Centre for Global Economic History
RECONSTRUCTING THE HISTORY OF GLOBAL INEQUALITY

Universiteit Utrecht

CGEH Working Paper Series

The historical determinants of language status in Sub-Saharan Africa

Katalin Buzasi, Utrecht University

May 2015

Working paper no. 66

www.cgeh.nl/working-paper-series/

The historical determinants of language status in Sub-Saharan Africa

Katalin Buzasi, Utrecht University

Abstract: Languages are one of the most naturally evolving human institutions. Although the status of languages is closely associated with the well-being of their speakers in multilingual societies, this issue gains only a marginal attention in economics and development studies. This paper aims to reveal the long-term determinants of the status of languages in Sub-Saharan Africa, one of the most linguistically fragmented areas of the world. Based on economic, anthropological and historical studies, we identify the following factors that are likely to have long-term effect on the current status of African languages: geography, precolonial contact with Europeans and the Arabs (Islam), precolonial development of indigenous societies, Christian missions and colonial policies. The main data sources are the Ethnologue, the Joshua Project, Murdock's Ethnographic Atlas, Roome's map on the location of missions, various sources on the first Bible translations in African languages, and geographical data available online in shapefile and raster format. Using OLS and IV estimation techniques, we find that indigenous groups with relatively high socio-economic development before the European dominance, early Bible translation and relatively large share within current country borders are less likely to have their language in an endangered state today. Geographical variables and the nature of colonial policy seem to affect current language status indirectly through their impact on socio-economic development and missionary activities. The counterfactual analysis suggests that the contact with Europeans contributed to higher polarization in terms of language status.

Keywords: Sub-Saharan Africa, language status, precolonial socio-economic development, European colonization, Christian missions, counterfactual analysis.

JEL Codes: F54, Z13.

Corresponding author: Katalin Buzasi, K.Buzasi@uu.nl.

Acknowledgements: The author is grateful to Jan Luiten van Zanden (Utrecht University), Maarten Mous (Leiden University), and the participants at the Utrecht Graduate Seminar (Utrecht, The Netherlands, 2015), the FRESH Meeting (Trinity College, Dublin, Ireland, 2014) and the Oxford Economic and Social History Graduate Workshop (University of Oxford, UK, 2014) for useful comments. A different version of this paper written together with Peter Foldvari is going to be presented at the European Economic Association Conference (Mannheim, Germany, August 2015) and the European Historical Economics Society Congress (University of Pisa, Italy, September 2015)

1. Introduction

Sub-Saharan Africa is the most underdeveloped region of the world. According to the World Bank (2013), the fifteen poorest countries measured with per capita GDP (in current USD) are located in this region. Moreover, Sub-Saharan Africa exhibits the lowest average Human Development Index (United Nations 2014) and the perceived level of corruption is also strikingly high there (Transparency International 2014).

The existing literature provides various explanations for this extremely poor performance. Ethnolinguistic fragmentation (the probability that two randomly selected people in the society belong to different ethnolinguistic groups) has been found to undermine economic growth by increasing corruption (Mauro 1995) and the probability of insufficient public good provision (Easterly and Levine 1997). Another strand of the literature highlights the role of history in explaining Africa's exceptional (under)development path. Acemoglu et al. (2001) argue that where they could not settle due to unfavorable disease environment, European colonizers established extractive institutions which persisted to the present. Consequently, countries with higher settler mortality during the colonial era tend to have, for instance, worse property rights and weaker rule of law today. The proportion of white settlers is found to be a good predictor of contemporary institutional quality (Angeles and Neanidis 2015, Wietzke 2015). The nationality of colonizer also matters. Colonial policies implemented by different nations have different long-term economic and social impacts (Bertocchi and Canova 2002, Brown 2000, Grier 1999). Missionary activities, which include education (literacy, agricultural and technological knowledge) and medical care, have been established to be responsible for development differences within Africa. In areas where colonial rules did not restrict the activities of the different denominations disproportionately, competition between missionaries resulted in better schooling outcomes (Gallego and Woodberry 2010). And finally, the precolonial diversities in societal organization across ethnic groups seem to explain contemporary between- and within-country development differences (Michalopoulos and Papaioannou 2015 and 2013, Bandyopadhyay and Green 2012).

What is quite striking is that although the regulation of language use in public administration and schooling was a crucial element of colonial and post-colonial policies and it was missionaries who initiated the first linguistic works on the continent, language-related issues, other than ethnolinguistic diversity, gain only marginal attention in economics, economic history, and development studies. Languages, however, determine individual well-being and social welfare through various channels. In economics, language (in its abstract sense) is viewed as an institution which makes information flow cheaper and easier, thus facilitates communication and cooperation between individuals and groups (Smith 2010, Ostrom 2000).

Better and less expensive communication and cooperation at the level of individuals can result in better organization at the macro or societal level. Due to lower translation costs and benefits from the economies of scale, markets are expected to work more efficiently in linguistically homogenous countries (Martinez-Zarzoso 2003, Egger 2002). Language learning is a form of human capital accumulation which is supposed to increase personal earnings (Grenier 1982). However, certain languages have higher labor market value than others (Levinsohn 2007). Language as an observable cultural marker serves not only as a basis for social fragmentation, stressed by the ethnolinguistic diversity literature, but also as a basis for discrimination on the labor market (Pendakur and Pendakur 2002). Moreover, the speakers of the officially not recognized languages are likely to perform worse in schools, know less about infection and disease prevention, and less able to participate in political decision making (Harbert et al. 2009, Batibo 2005). Overall, linguistically fragmented regions are more likely to suffer from severe poverty (Romaine 2009).

This paper gives insight on how history has shaped the linguistic situation, or more precisely, one of its aspects, in Sub-Saharan Africa. The study has several novelties. First, instead of the ethnolinguistic fragmentation, which has already been extensively researched, the status of languages is put into the center of investigation. We analyze the role of the precolonial socio-economic development level of indigenous groups, European colonization and missionary activities in explaining why certain languages are officially recognized, standardized and have higher social prestige than others. Second, the units of analysis are the linguistic groups in this paper. Focusing on linguistic groups instead of countries makes it possible to understand within-country social differences (explained in more details in the next paragraph). Development studies explaining the impacts of ethnolinguistic fragmentation usually compare countries. Third, unlike historical and sociolinguistic studies that conduct case studies limited to certain languages or countries, our empirical analysis attempts to arrive at more generalizable results by incorporating almost 400 languages located in 47 Sub-Saharan African countries.

