern Ethiosemitic languages have undergone a complete typological makeover, presumably due to contact with or shift from languages in situ, to being verb-final with associated characteristics (Crass and Meyer 2011). The Ethiopian civilization had its own grammatical tradition, where a few other languages in addition to Amharic were described (Mulugetta 2010). Ge’ez, Amharic, and further languages of Ethiopia are written with fidel, a distinctive script of the South Semitic family of scripts (Bender et al. 1976).

Apart from Ethiosemitic, the other widespread (sub-)family in the Ethiopian highlands is Cushitic. The Northern Cushitic Branch is represented only by the Muslim pastoralist Beja people in present-day Sudan and Eritrea. Four endangered languages in the highlands make up the Central Cushitic branch (Appleyard 2006). One of them, Qwara (aka Qimant) (properly part of the Kemant dialect cluster), is the traditional language of the Falasha, who have received some notoriety as a group of Ethiopians who have long self-identified as Jews (Appleyard 1998).

The East Cushitic languages are found both in the Highlands and the Lowlands, spoken largely by terrace-farming communities. The largest language is Oromo, which is also the largest language in Ethiopia in terms of L1 speakers and also stretches down into Kenya. It has a Roman-script written standard that has some use in schools, administration and literature (Griefenow-Mewis 2001). A particular age-grading system known as Gada is found in a number of East Cushitic cultures. Jensen (1954) describes the system for the Konso. The entire population is divided into the two classes Galgússa and Hirba, whereby membership is passed on patrilineally. Within each division there are four age-grades: (1) Fareita, (2) Chela, (3) Gada and (4) Orschada. Every 18 years a celebration is held where members are promoted one step up, retaining a fixed gap of two grades between father and sons. Thus, if a man is Gada (3), all his sons are Fareita (1), until the man is promoted to Orschada (4), when his sons step up to Chela (2). The age-grade and division regulate all aspects of Konso traditional life, i.e., labor tasks, ceremonial functions, marriage order, dancing conventions, and so on. Another strong East Cushitic language is Somali, spoken in Somalia and adjacent countries, by muslims of a traditionally pastoralist economy. Somali has many “dia-

\(^6\) The first wordlist of Amharic, collected in Jerusalem, dates from the 15th century (Cohen 1931; Muth 2010).
lects” that may not all be inherently intelligible (Tosco 2012). Somali has a Roman written standard as per a decision in 1972 to switch from an Arabic-based script.

The area southwest of the Ethiopian Highlands is drained by the Omo River. The language density here is among the highest anywhere on the continent (Pauli 1950), and the area also harbors considerable cultural and genealogical diversity. The (sub-)families represented are Surmic, Maji aka Dizoid, Nilotic, East Cushitic, Ta-Ne Omotic, Ari-Banna, and the isolates Shabo and Ongota. There are also scraps of data on a further (vanished or spurious?) unclassified group, named Dima by Bottegò (Rossini 1927). Most Surmic and Nilotic groups as well as the East Cushitic Dhaasanac are predominantly pastoralists, and with a few exceptions (see below) the remaining groups are predominantly agricultural. Many of the ethnic groups in the Omo Valley maintain their traditional lifestyle and appearance. Because of the elaborate bodily decorations, including lip discs, many collections of photos have been published in popular and/or art-oriented press (e. g., Sullivan 2012).

A number of languages in Southwest Ethiopia are highly tonal, culminating in Bench, a language with five distinct tone levels, which can be whistled as well as played (!) on a five-stringed guitar at least to the degree sufficient to describe the location of an object (Wedekind 1983: 148–151).

Agriculture/pastoralism is now the norm all over the Ethiopian sphere, but pockets of hunting and gathering subsistence modes survive, likely reflecting an earlier more widespread presence. The Shabo live embedded as hunters and gatherers in Majang (Surmic) society (Stauder 1971) and speak a language with no known relatives. The Ongota are an ethnic group of 83 individuals who now live in close association with the Ts’amakko (East Cushitic) and have a similar agropastoral culture. Some earlier testimonies describe the group as cattle-less, and predominantly hunting, but detailed testimonies of a foraging subsistence mode is lacking (Savà and Thubauville 2010). The language, still known by fewer than a dozen older members of the community, is unrelated to its neighbors, save for lexical loans. Boon is a name used for remnant (former) hunting and gathering communities in the Somali sphere. Apart from varieties of Somali with special vocabulary, one highly endangered Boon language has been briefly recorded. As far as can be told it represents an East Cushitic language (Lamberti 1986). The Weyto live at Lake Tana and were hippopotamus hunters until the last century (Gamst 1978). The early Scottish traveler James Bruce, who spoke Amharic, passed through the area around 1770 and reported that “the Wayto speak a language radically different from any of those in Abyssinia” (Bruce 1790: III: 403), but by the time Marcel Griaule visited in 1928 they had already switched to Amharic (Darmon 2010). A number of Surmic groups in the Omo Valley have hunter-gatherer castes (Cerulli 1929), but whenever their language variety is known, it is on par with other Surmic lects (Haberland 1966; Hieda 1990). The Mangio form a caste among the Kafa (Ta/Ne-Omotic) associated with hunting, and now speak Kafa, but some special vocabulary relating to their speciality has been recorded (Cerulli 1951). Further
castes are known among other highland Ethiopian peoples, not only as hunters but also as potters or tanners. Likely, with diminishing possibilities for hunting and gathering, such peoples were driven to become specialists in labors pertaining originally to a foreign culture (Freeman and Pankhurst 2003).

1.2.4. Sudanic West Africa

Sahelian West Africa was home to several medieval empires that flourished on trade, typically in gold and salt. Many of them are described by Arab travelers (such as Ibn Batūta; Hamdun and King 1994 [1353]) as prosperous, but their success is most clearly reflected today in the continued existence of various widely spoken languages. The country of Mali owes its name to the Mali empire, a Muslim kingdom that at its peak stretched from the coast of Senegal to Timbuktu (Niane 1975). The Manding subfamily owes its spread to this empire, with the three largest varieties Bambara (aka Bamanankan), Mandinka and Jula accounting for some eight million L1 speakers and twice as many L2 speakers in present-day Mali, Ivory Coast, Senegal and The Gambia. The close-knit Manding group languages form part of a large and deep family called Mande, which stretches as far east as northwestern Nigeria, with Busa spoken on the Niger River. The northwestern Mande language Bozo spoken in Kelinga (a variety of Hanxyaxo Bozo [bzx]) is noteworthy as a language with an unusual noun-verb derivational relationship. Throughout the lexicon, referential use of a lexeme requires one tonal pattern while predicative use requires an inverted tonal pattern (low becomes high, high becomes low, falling becomes rising and rising becomes falling). In contrast to languages with overt derivational morphemes, no natural direction of derivation suggests itself, i.e., Kelengaxo Bozo does not specify whether verbs are derived from nouns or vice versa (Ebermann 2005).

In westernmost West Africa, the languages that are not part of the Mande family all have concordial noun class systems but do not otherwise form a coherent group. A cover term for the five coordinate genealogical groups is Atlantic. The biggest, North-Central Atlantic, encompasses the bulk of indigenous languages of Senegal, The Gambia and the Guineas. The most important language is Wolof, with two mutually intelligible written standards in Senegal (with francophone loans) and The Gambia (with anglophone loans).

Although pre-Arabic and pre-Tifinagh writing is attested in the Sahara (Monod 1993), the popularity of writing was conditioned by the incursion of Arabic and Islam. The practice of writing an indigenous African language using (possibly a modified version of) the Arabic script is now known as Ajami (Mumin and Versteegh 2014). Large collections of writing of various genres exist in large West African languages such as Fulani or Bambara, but further manuscripts exist with writing in smaller, e.g., Mande, languages that have only just begun to be studied philologically (Ogorodnikova 2014).
Most West African traditional societies subsist mainly on agriculture, aided whenever possible by herding, hunting and gathering. But this is not universally the case. A number of predominantly fishing populations exist along the Niger River (the Bozo, Ligers 1969), in coastal areas (e.g., Lebou, Mercier and Balandier 1952 or Imraguen, Athonioz 1967), as well as in the lagoon area of Ivory Coast (e.g., Avikam, Zwernemann 1979). The Némadi of Mauretania is the last hunting and gathering (sub)group remaining in West Africa, but if the Némadi had a remnant language in the past, there is little trace of it today. Endogamous castes specializing in occupations such as metalwork, bardship, leatherwork, woodcarving and weaving are found throughout the area (Tamari 1988).

The Fulani are a predominantly herding people who are found in a more or less continuous belt from Senegal to western Central African Republic. Naturally, the Fulani move with their herds, typically over territories seasonally occupied by other, sedentary, ethnic groups with whom the Fula have a symbiotic and sometimes hostile relationship. All Fulani languages are closely related, betraying a very fast expansion over a quite enormous territory. The Fula languages are most closely related to Serer in Senegal, which implies a far western origin of the Fulani. The Fula languages have full-fledged noun class systems with alliterative concord. In various areas of West Africa, the local Fula variety is an important language, spoken also by other ethnic groups, and in Cameroon, a simplified Fula is attested (Lacroix 1959) that has lost the noun classes.

Nearly all West African languages between Liberia and Nigeria belong to the Mande (roughly Northwest), Kru (roughly West), Kwa (roughly East) and Gur (roughly North) subfamilies. The most important Kru language is Grebo, spoken with many regional variants in Liberia. The Cape Palmas variety was studied relatively early by missionaries (Auer 1870).

Liberia has seen the development of no less than five different scripts, all developed by visionary native speakers of Vai, Mende, Loma, Kpelle (all Mande family) and Bassa (Kru family) respectively. The oldest one, Vai, was developed in the 1830s while the others were devised in the 1920s or slightly later. Only the Vai script is still in use today. The scripts are not variants of each other or of European/Arabic scripts. Only the Bassa script is alphabetical and records tone, while the others are syllabic scripts. Mende was written from right to left, while the others left to right (Dalby 1967).

Most of the many dozens of Gur languages are spoken in or around Burkina Faso. The largest Gur language is Môôré, with around 5 million speakers in and around the capital Ouagadougou. Kwa is also a large group with many smaller languages, but has a few larger coastal languages (see below). Some Kwa (as well as some Mande) languages lack phonemic nasal consonants, as all phonetic nasals can be felicitously analyzed as nasal allophones conditioned by a nasal vowel (Saout 1973; Bolé-Richard 1984). There are surprisingly few traces in West Africa of what would have been the languages before the expansion of the Niger-Congo
and Mande families. Ega, Mpra and Pere are languages that have been difficult to place in the larger subgroups but are nevertheless arguably part of the Niger-Congo family (see Chapter 2 of the present volume).

A number of large West African languages owe their expanse to pre-Colonial empires. The Akan (the Ashanti Kingdom), Ewe in Ghana and Togo (the Dahomey empire), Yoruba in southwestern Nigeria (the Oyo empire) and Edo/Bini in southwestern Nigeria (the Benin empire) were coastal states that profited from trade with the Europeans after 1500, first with the Portuguese and subsequently with the British. In the 1500s vocabularies of coastal West African languages began to appear in travelogues and the like (Cole 1971). The transatlantic slave trade brought slave raids into the interior of West Africa. For many languages in the interior, the first published data were collected from slaves ex situ. Notably Koelle’s (1854) compilation of wordlists from 180 languages encountered among freed slaves in Liberia is remarkable for its accuracy and the detail paid to phonetic transcription and geographical provenance. Furthermore, it is organized according to genealogical subgroups with insight far ahead of its time.