The results are broadly linked to the following research fields. Since the status of languages is closely associated with the well-being and capacities of their speakers, revealing the long-term determinants of the status of languages contributes to our understanding on the historical roots of socio-economic inequalities. Moreover, since languages are generally concentrated on fixed geographical territories, the factors that are responsible for the differences in language status are, at the same time, expected to explain regional (within-country) development differences to a certain extent. Finally, although it is highlighted above that this paper focuses on an aspect of the linguistic situation other than ethnolinguistic diversity, our results are indirectly related to that research direction as well. While anthropological and economic studies have revealed that certain geographical and climatic factors (proximity to the Equator, good land quality, and low

variability in rainfall) imply higher ethnolinguistic diversity (Kaufman 2015, Green 2013, Michalopoulos 2012, Cashdan 2001, Nettle 1999), the role of history is less known in this respect. Sociolinguistics provide evidence that if the expected benefits from identifying with a language are not high enough, people are likely to abandon that language and acquire a new one with higher economic or social values (Mesthrie et al. 2009: 248-251, Batibo 2005: 93-94). Thus, since the status of languages is positively associated with the rewards from using that language, historical factors that influence the status of languages also determine which languages are maintained and abandoned. In the end, these individual decisions are expected to influence the size and the number of linguistic groups within a territory, i. e. linguistic diversity.

The paper proceeds as follows. The next section provides an overview on the most important historical factors that have shaped the linguistic situation of Africa in the past centuries. Also, we derive five hypotheses that are going to be tested in the empirical section. Section 3 describes our data. The empirical analysis and the discussion of the results are presented in Section 4. The last section concludes.

2. Historical background

This section provides an overview on the main periods of Africa's linguistic history which serves as the theoretical framework for the empirical analysis in the next section. We discuss the socio-economic development of indigenous societies, missionary activities and European colonization from the aspect of their linguistic impacts. Although contact with Europeans had been taken place continuously via trading and missions from the 15th century onward, the terms 'European colonization' and 'colonial era' in this paper refer to the period after the Berlin Conference in 1884-85 when the scramble for Africa began and the geographical boundaries were officially set up.

2.1. Linguistic situation before the colonial era

Before European colonization, language dynamics in Africa were predominantly driven by internal factors such as geographical and ecological conditions, invasion, and assimilation; and external factors such as the Arab conquest beginning in the 7th century and the establishment of Christian missions from the 15th century onward. Trade, which was determined by both internal and external factors, is special in this listing.

Empirical studies have shown that ethnolinguistic diversity is positively associated with closeness to the Equator (Cashdan 2001), longer mean growing season (Nettle 1999), and low variability in land quality, rainfall, temperature and elevation (Kaufman 2015, Green 2013, Michalopoulos 2012). The most commonly accepted theory to explain this relationship is that

ethnolinguistic diversity is higher in areas where geographical and climatic conditions make self-sustaining relatively easy, hence large scale cooperation and common actions between people or groups are not encouraged or necessary (Nettle 1999).

Although the climatic conditions of the Sahel region did not favor settled agriculture, the presence of rivers, the lack of physical barriers and the existence of salt and gold as valuable products of that time facilitated long-distance trade which became the basis of centralized kingdoms after the 8th century (Mansour 1993). The potential reward from trade served as the base for centralized states in other parts of the continent as well (Fenske 2014, Bates 1983). Wars and the expansion of these precolonial empires often resulted in major social, political and linguistic changes of the invaded societies. The enlargement of the Mali Empire, for instance, from the area of the current administrative region of Segou in the 13th century toward the oceanic coast in the West in the 16th century resulted in the massive spread of the Manding language (Mansour 1993).

Anthropological and historical documents argue that ethnic identity as understood by the modern Western world did not exist as a reality in Africa before the arrival of Europeans and 'group membership' was a highly fluid concept (van den Bersselaar 1997, Harris 1988). Ecological and economic shocks (famine and drought) often resulted in the displacement of people to new areas, reconfiguration of existing group ties and dependency of the dominant local groups (Stock 2012 p. 86.) which did not leave language use patterns intact. Murdock (1959) lists several precolonial examples on how certain groups shifted to the languages of their economically more developed neighbors. For instance, the Sanye, a hunting tribe in current Kenya took over the language of the neighboring Bararetta Galla (Oromo) from whom they adopted the rudiments of animal husbandry. Other examples include the Beni-Amer (current Eritrea) that adopted Tigre and the Shabelle (current Somalia) who completely acculturated to the dominant Somali.

The first missionaries arriving in the mid-15th century with the main aim to explore Africa were predominantly Catholics sponsored by the Portuguese. Protestant missions started to operate only in the mid-18th century (Asafo 1997, Welmers 1974). Beyond spreading the Word of God and translating the Bible to local languages (discussed later), they provided agricultural and technical education.

Although trade was essential at the local level as well, it is trading between neighboring and often culturally distinct groups and international trade (the export of slaves, gold and ivory to Europe and America and the import of clothes) that are expected to have had more important linguistic consequences.[1] First, intergroup and interregional trading fostered the emergence of

[1] A good review on the geographical extension and the main products of local, intergroup and long-distant trade in Tropical Africa is found for instance in Konczacki and Konczacki (1977).

indigenous lingua francas (Heine 1970) such as Hausa (Nigeria, Niger), Songhay and Manding languages (Mali) and Swahili (Tanzania, Kenya, Uganda) which still have high esteem and important role in interethnic communication in contemporary African societies (Lewis et al. 2014, Mansour 1993). Second, due to the strong Muslim Arabic conquest associated with high political and religious centralization and control over the trans-Saharan and Red Sea trade routes, Arabic, a non-indigenous language, was also used as a lingua franca. Third, Africa provides several examples that lingua francas in the long run had spread to the expense of local languages and became primary languages for certain groups. The founders of the Niger-Gambia trade axis were the Soninke people who adopted Manding in the heyday of the Mali Empire (Mansour 1993). The speakers of Tumzabt which is a threatened language in Algeria also use Algerian Arabic as primary language (Lewis et al. 2014). Fourth, trade with Europeans accompanied with agricultural and technical education, and medical services offered by Christian missionaries (Welbourn 1971) contributed to positive attitude toward European languages. Several historical sources document the demand of indigenous people for European languages and missionary education before the establishment of colonial administration. In a letter to a trading firm with the goal to get missionaries around 1850, the chiefs of Bonny (Nigeria) wrote: 'We expect that those gentlemen to be sent to us shall be capable of instructing our young people in the English language' (Ajayi 1965 p. 56.).