Yoruba is the paramount language of the region around Lagos, the economic capital of Nigeria. Yoruba has over 20 million L1 speakers and several million L2 speakers. It has a Roman script orthography that was created in 1850 and adjusted to its current form in 1966 (Bamgbose 1965). Yoruba is used in all domains, including newspapers and movies. In the slave trade diaspora some Yoruba presence survives in the new world. Lucumí, a language spoken in Cuba by practitioners of the religion known as Santería, is a variety of Yoruba (Olmsted 1953; Concordia 2012).

Akan, dominant in the southern half of Ghana, has over 10 million L1 speakers and another few million L2 speakers. Three mutually intelligible dialects have been developed as literary standards with distinct orthographies: Asante, Akuapem (together called Twi), and Fante. The Akan convention of naming children after the day of the week on which they are born has led to the proliferation of certain names, e.g., Kofi is the given name for a male born on a Friday (Dolphyne 1988).

Ewe, with over three million speakers, is an official language of Togo and Ghana. It is part of the larger Gbe cluster, which dominates the coastal areas between Ghana and Nigeria (Manoukian 1952).

Further east along the Nigerian coast and hinterland we find the smaller (sub) families of Edoid, Ijoid, Igboi and Cross River languages (Williamson 1971). The principal means of subsistence for these people is the cultivation of yams and cassava (Forde and Jones 1950; Bradbury 1957).
1.2.5. Central Sudan and Cameroon Grassland

1.2.5.1. Central Sudan

The area between the Niger and Logone Rivers is a tableland savanna broken up by a number of mountain massifs, the Mandara Mountains, the mountainous Adamawa Highlands and the Nigerian Plateau. The language diversity in these highland and mountainous areas is staggering. Most of the area above the Niger River in Nigeria is covered by different subgroups of Niger-Congo, the large Kainji subgroup, in the western part, near the lake of the same name, mid-south is Nupe and a handful of closely related languages forming the Nupoid subgroup, the large Plateau subgroup in mid-central Nigeria and the Jukunoid subgroup in the mideast. There are a couple of in-depth descriptions of individual Kainji (McGill 2009) and Plateau (Bouquiaux 1970) languages, but overall, these groups belong to the least-studied subfamilies in all of Africa (Sands in press). Most of these languages have noun classes with alliterative concord and the remainder are thought to have lost such a system. The merit of studying these languages is effectively illustrated by the discovery of a phonemic “explosive bilabial nasal”, provisionally transcribed as [ŋ̊ʘʷ] (Harley 2012: 59–62) in Ninkyob, a sound not known to occur in any other language.

Jukun, of the Jukunoid subgroup, is a relatively widely spoken language, owing to the (non-Islamic) Kororofa empire along the Benue River in the centuries after 1500.

To the north and northeast, around the mountainous border with Cameroon, we find a large array of languages from the Adamawa pool of Niger-Congo and the Chadic subgroup of Afroasiatic. In the west, predominantly Nigeria, the languages belong to the West Chadic subgroup. Although a subgroup, its internal depth is frequently compared with Indo-European as a whole (Schuh 2003). The Central or Biu-Mandara subgroup is much less deep. Hausa is the paramount Chadic language, whose triumph started with the Hausa empire in the middle ages, and now counts some 35 million native speakers and another 15 million L2 speakers. Hausa was written in an Arabic-derived script but has now adopted a Roman-based orthography. Hausa is dominant over all of northern Nigeria and adjacent regions in neighboring countries, and is the target of language shift for many of the innumerable small languages in the area, be they Niger-Congo or Chadic, which are increasingly becoming endangered as a result (Blench 2007b). Hausa spoken as an L2 language with differences from standard Hausa has been recorded by linguists (Feyer 1947; Hodge 1960) but may or may not represent stable pidgins.

Nearly all the peoples in the central Sudanic area are farmers of various kinds and intensities. Although, again, along the major rivers, the Niger and the Logone, as well as the Buduma around Lake Chad, we find predominantly fishing populations. Laal is a fishing-farming language on the Chari River discovered only
in the 1970s (Boyeldieu 1979) with no demonstrable relatives (Lionnet 2010). Sorghum and millet are the principal crops in the north, while yams become more important the further south one moves. A common kind of hereditary professional specialization was (black)smiting, and endogamous smith clans are found across the area. Occasionally, such groups have a language of their own, unintelligible to outsiders, e.g., Kawaway in southern Chad (Lionnet and Hoinathy 2015), Uneme in southwestern Nigeria (Bradbury 1957: 123–129) or Kpeego in southwestern Burkina Faso (Zwernemann 1996).

Islam is the most common religion across the central Sudanic area, though scattered non-islamic areas exist, called Maguzawa in the Hausa sphere and Kirdi in the Mandara sphere.

1.2.5.2. Cameroon Grassfields

A grassy high plateau aptly labeled the Grassfields provides home to a large number of languages belonging to the various constellations of the Bantoid subfamily. The area was previously covered by forest but was tamed to its current state for greater agricultural productivity (Chilver and Kaberry 1967). The linguistic diversity in the hilly areas of Cameroon is certainly impressive, so that several authors have suspected deliberate language change to play a role, which is consistent with data from ethnography and oral history (Dicarlo 2011).

An indigenous script for the Bamun language was developed, starting in 1896 by a royal decree from King Njoya. The script evolved from a pictographic type to a partially alphabetic syllabic script in the space of 14 years, from 1896 to 1910. The latter version contained 73 syllable characters plus 10 signs for numerals, and tonal indications were used to differentiate between otherwise homonymous words. After he had invented his script, Njoya opened a school and taught a large number of people to write. Scribes were installed in the courts and cases were recorded, but already around 1931 the script fell into disuse. King Njoya attributes the idea of devising a script to a dream, but the idea of writing was not unknown, as Hausa merchants had brought books with Arabic script to the area around the mid-19th century, and German and English books somewhat later (Dugast and Jeffreys 1950).

1.2.5.3. The Central African area

Most of southern Chad is inhabited by peoples speaking languages of the Sara-Bongo-Bagirmi subgroup. Bagirmi is the language of the Kingdom of Bagirmi (1522–1897; Lebeuf 1978), and while the number of L1 speakers does not exceed 100,000, many speakers of smaller languages of the area speak Bagirmi as L2. A few Sara-Bongo-Bagirmi languages are also found across the border in South Sudan and in Central African Republic (CAR). One Sara language, Furu, is found
at the very southern border of CAR, at a considerable distance from its congeners (Boyeldieu 1990).

The western part of CAR is dominated by the close-knit Gbaya subfamily of Niger-Congo, which also extends into Cameroon. La’bi, an initiation language learned by men among many groups of Gbaya, is documented. The lexicon of Gbaya consists of predominantly non-Gbaya lexical items, but these often have similarities with, and probably derive from, non-Gbaya languages in the vicinity such as Sara or one of the Adamawa languages. The grammar, as far as is known, is identical to Gbaya (Tessmann 1931; Moñino 1977).

The eastern part of CAR is dominated by another close-knit subfamily, the Banda languages, which also extend down into Democratic Republic of Congo (DRC) and western South Sudan (Olson 1996). The southward spread of Banda is in part due to migration to escape slave traders in the 19th century (Maes 1984: 87).

The indigenous languages of the very remote western part of South Sudan are among the least studied in all of Africa. In fact, that which is known is largely due to one indefatigable individual. The missionary Stefano Santandrea wrote no fewer than 25 grammar sketches, along with lexical and ethnographic documentation of the little-known languages and peoples of southwest (South) Sudan. The area is quite diverse in terms of genealogical (sub)families. A handful of languages belong to the Sara-Bongo-Bagirmi group, another handful to the Mundu-Baka group, nine languages form the Sereic group and another few to the Zandeic group (cf. Chapter 2 of the present volume). The small Kresh-Aja group (Santandrea 1976), the highly endangered Birri language (Santandrea 1966) and the Moru-Ma’di group of languages are thought to be distantly related, forming a Central Sudanic family together with Sara-Bongo-Bagirmi and the Ngiti-Lendu, Mangbetu-Asua and Membi-Mangbutu-Efe subgroups of the northeastern DRC (see Chapter 2 of the present volume). The Moru-Ma’di languages extend into DRC and Uganda, and comprise several languages with a sizable speaker population, such as Ma’di (300,000 speakers) and Lugbara (over a million speakers).

All over the area around the CAR, South Sudan and DRC tricrossing border, the Zande language prevails with over a million speakers. The Zande have been made famous by the British anthropologist E. E. Evans-Pritchard (1937) for their traditional beliefs in witchcraft.

1.2.6. Equatorial rain forest

The great equatorial forest is shared between nine countries: Angola, Cameroon, Central African Republic, Democratic Republic of the Congo, Republic of the Congo, Burundi, Rwanda, Tanzania and Zambia. Surviving in the forest is not easy (Vansina 1990: 35–46) and it is inhabited today by humans who have adapted and learned to live in it. Today the Bantu languages occupy most of the forest, but along the northern edge we find the small Ngbandi group and the Mundu-Baka
languages; inside the Bantu area in northern DRC, we find the small group of Mbaic languages; and finally, in the Ituri forest area of far northeast DRC we find three coordinate groups of Central Sudanic languages, the Ngiti-Lendu, the Mangbetu-Asua and the Membi-Mangbutu-Efe.

We know that the Bantu are relatively recent arrivals into the rain forest area. First, the Bantu carry with them a farming culture and the archeological record shows such a distinction entering from the northwest some 5,000–3,000 years ago (Phillipson 2005: 201–202, 245–269; cf. Bahuchet 1993: 104; Oslisly et al. 2013). Second, the earliest split-offs of Bantu, e.g., Bube on Fernando Poo island, as well as the closest outside relatives of the Bantu languages, are found successively outside the Bantu area to the northwest. It thus stands to reason that the Bantu entered from the northwest. Exactly when and how the forest area was settled is a much more difficult question to answer (Vansina 1990; Lupo et al. 2014). One approach is to model the expansion of Bantu languages as a tree where every internal node is located in time and space. Given some modeling assumptions, such a tree can be built on lexical data on Bantu languages using their present locations and a few temporal calibration points. Grollemund et al. (2015) find that the backbone expansion starts out going southeast to the mid-southern Congo, where subsequent diversification takes place to the east and south. This result is inconsistent with earlier suggestions that had emphasized migrations straight south along the coast and straight east along the northern border of Congo. Concerning the penetration of the forest area, a significant finding is that a majority of ancestral locations in the backbone migration coincide in time and space with paleoenvironmental data that indicate a temporary retraction of the Congo rainforest at its periphery and that the rate of migration slows down when moving from savanna into rainforest.

Like its congeners, the Bantu languages are known for their large concordial noun class systems whose existence has proven to be remarkably stable. Out of upwards of 500 Bantu languages only Mbati-Pande (Maho 1999: 137), the Bira group (Kutsch Lojenga 2003) and Nzadi (Crane et al. 2011) are known to have lost the noun class system (in the sense of not showing agreement in the noun or verb phrase).