Written form of languages was far from common and highly concentrated to the North, the Horn of Africa and the Eastern coastal region. Although a few languages developed their own scripts (for instance the Ge'ez of Amharic and Oromo in Ethiopia and the Tifinagh of Berber languages in North Africa), the majority of tribes or kingdoms that had written languages before the arrival of Christian missions were those participating in the trans-Saharan or the Indian (slave) trade and strongly influenced by the Islamic (Arabic) culture. Hausa (Nigeria, Niger, Benin), Tamasheq (Mali), Kanuri (Nigeria), Wolof (Senegal), Soninke (Mali, Senegal), Fulfulde (Nigeria), Swahili (Tanzania, Kenya, Uganda) and Songhai (Mali, Niger) languages used a version of the Arabic script (ajami) that was adopted to the writing of non-Arabic languages (Albaugh 2014, p. 23, Warren-Rothlin 2009, Adegbija 1994). The earliest linguistic works (Bible translation, dictionaries and the development of Latin script for African languages) by the Christian missions were related to languages already existing in written form. According to Groves (1964), the first Bible portion (Psalms) translation on the African continent was compiled in Abyssinia in 1513. Several Coptic translations appeared in 1663, 1786, and 1811. Bible portions were translated to Amharic in 1824, Oromo in 1839, Swahili in 1847, Hausa and Kanuri in 1853, and Wolof in 1873. Missionary linguistic works related to languages along the coast and located around early trading ports were implemented also relatively early. Although Fanti (one of the Akan dialects in current Ghana) is the only non-Sahel region language in which

Bible portions were available before 1800, several parts of the Script were translated for instance to Ga (Ghana) in 1805, Susu (Guinea and Sierra Leone) in 1816, Nama (Namibia, South Africa) in 1831, Xhosa (South Africa) in 1833, Grebo (Liberia) in 1839, and Zulu (South Africa) in 1846.

Numerous historical sources inform us on the nature and intensity of the relationship between European missions and local people, or more precisely, kings, chiefs and the elite in the 19th century. In 1823, Radama I (1788-1828), the Merina king on Madagascar, decided to modernize his country to keep Europeans at a distance. He proclaimed the Latin script of Malagasy introduced by Reverend Jones (London Missionary Society) to be the official alphabet of the kingdom and allowed the LMS to establish schools, develop teaching materials and promote literacy. However, schools concentrated on the central highland and were available only to the Merina nobility (Steinhauer 2005 pp. 78-80). Crowder (1968 p. 9.) points out that the chiefs in Sierra Leone were asking for missionary schools twenty years before the establishment of the Protectorate in 1896. Ajayi (1965) reports an excellent relationship between the Efik king (Eyo) and the mission under Hope Waddell (1804-1895) arriving in Calabar (Nigeria) in 1846. To adapt to the changes in international commerce, Eyo wished to support the transition from a former slave-based economy to palm oil production through education provided by missionaries.

2.2. The linguistic effects of European colonization

The 'scramble for Africa' beginning in the 1880s not only opened up a new era in African economic, social and political history, but significantly influenced the linguistic landscape of the continent.

Colonial boundaries settled in 1884-85 at the Berlin Conference are commonly accepted as a result of arbitrary decisions (Hargreaves 1985, Englebert et al. 2002) without taking natural geographical circumstances, original social organizations and indigenous group distribution into account. An important consequence of the fixed colonial borders is that they anchored or froze the previously dynamic economic and political relationship between indigenous groups and determined which groups were to compete with each other for the colonial administrative positions, education and other missionary services. Dividing some groups between countries might have resulted in the weakening of certain groups and the strengthening of others. As Posner (2004) shows, the current political relevance of the Chewa and Tumbuka people (both divided between Zambia and Malawi) is highly dependent on their relative size within the current country borders.

Although, as it is discussed above, missionary activities were significant from the 15th centuries, the massive penetration of the inland regions of Africa and organized linguistic works

started only in the late 19th and early 20th century. Missionaries were dedicated to language standardization and development through establishing the written form of certain African languages (Latin script), compiling dictionaries, grammar books and orthographies that served education, literacy promotion and Bible translation purposes (see for instance Peterson (1997) on the process of creating the Gikuyu dictionaries in Kenya or Doke's report (1931) on the creation of a single Shona orthography).

Due to resource scarcity and other practical reasons, missionaries often had to decide which languages or dialects should be developed. For instance, Hausa was selected for standardization for its beneficial properties. Since it already existed in written form and due to its function as a lingua franca in trade, the cost and benefit (reaching several tribes with one language) ratio of its development, standardization and promotion was expected to be smaller compared to its non-written and locally concentrated counterparts (Adegbija 1994). However, there were other decisive linguistic characteristics. In relation to Southern Africa, Gilmour (2007) highlights that some languages (such as Xhosa and Tswana) were found to be sufficient and promoted for religious translation, others were considered ineligible (lacking of crucial religious concepts) and inefficient (certain terms needed to be expressed with circumscription), thus became discouraged (such as Khoi and San languages). In certain cases a single dialect was chosen as the basis for standardization (such as Ki-Unguja (Zanzibar) dialect of Swahili over the Mombasa dialect), in other cases the unified orthography was based on various dialects (the Shona language was a distilled variety from the Zezuru, Korekore, Karanga, Manyika and Ndau dialects) (Chimhundu 1992, Ansre 1974, Whiteley 1956).

However, missionary activities cannot be viewed isolated from the system of colonial rules in which they operated. Missions were not only the receptive subjects of the colonial administration with the exclusive aim of spreading the Word of God but they represented significant economic and political power. They contributed to setting up economic and social conditions (such as creating new markets) that made colonial utilization profitable (through training masses and local administrators) (Fabian 1983). An old Gikuyu (ethnic group in Kenya) proverb 'There is no difference between a missionary and a settler' (Oliver 1952) illustrates that Christian mission was perceived by local people as the representative of the colonial power.

Each element of the colonial practice (including language and education policy and the regulation of missionary activities) implemented by the different European empires were designed to match their overall underlying philosophies. The French and Portuguese considered Africans as people that need to be civilized (Conklin 1997) and to be assimilated into the broad metropolitan community (Betts 2005, Bokamba 1991, Spencer 1974) in the long run. The exclusive use of French and Portuguese in the colonial administration and at all levels of education was seen as a main tool to achieve these goals. Local languages were not only not

promoted or ignored but actively discouraged (Bokamba 1991, Spencer 1974). In contrast, the British acknowledged local circumstances to a greater extent. Local languages were extensively studied, documented and standardized and some of them (mostly dominant local vernaculars) were used as the language of education in primary schooling. Mother tongue education was seen as a tool to support the success of English learning in the long run (Atkinson 1987, Ward 1940). However, English served as a language of higher education and the colonial administration (Berman 1975).

Education policy (beyond the rules regarding the language of instruction) was also designed in line with the general colonial vision. Although the French government established numerous public schools and allowed limited (mostly Catholic) missionary activities, Hailey (1945 p. 1261) points out that missionary activity remained sufficient and the proportion of students enrolled in public schools in French West Africa did not exceed 15%. British preferred to 'outsource' most of their education to missions irrespective of their denomination (Frankema 2012) and all religious groups could apply for grants from the state (Gallego and Woodberry 2010). The French colonial education system is often described as highly selective and 'elitist' which provided disproportionate access to the French type education, and consequently, to prestigious state positions for certain ethnic groups (Blanton et al. 2001 p. 478). Compared to the French, the British enrolled more students into the education system and invested higher percentage of their budget in schooling (Frankema 2012, White 1996). However, due to high selectivity, the French system resulted in higher per capita education expenses (White 1996).