The largest of the languages of the Congo rain forest area are Lingala, Luba-Kasai and (the many varieties of) Kongo. Lingala is spoken natively by some 15 million speakers and another 10 million second language speakers (Meeuwis 2013: 25). It has its origins in the Bangi language, which was used as a riverine trade language on the western part of the Congo River before the European occupation in the 1870s and 1880s. Europeans officers and workers recruited from various other parts of Africa acquired a simplified version of the Bangi trade language and expanded its sphere of influence, including the important state post Bangala (later Nouvelle-Anvers, now Mankanza). At Bangala the language underwent significant influence from local (closely related Bantu) languages such as
Boko, Mabaale, Libinza, Boloki and Lusengo and various missionary attempts to standardize or even set a prescriptive standard for its form. Given its association with the Bangala station it soon acquired the glossonym Bangala, and spread east, northeast and south as far as Leopoldville (now Kinshasa). The name Lingala was later introduced by missionaries on the analogy of the regionally common pattern with a ba- prefix for the ethnonym and li- for the glossonym. Luba-Kasai (or Western Luba) is spoken in the center of the DRC in an area extending along the Kasai River (Burssens 1946). Luba-Kasai has some 6.3 million L1 speakers and almost one million more L2 speakers. The present-day Kongo varieties reflect the dispersal of the kingdom of the same name at the mouth of the Congo River (de Schryver et al. 2015). Given its strategic location, the kingdom had early contacts with European powers and thus the oldest known dictionary of a Bantu language (de Gheel 1652) was compiled at the mission station of San Salvador. Simplified varieties of Kongo spread up the river and are now widely spoken as second languages on both sides of the river (Mufwene 1997). Under the bend of the Congo River we find the Mongo people with the language Lomongo (aka Mongo-Nkundu). Excepting languages with a dedicated academy, Lomongo is the most extensively described language on the planet, thanks to the work of especially Gustaaf Hulstaert. Hulstaert authored a grammatical description totaling almost 2,000 pages (Hulstaert 1966, 1988), a dictionary of over 2,000 pages (Hulstaert 1957, 1985) and an extensive dialectology (Hulstaert 1999).

Most of the people who inhabited the Central African rain forest are fisher-farmers, with a significant proportion of hunting. Curiously, some of the important crops such as the banana (from the east) and maize (from the New World) were introduced relatively late, such that it is possible to trace their spread through linguistic data (Bahuchet and Philippson 1998). But also, all across the forest area, interspersed with the farmers, we find forest specialists who exploit the forest’s resources by foraging. The foraging groups are always in a patron-client relationship trading forest produce with a farming community. Most of the forest specialists have significantly shorter stature (Bahuchet 1993: 89–90), and are commonly called Pygmies, while other groups appear to be mixed with the taller farming populations and are called Pygmoids. While size is just one physical attribute, the Pygmies are genetically extremely different from the farming neighbors (Verdu et al. 2009). The current Pygmy population is estimated to be at least (several) hundred thousand but could potentially be closer to a million judging from the environmental carrying capacity (Olivero et al. 2016).

One would expect with such genetic and ethnographic differences that the Pygmies would either have vanished as distinct ethnicities or that they would speak language(s) unrelated to, or at least genealogically far removed from, the their present farming neighbors. Yet the ethnolinguistic conundrum is that all attested Pygmy languages are rather closely related to a farming language in their vicinity (see Table 5). The inescapable conclusion is then that the Pygmies shifted
Table 5: Pygmy languages with family classification, relation to neighboring farming languages and data sources.

<table>
<thead>
<tr>
<th>Pygmy Ethnic Group</th>
<th>Country</th>
<th>Language</th>
<th>(Sub)family</th>
<th>Main Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gyeli</td>
<td>Cameroon</td>
<td>Borderline intelligible to the neighboring Kwaisio language</td>
<td>Bantu A80</td>
<td>Rénaud 1976</td>
</tr>
<tr>
<td>Baka-Gundi-Ganzi</td>
<td>Cameroon and CAR</td>
<td>Closely related to or even intelligible with the fishing-farming</td>
<td>Mundu-Baka</td>
<td>Ouzilleau 1911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>communities of Limassa and Ngundi neighboring to the east</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aka-Mikaya</td>
<td>Congo Republic and</td>
<td>A language (or dialect chain) unintelligible to all other groups</td>
<td>Bantu C10</td>
<td>Klieman 2003</td>
</tr>
<tr>
<td></td>
<td>CAR</td>
<td>Variety of the neighboring farming Bofi language</td>
<td>Gbaya</td>
<td>Fouts 2002</td>
</tr>
<tr>
<td>Bofi Pygmies</td>
<td>CAR</td>
<td>Varieties of the respective farming groups</td>
<td>Bantu B</td>
<td>Raponda Walker 1937; Mayer 1987;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Klieman 2003; Medjo Mvé 2011;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Larochette 1958; Démolin 1992;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Harvey 1997</td>
</tr>
<tr>
<td>Babongo</td>
<td>Gabon and Congo</td>
<td>Variety of the neighboring Mangbetu</td>
<td>Central Sudanic</td>
<td>Schebesta 1952; Vorbichler 1974;</td>
</tr>
<tr>
<td></td>
<td>Republic</td>
<td></td>
<td>(Mangbetu-Asua)</td>
<td>Schebesta 1952; Harvey 1997</td>
</tr>
<tr>
<td>Asua</td>
<td>DR Congo</td>
<td>Variety of the neighboring Lese</td>
<td>Central Sudanic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Membi-Mangbutu-Efe)</td>
<td></td>
</tr>
<tr>
<td>Efe</td>
<td>DR Congo</td>
<td>Variety of the neighboring Lese</td>
<td>Central Sudanic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Mangbutu-Efe)</td>
<td></td>
</tr>
<tr>
<td>Kango</td>
<td>DR Congo</td>
<td>Borderline intelligible to the (non-neighboring) Komo</td>
<td>Bantu B20</td>
<td>Schebesta 1952; Harvey 1997</td>
</tr>
<tr>
<td>Mbuti Pygmy Bila</td>
<td>DR Congo</td>
<td>Intelligible with neighboring Forest Bila</td>
<td>Bantu D30</td>
<td>Turnbull 1965; Harvey 1997</td>
</tr>
<tr>
<td>Rundi Kitwa</td>
<td>Rwanda-Uganda</td>
<td>Borderline intelligible to neighboring Rundi</td>
<td>Bantu D60</td>
<td>van der Burgt 1902</td>
</tr>
</tbody>
</table>
language, not once, but on multiple occasions. Only in the case of Bofi did the shift happen recently enough to make it into living memory (Fouts 2002: 44; this shift was from Aka to Bofi). In the case of Aka, the shift must have happened at least a millennium ago, as no other present population speaks (a language intelligible to) Aka, let alone an adjacent people. Particularly interesting is the case of Aka and Baka, two adjacent Pygmy populations, one speaking a Bantu language and one speaking a Mundu-Baka language. Bahuchet (1989) has shown that the two nevertheless share a significant portion of vocabulary relating specifically to forest foraging (such niche vocabulary may not even exist in farming languages) and attributes this to a period of ancient shared history.

1.2.7. Eastern savanna

The Bantu languages are also omnipresent in East Africa. The largest of all is Swahili, a Bantu language with a very large proportion of Arabic loans, betraying its origin in coastal settlements with Arab trade. In the last centuries, the coast-based trade intensified into the interior of Africa, and this way Swahili spread to become the lingua franca all the way from the coast to eastern Congo. For a briefer period, another Bantu language, Sukuma, was used as a lingua franca in the interior of Tanzania, before Swahili took over. The chief items of export trade from eastern Africa were ivory, horn and skins, as well as slaves. Beads, pottery, glass, cloth and other luxury manufactures were the principal imports.

Swahili has some 15 million L1 speakers and at least 50 million L2 speakers. It is the medium of instruction in countries like Tanzania and Kenya and spoken as an L2 by most of the inhabitants of these countries. In post-independence Tanzania it is the express policy to make the (non-European) Swahili the language unifying the nation (Blommaert 2014), for instance by enforcing Swahili in elementary school instruction, even in areas where the younger pupils do not yet understand Swahili. The success of Swahili as the national language naturally comes at the expense of the over 100 smaller, mostly Bantu, languages of rural Tanzania, intensifying their endangeredness. With the rapid spread of Swahili as a lingua franca in the last century, the high proportion of L2 speakers in DRC appears to have had some leveling effect on the Swahili spoken in the DRC (Mwamba Kapanga 1993). There are also pidgin Swahili varieties attested among settlers (Mutonya and Parsons 2004) as well as in the army (Vitale 1980). One African community of those brought to what is now India and Pakistan used a form of Swahili (Burton 1851: 372–374) in the mid-19th century but no longer do so (Lodhi 2008).

The largest other Bantu languages are Kinyarwanda-Rundi (totaling some 20 million L1 speakers) in the fertile highlands of Rwanda and Burundi and Kikuyu (over 6 million L1 speakers), the native language of the area around Kenya’s capital Nairobi.

Apart from the relatively late-arriving Bantu languages, other lineages are
found in East Africa, namely the South Cushitic subfamily, the Nilotic languages, and the language isolates Hadza and Sandawe. Hadza is spoken by a hunting and gathering community of less than a thousand individuals who live near Lake Eyasi and are famous among anthropologists for their disinterest in accumulating individual material wealth (Woodburn 1982). Many Hadza are monolingual and Hadza is still being transmitted to children, though there are now many Hadza who speak Swahili and/or other neighboring languages. Hadza is a language with phonemic clicks and its grammar has been studied, though we are still lacking an extensive grammatical description (Sands 2013). Sandawe, also in Tanzania, is a much larger language with some 40,000 speakers. It also has phonemic clicks and now has two extensive grammatical descriptions (Eaton 2010; Steeman 2012). The Sandawe are a farming community, though elements in their culture has led Newman (1970) to argue that they adopted farming only in the last centuries.

Four South Cushitic languages, Burunge, Alagwa, Iraqw and Gorowa, are still spoken in Tanzania by farming communities. Kw’adza, Aasax and Ma’a are another three languages often subsumed with South Cushitic. Kw’adza and Aasax went extinct before extensive data could be gathered (Kohl-Larsen 1943; Petrollino and Mous 2010). The Aasax were a hunting and gathering community submerged among the pastoral Maasai (Merker 1910: 229–269). The Ma’a language represents one of the world’s most curious cases of language mixing, which was finally clarified thanks to the fieldwork of Mous (2003). The ethnic group speak two different registers, one, which we may call Mbugu, is a pure Bantu language closely related to Pare (aka Asu), while the other, which we may call Ma’a, is identical to Mbugu in grammar but has a divergent core vocabulary, a large section of which has Cushitic parallels. Different scenarios have been proposed to account for this situation. The strongest case can be made for a Cushitic community who shifted to a Bantu language, but consciously decided to stop, or even reverse, the shift when it was already nearly complete (Mous 2003).

In Kenya three traditionally hunting and gathering communities, Yaaku, Elmolo and Dahalo, speak (or spoke until recently) Cushitic languages. In both the Yaaku (Carrier 2011) and Elmolo (Tosco 2015) communities the last fluent speakers have now passed away following a long process of cultural assimilation, but there are individuals in both communities who are actively engaged in reviving the languages. Although the traditional lifestyle is no longer practiced, Dahalo still has fluent speakers. Dahalo, uniquely for its family, contains about 40 words with phonemic clicks, presumably traces of contact with or shift from now vanished northernmost click languages (Tosco 1991).