European colonization induced several changes in the economic and political organization of traditional societies. Based on a fieldwork conducted between 1949 and 1952, Middleton (1971) provides an overview on the consequences of European colonization in the Lugbara (Uganda and the Democratic Republic of Congo) society. Due to railway and the introduction of the tobacco as a cash crop during the era of the British administration (1894-1962), from a subsistence economy, the Lugbara people transitioned to a peasant economy with larger and continuous settlements, labor work and institutionalized markets. Instead of hereditary succession, chiefs that ruled certain geographical units (counties, sub-counties and parishes) with fixed borders were appointed by the government. Although older generations considered this period as the destruction of traditional values, younger people saw its economic advantages.

From the aspect of this paper, the most important consequence of the above discussed policies is that they changed the costs and benefits of maintaining local languages which determine language dynamics (language maintenance and decline) in the long run (Harbert et al. 2009). Since they were attributed with political power and prestigious socio-economic status, both French and English enjoyed favorable attitudes (Bunyi 1999), which is expected to have increased the incentive to learn them. Yet, this encouragement could have been counterbalanced

in former French colonies by two factors. First, since ethnic groups experienced unequal chances to gain state positions, individual motivation to learn French might have remained low among the members of less preferred ethnic groups. Second, if lower proficiency in mother tongue is associated with lower success in second language acquisition as suggested by socio-linguistic literature, the French system with the language of the colonizer as the exclusive mean of instruction is expected to work less efficiently compared to the British system. Local languages selected for education purposes also gained considerable esteem (for instance the Efik in Nigeria) (Adegbija 1994).

2.3. The postcolonial era

Since independence did not result in essential changes in the nature of language policies[2], the aforementioned colonial practices have several long-term linguistic and language-related social consequences. Due to the ignorance and discouragement of local languages, former French West Africa still had the least developed lingua francas in the world in the beginning of the 1990s (Bokamba 1991). While the share of population with a written language reached 76%, 79% and 80% in 1950 in British, Belgian and Portuguese colonies respectively, the corresponding share was only 58% on the French territories (Albaugh 2014 p. 27 and p. 70). Recent empirical studies show that former British colonies still exhibit higher average levels of education and literacy (Cogneau 2003, Brown 2000) which is partly the consequence of higher competition between the Protestant and Catholic missions (Gallego and Woodberry 2010).

The most remarkable novelty of the past two decades in linguistic terms has been the increased focus on endangered and minority languages. Having acknowledged the severity of social and ecological issues related to language decline (identity and culture loss, difficulty in accessing education and health services, the high possibility to get excluded from political decision making, and the interconnectedness between biodiversity, linguistic diversity, and poverty) a number of international organizations and research groups have committed to elaborate policy guidelines to support language revitalization programs that meet global, regional or local needs.[3] In parallel to this worldwide phenomenon, the majority of former

[2] The exclusive use of French was maintained after independence in most countries (except for Northern African countries and Madagascar) to avoid tribal conflicts and providing access to global development by using a language of wider communication (Bokamba and Tlou 1977).

[3] In the past decades language decline and the loss of linguistic diversity have been acknowledged a serious problem affecting various areas of the globe. According to the latest edition of the Ethnologue, 34 percent of the currently known languages are threatened (likely to die out in the short run) and about 370 languages have died out (lost all native speakers) since the 1950s. The UNESCO launched its Endangered Languages Program in the early 1990s and beyond helping policy makers through professional advices and meetings it publishes results as the UNESCO Atlas of Endangered Languages (Moseley 2010). The SIL International (http://www.sil.org) as the largest Christian non-profit organization is engaged in studying and documenting languages of the world in order to promote literacy, support Bible translation and

French colonies have recently moved toward the recognition of local languages in education (Albaugh 2014).

2.4. Hypotheses

Based on the literature discussed above, we derive the following five hypotheses in relation to the current status of Sub-Saharan African languages. In this paper, language status is understood as a complex concept which is defined in Section 3.1 in more detail. For now, the following 'definition' is sufficient. Languages with official recognition in public administration and/or education, with established orthography and high social esteem are considered to have higher status compared to those that are officially not recognized, unwritten and not favored by the society.

First, we expect that the languages of those groups that were more developed in social and economic terms before the European dominance have higher status today (Hypothesis 1). This argument is supported by the sociolinguistic and economic literature. Batibo (2005 pp. 93-94) argues that language shift takes place when a weaker linguistic group comes into contact with another language that has higher demographic, economic, cultural or political value; the speakers of the weaker language want to identify with the other group in order to share in the benefits attached to that language. This is in line with 'identity economics' (Bodenhorn and Ruebeck 2003) which views identity as a fluid concept that is not 'determined by nature' but dependent on individual choices. If the benefits from acquiring a new identity are higher than the costs of abandoning the existing one, people are likely to give this latter up and choose the more advantageous strategy.

Second, size matters. It is commonly accepted that larger communities have higher chances to survive and spread widely since they execute demographic pressure on smaller languages (Batibo 2005). Also, more populous ethnic groups have more bargaining power to secure a higher legal status and financial sources for their languages. More speakers may also give rise to a kind of economies of scale, since with the same investment in learning a language more possible partners can be reached (or, alternatively, this can be seen as a network externality). As a result, we expect that the higher the share of a linguistic group within a country is the higher the language status should be (Hypothesis 2).

Third, we also expect that languages that were standardized earlier have higher current status (Hypothesis 3). As the historical evidence suggests, due to scarce resources, missionaries

language development policies. Their results are published in a comprehensive catalog titled Ethnologue: Languages of the world that serves as reference work in several disciplines addressing language related issues. The Catalogue of Endangered Languages (ELCat, http://www.endangeredlanguages.com/), joint project of The LINGUIST List and the University of Hawai'i Mănoa is a recent attempt with similar goals as the above mentioned two examples.

did not develop each language equally (thereby they optimized their use of scarce resources). High initial costs of language development encouraged missionaries to focus on a few selected languages instead of many, while the possible increasing returns made it profitable to focus on those languages which had already achieved a higher degree of development. They also promoted the use of such preferred languages among other groups originally with different primary language.

Fourth, the number of missions on a territory is associated with the amount of human resources that can be utilized for linguistic works and the bargaining power within the colonial government. Thus, we expect that the higher concentration of missionary activities is associated with higher current language status (Hypothesis 4).

And finally, due to the promotion of local languages and the higher investment in primary education discussed in Section 2, we expect that the current status of African languages is higher in former British colonies (Hypothesis 5).

3. Data, variables and limitations

3.1. The dependent variable: language status

Similarly to most phenomena investigated by social and political scientists, the status of languages can be approached from various aspects. On the one hand, it is possible to measure status from one angle only (minimalist approach). By their legal status, languages can be officially recognized as the language of national or regional affairs and education or officially not recognized in any socio-economic domain. Languages without established orthography and available written material are usually considered inferior to standardized languages. Moreover, due to the number of speakers and their traditional role in trade for instance, some languages have higher social prestige (e.g. lingua francas) than others. On the other hand, the status of languages can be understood as a complex concept determined by various factors including those just mentioned above (maximalist approach).