The Nilotic languages are spoken over a wide, vertically oriented area, stretching from mid-Sudan through to northern Uganda, western Kenya and adjacent areas in Tanzania. The Nilotes have a distinctive physical type and a decidedly pastoral lifestyle. The importance of cattle can hardly be underestimated in the classic Nilotic ethnographies. Cattle is a source of food, wealth, marriage and religious
fulfilment through sacrifice (Evans-Pritchard 1940; Ocholla-Ayayo 1980; Burton 1987) and the Nilotic languages typically have large vocabularies to describe the color and pattern of their livestock (Coote 1994). The largest Nilotic languages are Dholuo in Kenya and Tanzania with four million speakers and Maasai with over a million. The Western Nilotic languages are famous among researchers for being exceedingly difficult for an untrained European ear, with three vowel lengths as well as tone in Dinka (Remijsen 2014). Some of the Western Nilotic languages of the Northern Lwoo subgroup exhibit another rarity in that they have OVS basic constituent order and are ergative (Andersen 1988). This unusual state of affairs was captured early on by Westermann (1911: 27) as “the Shilluk prefers to speak in the passive voice” but, perhaps because of the absence of a specific terminology, later researchers assumed ergativity and OVS to be absent from the African continent (Andersen 1988: 289–290, 320). Two Southern Nilotic languages, Okiek (Blackburn 1971) and Akie (Kaare 1996), are spoken by hunting and gathering communities interspersed with other Nilotes with sizable cattle herds.

1.2.8. Madagascar

Madagascar, situated in the Indian Ocean some 400 kilometers off the coast of Mozambique, is the world’s third-largest island. The Swahili sphere once reached the northern parts of Madagascar (Nurse and Hinnebusch 1993: 14, 22, 559) but the language is no longer spoken there (except by the older generation of a relatively recent Zanzibar offshoot on the small island of Nosse Be, see Gueunier 1989). On the contrary, it has been known since the early 17th century that the affinities of the languages of Madagascar lie not with the languages of the African mainland but with the large family of Austronesian, comprising some 1,200 languages found in an area stretching from Madagascar in the west to Easter Island in the east, and from the Formosan Islands in the north to New Zealand in the south. More specifically, as famously shown by Dahl (1951), the Malagasic languages belong more precisely to the Greater Barito subgroup of Austronesian, whose languages are spoken in Borneo. On the basis of careful study of the Malagasy lexicon and the issue of loanwords involving datable indirect borrowings from Sanskrit and Arabic, Malagasy is now thought to have arrived in Madagascar in the 7th century AD, brought by a Malay-speaking maritime economy (Adelaar 1995). There is also evidence of lexical interaction with Bantu languages, but this is likely to have occurred afterwards, and involved the Comorian Swahili varieties Ngazidja and Ndzwani (Dahl 1988), spoken on the islands northwest of Madagascar. Phenotypically, Malagasy populations today reflect the spectrum from African to Indonesian physical appearances not least due to the presence of the slave trade off the African coast in the last millennium.

Malagasy subsequently diversified in situ into a small family comprising 12 different languages in the count of Lewis et al. (2015), traditionally divided
amongst 18 named ethnic groups. The largest ethnic group in terms of numbers is the Merina in the northern central mountains (hence counted as Plateau Malagasy, 7.5 million speakers), the Betsimisaraka (1.8 million and 1 million speakers for the southern versus northern division) along the east coast and the Betsileo in the southern central mountains (also counted as Plateau Malagasy, 7.5 million speakers). The Malagasic peoples in the interior and east of the island, e. g., the Merina, are predominantly agriculturists whose principal crops are rice, manioc and yams. In contrast, the traditional subsistence of the Antakarana in the north, the Sakalava in the west and the Bara, Mahafaly, Antandroy and Karimbola in the south is herding, in addition to agriculture (Schomerus-Gernböck 1975). Vezo is a predominantly fishing population living on the southwest coast (Astuti 1991), and close to them are traditionally hunting and gathering groups variously called Vazimba, Beosi and Mikea (Birkeli 1936; Stiles 1991) who have now largely abandoned their presumed ancestral way of life. Birkeli (1936) was able to record a scattered array of vocabulary items of Beosi, which he insisted was not only a language different from its neighbors but also contained non-Malagasy items, possible relics of an aboriginal population. Later researchers have found no language particular to the traditionally foraging ethnic groups (Tucker 2001) and find unproven the idea that these groups reflect a pre-Malagasy aboriginal population.

Until the 19th century Malagasy was rarely written, but an Arabic script was used not least by the Antemoro in the south (Dahl 1983) from the 15th century until the introduction of a Roman script for various Malagasy varieties by missionaries in the 19th century.

1.2.9. Southern tip

The indigenous languages of the southern tip of Africa fall into two sharply distinct layers: click languages and Bantu languages. Again, the Bantu layer can be inferred to be the last one. The click languages belonging to the Kx’a and Tuu families, spoken by hunting and gathering communities now confined to the Kalahari Desert, probably represent the oldest layer. The remaining non-Bantu click language family is the Khoe-Kwadi family. Güldemann (2008) argues that they entered southern Africa from East Africa, traveling with livestock herds through the narrow corridor with a sparse distribution of the Tsetse fly (Leak 1999: 79–90). If pastoralism was brought to southern Africa by the proto-Khoe-Kwadi speakers, some groups must have shifted subsistence or, alternatively, some hunter-gatherers shifted language, as there are Khoe-Kwadi languages presently spoken by traditionally hunter-gatherer communities. The largest non-Bantu click language, Namibian Khoekhoe (also known as Nama-Damara, of the Khoe-Kwadi family, 250,000 speakers), is spoken in Namibia and adjacent regions in Botswana and South Africa. In Namibia it is used in radio broadcasts, in the public administration and for teaching up to the university level.
Little remains of the click languages once spoken in the country of South Africa. Traill (1996) explains in detail how the Tuu language ǀXam, once spoken in the northern Cape area and beyond, was exterminated within the span of a century, as their ancestral lands were taken over and the population shattered into farm laborers or outright killed. Nevertheless, the ǀXam language is amply recorded through the work of Wilhelm Bleek and Lucy Lloyd, who interviewed Bushman prisoners beginning in 1870. For Nǁng, another Tuu family language of South Africa, the speakers were similarly dispersed and the language ceased to be used, and was declared extinct, until a series of aged speakers became known to researchers in the 1990s (Crawhall 2004).