This paper relies on the maximalist approach for the following reasons. First, since we aim to link our results to the broader development and ethnolinguistic diversity literature, we have to take into account that the well-being of the speakers and the linguistic choices of individuals are determined by linguistic factors other than their official recognition. For instance, although the Dioula language in Burkina Faso is not recognized officially, as the traditional trade language, it is widely used for interethnic communication and has high prestige. The second reason is that there is an available appropriate measure.

In this paper, language status is measured with the Expanded Graded Intergenerational Disruption Scale (EGIDS) (Lewis and Simons 2010) from Ethnologue (Lewis et al. 2014), which

summarizes numerous potential aspects of the language status in a single indicator. The original purpose of the EGIDS is to provide an appropriate tool that helps evaluate the severity of language endangerment and to determine the nature of actions that are required to overcome language decline (Footnote 4). The vitality (or endangerment) of languages is evaluated along five dimensions. Languages are considered less endangered if they are actively used as first language (sociolinguistic abbreviation is L1) and vehicular or lingua franca (key concept #1: identity function), officially recognized at the national or regional level (key concept #2: official recognition), transmitted uninterruptedly from generation to generation as first language (key concept #3: transmission), taught and/or used in the education system (key concept #4: literacy status), and younger generations speak it actively as L1 (key concept #5: youngest generation). According to these concepts, a 13-level scale is derived (see Table A.1 in the Appendix) where higher values reflect more disruption, i. e. higher chance to die out in the short run. The scale levels are hierarchical in nature, which makes it possible to use it as an ordered (ordinal) dependent variable in the empirical models estimated in the next section. With only one exception, the scale assumes that each stronger level of vitality entails the characteristics of the levels below. The one exception to this principle is level 3 (wider communication) where the vehicularity of languages is counted as being more important than the existence of an orthography and the use of language in education. Some languages that are widely used for intergroup communication are not used in education and have no written materials. Were these languages to lose that vehicularity, they would drop directly to level 6a. The main benefit of the aforementioned hierarchical structure is that if one of the underlying aspects changes, the status of the language also changes by holding the remaining aspects constant.

Compared to other existing language endangerment assessment methods (the Graded Intergenerational Disruption Scale (GIDS, Fishman 1991), the UNESCO's language endangerment classification scheme (UNESCO 2003, Moseley 2010), and the language status measures applied in previous editions of the Ethnologue), the EGIDS has various advantages which make it the best available language status measure for the purpose of this study. Unlike the UNESCO classification which handles safe languages as a single category and the GIDS by Fishman (1991) which does not provide an adequate description of the possible language statuses at the lower end of the scale, the EGIDS classifies languages into 7 non-endangered (from international (0) to vigorous (6a)) and 6 endangered (from threatened (6b) to extinct (10)) categories. Another benefit of the EGIDS compared to GIDS is that endangerment is not only dependent on the characteristics of the language itself (internal or micro factors such as intergenerational transmission) but on external or macro-level factors such as the attitude, government support, and the institutional context. As Lewis and Simons (2010) argue, although the nature of intergenerational transmission is an eligible factor to distinguish between the different level of

endangerment, at the upper end of the scale the level of institutionalization might be a more appropriate criteria to distinguish between the more developed levels.

Among the other possible classification schemes, the EGIDS is the most dynamic: it does not only assign high importance to intergenerational transmission as the GIDS does, but the final categories are designed to indicate whether a language is on the way to language shift or language development.

And finally, due to its hierarchical and refined structure, the EGIDS can easily be collapsed into only one aspect of the language status. This property is useful in two situations. First, if the research question requires information only on the official status, for instance, the researcher can consider all languages with EGIDS score above 5 as recognized languages. Similarly, all languages above EGIDS 6a have established orthographies. Or, all languages with EGIDS score above 6b can be seen as safe, while the remaining ones as threatened languages. Second, complex social, economic and political indicators (e.g. the Human Development Index[4] or the polityIV[5] which is designed to measure democracy) similar to EGIDS are often criticized that they are too vague or rely on arbitrarily selected components. For the beneficial characteristics discussed above, the EGIDS can be applied even in case the researcher prefers the minimalist approach.

3.2. Sample design

The basis of our dataset is a map by Murdock (1959) (available in shapefile format on Nathan Nunn's[6] website) that shows the geographical location of the indigenous ethnic groups (835 groups) in Africa and provides an appropriate framework to store and organize the data derived from various sources for an empirical analysis with individual languages (or linguistic groups) as the units of analysis (Figure A.1). Although the map and most of our sources would make it possible to conduct an investigation covering the whole African continent, the availability of information on the social and economic characteristics of indigenous African societies enables only a more limited analysis.

Data on the dimensions of socio-economic development of local societies are derived from the Ethnographic Atlas (Murdock 1967) which contains anthropological data on 299 African groups. But since we aim to control for the effects of colonial rules practiced by different European empires and the potential effect of borders, we account for the partitioning of groups between countries. Thus, for instance, the Kung group which is divided between Namibia and Botswana is included in our database twice. This procedure also makes it possible to account for

[4] http://hdr.undp.org/en/content/human-development-index-hdi Accessed: 05.05.2015
[5] http://www.systemicpeace.org/polityproject.html Accessed: 05.05.2015
[6] http://scholar.harvard.edu/nunn/home. Accessed: 12.11.2014.

the fact that the status of the same language often varies by country. For instance, the status of the Venda language is level 1 on the EGIDS in South Africa, while in Zimbabwe it is only level 5.

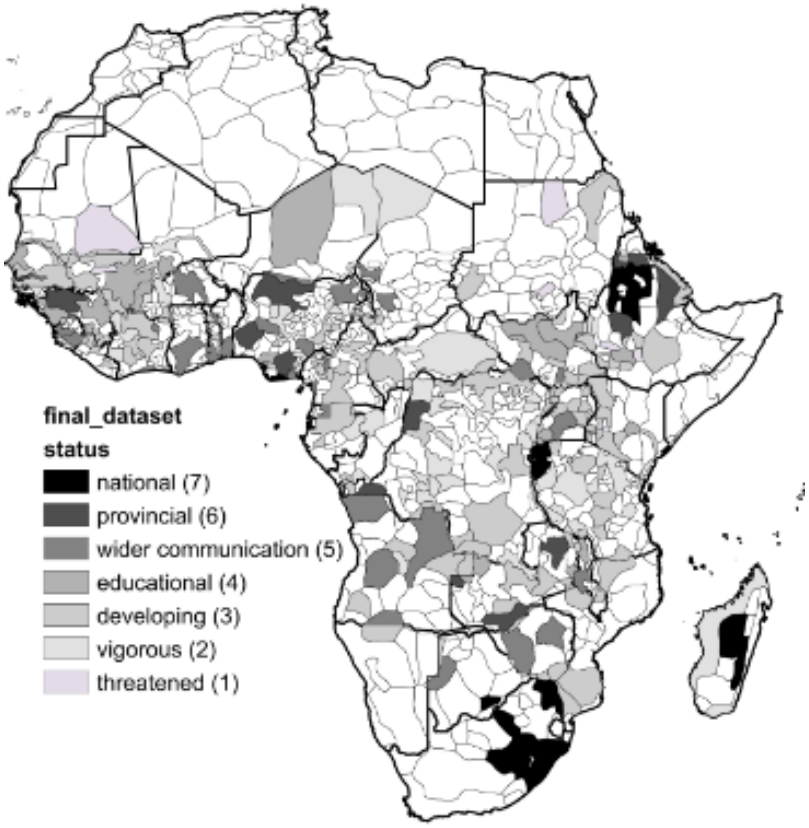
Some groups are excluded from the final dataset and the empirical analysis for certain reasons. Due to the presence of Arabic long before the European colonization, six groups located in Algeria, Tunisia, Libya, Western Sahara, Morocco, and Egypt (Ahaggaren, Barabra, Kabyle, Mzab, Riffian, and Siwans) are left out. Some other groups are ignored for more practical reasons. Certain groups represented in the Ethnographic Atlas are considered as only dialects and not separate languages in the Ethnologue. In this case we followed two strategies. First, if the group in the Ethnographic Atlas is a dialect in Ethnologue and the related main language is not included in the Ethnographic Atlas (for instance Pondo as a dialect of Xhosa and Lovedu as a dialect of Pedi in South Africa, and Xhosa and Pedi themselves are not included in the Ethnographic Atlas), we use the information on the dialect group as a proxy for the main language. Second, if both the main language and its dialect(s) are represented in the Ethnographic Atlas (for instance Afikpo as a dialect of Igbo in Nigeria or the Luapula dialect of Bemba in Zambia), dialect groups are eliminated. These aforementioned restrictions and the controlling for the country borders result in 389 Sub-Saharan African groups in the final dataset.

Then, the 389 groups from the Ethnographic Atlas are assigned with an EGIDS level of language status derived from Ethnologue discussed in the previous subsection. The distribution of languages per status level is presented in Table A.1 in the Appendix. The distribution of languages across status categories in our sample is somewhat different from that what we find in Ethnologue regarding the whole continent. Instead of EGIDS level 6a (vigorous), EGIDS level 5 is the most populous category in our sample. Our sample does not include international, nearly extinct, dormant and extinct languages. Due to the small number of languages, level 6b, 7 and 8a are combined and understood as endangered or tongues with the lowest status. Moreover, since our analysis is concerned with the status of languages instead of their endangerment level, the scaling is reversed in order to ease the interpretation of the coefficients in the empirical analysis. Data availability is shown in Figure 1.

At this point, however, we have to remark that combining the Ethnographic Atlas that refers to ethno-cultural groups with Ethnologue which basic units are languages requires some compromises. Our strategy implicitly accepts the common simplification of development studies that ethnic and linguistic groups are identical. Empirical works relying on national censuses and other surveys often use linguistic data to proxy the ethnicity of respondents when information on this latter is not available (Cheeseman and Ford 2007). Nevertheless, this is not necessarily true, since during the course of history some linguistic groups (mostly minority language groups) have taken over the language of another group as primary tongue or several groups belonging to the same ethnicity or cultural group speak different primary languages today (some

examples are given in Section 2.1.). As a consequence, groups in the Ethnographic Atlas and languages in the Ethnologue are not identical: certain groups in the Ethnographic Atlas could be assigned with several languages. When this is the case, the language with the highest reversed EGIDS level is assigned to the Ethnographic Atlas group. For instance, the Bete people in Ivory Coast in the Ethnographic Atlas can be assigned with three languages in the Ethnologue: Bete-Daloa and Bete-Guiberoua with EGIDS level 5 (reversed score is 3) and Bete-Gagnoa with EGIDS level 6a (reversed score is 2), which, according to our strategy, results in the final language status of 5 (reversed score is 3).[7]

Figure 1
Data availability



3.3. Socio-economic development of indigenous societies

In order to test Hypothesis 1, we need a variable to proxy the degree of socio-economic development of indigenous societies before the European dominance. By today's standards, a society is considered more developed if it exhibits higher GDP per capita and urbanization ratio, better quality institutions (less corruption, rule of law, democracy), higher average life

[7] This paper does not intend to evaluate the method according to which languages are decided to be distinctive or dialects in the Ethnologue. For instance, although it is said that Malagasy languages in Madagascar are mutually intelligible, the Ethnologue reports them as separate individual languages not dialects. Only the Merina (Malagasy, Plateau in Ethnologue) reaches the highest EGIDS level of 1.

expectancy and education level etc. Although, due to data unavailability and conceptual challenges, the development of precolonial African societies cannot be measured along these dimensions, recent studies have shown that certain traits of precolonial societies strongly correlate with current development measures. Empirical studies utilizing anthropological data from the Ethnographic Atlas by Murdock (1967) and the Corrected Ethnographic Atlas by Gray (1999) find that the variation in the precolonial centralization level of ethnic groups explains regional- and country-level differences in economic development (proxied with satellite images of light density at night) (Michalopoulos and Papaioannou 2013), public good provision and the quality of contemporary institutions (Gennaioli and Rainer 2007) across the African continent. Moreover, Bandyophadyay and Green (2012) show that precolonial centralization is highly correlated with several measures of development within Uganda, which point to the persistence of wealth and poverty from precolonial times to present.

Although it also utilizes information from the Ethnographic Atlas (Murdock 1967), contrary to previous studies listed above, our study is not limited to a single variable. While previous works were concerned with only one particular aspect of the social development of indigenous African groups, namely the existence of a centralized state (variable 33 in Gray (1999)), we consider development as a wider, multidimensional concept. We assume that the degree of the socio-economic development of indigenous societies[8] cannot be observed directly, but can be estimated from observable factors that are related to it. In this paper, the generalized structural equation model technique (gsem command in Stata 13) is applied to estimate the latent (unobserved) socio-economic development variable from various (observed) data in the Ethnographic Atlas that are related to the complexity of societal and economic organization and can be measured on an ordinal scale. Selected variables including the intensity of agriculture (variable 28), jurisdictional hierarchy beyond the community level (the second digit of variable 32), class stratification (variable 67), succession of the office of local headmen (variable 73), and the presence of high god (variable 34)) are described in Table A.2. All these variables are ordered and higher values represent higher complexity.

The underlying idea behind the measurement model is described in Figure 2. We assume that all observed variables of the socioeconomic development reflect a single underlying continuous latent variable. Hence, the observed indicators are all imperfect measurements of this underlying factor with independent errors (denoted by e). Consequently, their observed correlation is due to the common latent development factor only. Moreover, when more than one proxy of a single latent variable is available, using all available indicators to estimate the

[8] Since we can assume that the Ethnographic Atlas does not represent the precolonial characteristics of African societies purely (details provided in Section 4.2.) we name our variable as 'socio-economic development of indigenous societies' and refrain from the term 'precolonial development'.

latent factor is always preferable to a single proxy approach. The inclusion of more variables reduces the measurement error and avoids the bias resulting from focusing on only a single aspect of the latent factor.

The model outlined in Figure 2 operates under very similar assumptions to a factor analysis with a single factor, but while traditional factor analysis is based upon the assumption that all observed variables are continuous, here we allow them to be ordinal. This leads to more efficient and theoretically more appealing estimates. The coefficients are reported in Table 1.

The number of observations in Table 1 might require some explanation. It is said above that the number of languages (or groups) in the final dataset which takes colonial borders into account is 389. Still, the socio-economic development of indigenous societies (Table 1) is estimated on a sample that takes each group into account only once. Then, each division of the partitioned groups located in different countries is assigned with the same estimated score of the latent socio-economic development variable. There are two reasons for doing this. First, the development variable is aimed to grasp the socio-economic characteristics before colonialism when tribes were not divided. And second, including groups as many times as the number of countries they are assigned to by the colonial borders would give more weight to divided groups within the dataset, which would bias the estimation.

Figure 2

The measurement model of socioeconomic development

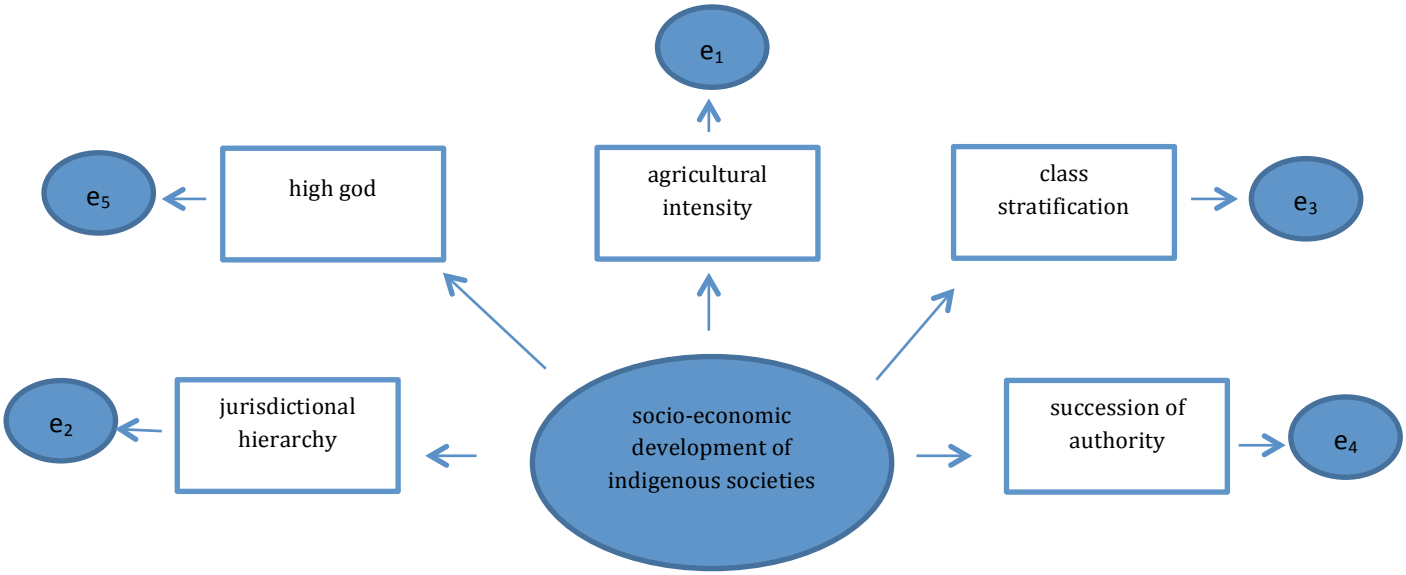


Table 1

Estimating the latent socioeconomic development of indigenous societies

	Model 1	Model 2	Model 3
	latent var: socio-economic development (dev 1)	latent var: socio-economic development (dev 2)	latent var: socio-economic development (dev 3)
agricultural intensity	**0.458*** (2.61)**	**0.611*** (3.10)**	**0.825*** (3.16)**
jurisdictional hierarchy (beyond community level)	**2.518** (2.42)**	**2.192*** (4.71)**	**2.049*** (4.27)**
class stratification	**2.509* (1.84)**	**2.962*** (2.87)**	**2.887*** (3.93)**
authority succession	-	**0.600*** (3.30)**	**0.413* (1.77)**
high god	-	-	**0.602** (2.35)**
number of observations	237	194	136
log-likelihood	-728.63856	-758.84932	-680.00877

Note: cuts are not reported, Z-stats in parentheses. . *, **, *** label significance at 10%, 5%, and 1% level respectively.

The three specifications in Table 1 are based on various sets of variables. It starts with a specification that leads to the highest number of observations (237 undivided groups) and continues with including additional components of development with less observations serving as robustness checks. We restrict the latent variable to have zero mean and unit standard deviation. The latent development factors are strongly correlated (Table A.3). While the procedure does not allow for a simple estimation of communality, Table A.4 reports the rank correlation coefficients between the estimated latent development variables and their respective components. The latent socio-economic development variable is positively related to all included variables. According to the results, the Rundi (Burundi), the Ruanda (Rwanda, Uganda, and the Democratic Republic of Congo), the Amhara (Ethiopia), the Songhay (Burkina Faso and Mali), and the Oyo Yoruba (Nigeria) were the five most developed or complex societies (Table A.5).

3.4. Population share, missionary activities and colonial policies

Ideally, testing Hypothesis 2 would require historical information on the share of ethnic groups within the colonial borders. Since these data are difficult to obtain, we can follow two, yet, imperfect strategies. First, we can utilize the area of each group in the Murdock map (1959) to proxy the share within the colonial borders. The benefit of this strategy is that it is based on historical information and the share of groups within the colonial borders effective before independence can be easily computed. The obvious drawback is that the appointed areas named after one certain group in the map might contain several other culturally related groups and the area share without information on population density does not necessarily capture the

population share. The second possibility, which is actually applied in this paper, is to rely on the Joshua Project which provide data on the current share of ethnic groups within country borders (*variable: ln(pop share*)). Although this option makes it possible to work with numbers in relation to each known ethnic and linguistic group, the data mirror the contemporary situation which must be handled (Section 4.3).

In order to test Hypothesis 3, the year of language standardization is proxied with the year of the first Bible translation (either portions, New Testament or the complete Bible) in African languages represented in the sample. We deduct the year of the earliest reported translation in our three available sources (Groves 1964, Lewis et al. 2014, worldbibles.org) from 2014 (the latest version of Ethnologue from which language status information is taken), which gives the final variable (*variable: bible age*) included in the empirical models. Groves (1964) reports the year of the first portion, New Testament and the complete Bible separately until the mid-1950s. The Ethnologue (Lewis et al. 2014) usually do not inform about portion translation, only on New Testament and the complete Script. The worldbibles.org dataset often reports unavailable information even in case of languages in which Bible is available according to the other two sources.

The location of Christian missionaries (Figure A.2), which is required for testing Hypothesis 4, is obtained from Roome's (1924) map that has been extensively used in studies seeking the long-term development impacts of missionary activities in Africa. Although the source would make it possible, we do not distinguish between Catholic and Protestant missions. We utilize the map from Murdock's Africa: Its people and their culture book (1959) digitalized by Nunn (2008) showing the geographical location and size of precolonial societies in the Ethnographic Atlas to calculate the number of missionaries per 100 square kilometers on their area (*variable: missions per 100 km²*). However, this variable is applied to proxy the European influence as discussed in Section 4.1.

The nature of colonial policies, which is the key variable in testing Hypothesis 5, including the use of languages in public administration and education is proxied with nationality of colonizer at independence (*variables: British, French, Belgian, Portuguese*) as provided by Bertocchi and Canova (2002).

3.5. Other variables: geography, climate and the spread of Islam

The empirical models discussed in the next section contain several control variables related to geography, climate and the spread of Islam. Geographic environment is controlled for with a number of variables. First, we control for the natural logarithm of the distance (in kilometer) of the geographic center of ethnic groups (calculated using ArcGIS 10.2.1) to the Equator (*variable:*

ln(distance to equator))[9], oceanic coasts (*variable: ln(distance to coast)*)[10] and inland waters (rivers and lakes)[11] (*variable: ln(distance to inland water)*). Second, each group is assigned with the areal mean and standard deviation of precipitation (in millimeter) (*variables: mean precipitation and std of precipitation*), mean temperature (in 0.1 Celsius degree) (*variables: mean temperature and std of temperature*) and altitude (in meter) (*variables: mean altitude and std of altitude*).[12]

As it is suggested by Bates (1983) and empirically tested by Fenske (2014), indigenous societies are likely to be more centralized where ecological diversity, hence the opportunity to profit from trade is higher. Following Fenske's strategy based on the vegetation map by White (1983), the ecological diversity (*variable: ecological diversity*) per ethnic group is calculated. The measure (based on the well-known Herfindahl-index) can be explained as the probability that two randomly selected points within the territory of an ethnic group belong to different major ecological zones.[13]

Retrieving historical data on the spread of Islam (*variable: islam*) before or during the colonial era in Africa is a challenge. But since the share of Islam has been relatively stable in the past hundred years and Christianity could spread rather to the expense of traditional religions (Asafo 1997), we use the current share of each group affiliated with Islam as the proxy of early Muslim Arabic influence from the Joshua Project.[14]

Summary statistics of the variables introduced above are found in Table 2. Table A.6 contains the data sources. Table 3 summarizes the expected signs of the main variables if the hypotheses in Section 2.4 hold.

---

[9] Shapefiles including tropical and polar circles, equator, and International Date Line is available at the website of Natural Earth. In this paper we use the shapefile with large scale resolution (1:1000000) directly downloadable at http://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-geographic-lines/.

[10] Shapefile on the coastline of Africa is obtained from http://omap.africanmarineatlas.org/BASE/pages/coastline.htm.

[11] Shapefiles (medium scale resolution, 1:50000000) containing the main lakes and rivers in Africa are obtained from the Natural Earth database directly downloadable at http://www.naturalearthdata.com/downloads/50m-physical-vectors/50m-lakes-reservoirs/ and http://www.naturalearthdata.com/downloads/50m-physical-vectors/50m-rivers-lake-centerlines/ respectively.

[12] These climatic variables based on information for the period between 1950 to 2000 obtained at climate stations (presented here: http://www.worldclim.org/methods) are downloaded from the WorldClim – Global Climate Data website in ESRI grid (raster) format with the smallest resolution (10 arc-minutes). Direct link: http://www.worldclim.org/current.

[13] The original 81 ecological zones in White (1983) are collapsed into 18 major types. Fenske (2014) applies alternative measures to proxy the potential gains from trade (such as ecological polarization, distance to ecological border, or using the Food and Agricultural Organization's ecosystem classification), however, his results are not sensitive to these types of changes in the empirical design. Thus, our analysis relies only on the ecological diversity indicator.

[14] joshuaproject.net accessed: 30.10.2014. 10:41
The Joshua Project is a research initiative (founded in 1995) that seeks to support Christian missionary activities among the least reached ethnic groups worldwide. They keep an up-to-date freely available database on the size of ethnic and linguistic groups and their religious affinity.

Table 2

Descriptive statistics

variable name	#	mean	std. dev.	min	max
language status	389	3.234	1.371	1	7
socio-economic development 1	237	0	1	-1.619	2.439
socio-economic development 2	194	0	1	-1.62	2.531
socio-economic development 3	136	0	1	-1.828	2.478
ln(pop share)	388	-4.502	2.045	-10.677	-0.162
bible age	389	73.913	53.475	0	250
missions per 100 km^2	389	0.014	0.04	0	0.448
British	389	0.442	0.497	0	1
French	389	0.306	0.461	0	1
Belgian	389	0.100	0.301	0	1
Portuguese	389	0.049	0.216	0	1
ln(distance to equator)	389	6.664	1.033	0.693	8.169
ln(distance to coast)	389	6.041	1.136	1.099	7.469
ln(distance to inland water)	389	4.412	1.179	0	6.779
mean temperature	389	238.22	24.81	135	287
std of temperature	389	10.68	8.04	0	49
mean precipitation	389	56.89	77.21	0	371
std of precipitation	389	8.89	11.72	0	92
mean altitude	389	699.24	496.98	9	2234
std of altitude	389	164.81	148.19	2	819
ecological diversity	389	0.319	0.223	0	0.802
islam	389	0.254	0.359	0	1

Table 3

The expected sign of the key variables in testing Hypotheses 1 to 5

	key variable	name in tables	expected sign
Hypothesis 1	socio-economic development of indigenous societies	socio-economic development	+
Hypothesis 2	share of the groups within the country population	ln(pop share)	+
Hypothesis 3	the age of the Bible translation	bible age	+
Hypothesis 4	the intensity of missionary activity	missions per 100km^2	+
Hypothesis 5	former British colony	British	+

4. Empirical results and discussion

4.1. Hypothesis testing

The hypotheses discussed in Section 2.4 are tested with an ordered logit model (Table 4) which dependent variable is the reversed EGIDS scores (Table A.1) and the main independent variables are those listed in Table 3. Each column of Table 4 relies on a different version of the socio-economic development variable of indigenous societies (introduced in Section 3.3). The specifications contain some additional controls for the spread of Islam and certain geographical

and climatic circumstances which have been identified as important sources of ethnolinguistic diversity (Table 2).

Table 4

Determinants of language status (ordered logit models with OLS estimation method)

	