The Bantu languages cover the remaining savanna territory. In Zimbabwe several dozen dialects commonly subsumed under the label Shona account for over 10 million speakers. The spread of Shona is likely related to the ancient empire of Zimbabwe (1220–1450), which boasted impressive stone structures and trade links to the east coast (Wieschhoff 2006). Gold was the trade item most demanded at the coast, while finds of beads and pottery from Asia in the Zimbabwe ruins testify to some of the reciprocal goods. Tswana has five million speakers in Botswana and South Africa. Although it is the largest indigenous language of Botswana, there are more speakers in South Africa. In the Caprivi area of Botswana and Namibia, one Bantu language, Yeyi, has adopted a large number of click consonants (Seidel 2008). A few other Bantu languages in the vicinity have clicks, but only in a small portion of their vocabulary (Bostoen and Sands 2012).

Though there are a number of large reference grammars and dictionaries produced mostly by missionaries, the Bantu languages of Angola and Zambia are among the least-known areas of Bantu languages. Two large Angolan languages are the similarly named Kimbundu, with 4 million speakers, and Umbundu, with 6 million speakers. The Bantu peoples of the Angolan and Zambian savannas are agropastoralist, save for the Kwisi (Estermann 1976: 20–30) and the Gciriku (Gibson 1981), who led a hunting-and-gathering-oriented lifestyle until recently, and the Ambo and Herero clusters, who are pastoralist.

The southernmost Bantu subgroup is the traditionally herding Nguni dialect cluster (Ownby 1985), with the principal languages being the mutually intelligible but sociopolitically separate Zulu (11 million speakers) and Xhosa (8 million speakers). Both Zulu and Xhosa have clicks, undoubtedly diffused from the click languages in situ. However, the words containing clicks are not confined to loanwords, but permeate the vocabulary more generally. In the case of the Nguni languages, there is a cultural practice called isihlonipho sabafazi that likely facilitated the replacement of vocabulary. According to isihlonipho, ‘women’s language of respect’, when a woman is married and moves into her husband’s family, she is taught new vocabulary by her sisters-in-law (and whenever necessary, advised by her mother-in-law). This new vocabulary is to replace words that contain syllables occurring in her husband’s family’s names which she may no longer use (Finlay-
son 1995). In this way, clicks entering the vocabulary may have been facilitated, especially if the wife is marrying in from a non-Bantu click-speaking ethnolinguistic group (Herbert 1995).

Also in southern Africa, warfare in the past centuries has altered the linguistic landscape. The Ngoni originate from Northern Zululand and spoke the corresponding Zulu dialect in the early 19th century. Due to the military expansion under Shaka Zulu, the Ngoni were pushed northeast. By that time, the Ngoni had military techniques and weapons superior to those of the local peoples and could establish themselves in the lands to the northeast, reaching present-day Malawi and finally southern Tanzania. As a result, ethnic groups called Ngoni are found along this route, but the languages they speak have assimilated to the local languages so that they are gradually more similar to them than to their original Zulu variety (Miti 1996; Ngonyani 2001).

1.3. Summary and outlook

The 2,000 languages of Africa manifest a vast range of typological, sociolinguistic and genealogical diversity, and undoubtedly more remains to be discovered. Despite centuries of interest, lack of documentation continues to be the main obstacle towards a full understanding of the languages of Africa. As a continent, Africa is the least-known area of the world, second only to the greater Melanesian area. With the advancing frontier of language endangerment, documentation and description are of the highest priority. Most African countries harbor un(der)documented and endangered languages, though Nigeria, especially the Hausa-dominated northern parts, stands out as the country with the largest numbers. Similarly urgent is the documentation of languages or registers relating to vanishing subsistence modes (e.g., hunting and gathering), professions (e.g., blacksmiths), ritual practices (e.g., initiation languages) and cultural expressions (e.g., drummed languages) as well as the study of sign languages.

All of the classic linguistic preoccupations such as undeciphered scripts, historical philology, language contact, multilingualism, language urbanization, areal typology, and comparative reconstruction can be found on the African continent. In many of these areas, the African data appear to be underutilized, either for their global linguistic significance or with respect to their value for other disciplines. The reconstruction of the Niger-Congo and Afro-Asiatic families (see Güldemann, this volume, Chapter 2) presents the greatest challenge of historical reconstruction in terms of the number of languages and presumed time depth. The quatrimestennial written record of Egyptian-Coptic represents the longest recorded vertical history of any language on the planet. The human genetic diversity of African populations is maximal and can be used to study the covariation between genes and languages (e.g., van der Veen et al. 2009; Dediu 2010). Africa provides ample
grounds for modeling the global emergence of linguistic diversity and its correlation with geophysical features, population density and state formations (cf. Nettle 1999; Coupé et al. 2013; Axelsen and Manrubia 2014). Fine-grained linguistic data can be profitably used to trace the spread of domesticated livestock and plants (Blench 2000) as well as technologies such as iron smelting (Lesage 2016) or pottery (Bostoen 2005).

Hopefully, all the abovementioned lines of research can be enhanced with more data, better access to data and computational support as African linguistics enters the digital age.

**Acknowledgements**

The present chapter has benefited from comments by Roger Blench, Mark Dingemanse and Tom Güldemann. The usual disclaimers apply.

**References**


A survey of African languages


A survey of African languages


Gheel, Georgius de. 1652. Vocabularium latinum, hispanicum et congense. Manuscript. Reprinted in revised form 1928 as “Les plus anciens dictionnaire bantu” by J. van Wing & C. Penders. “The question as to whether Joris van Gheel was actually the author or but a copyist is discussed at some length in the introduction to the [revised reprint] … There can be no doubt, however, that he copied a manuscript known to be in existence at the mission station of San Salvador before his arrival … Though the dictionary is probably not the work of a single person, it is practically certain that in the main it is to be ascribed to Robodero, a Spaniard whose name is the only one mentioned in the original text” (Doke 1959: 56).


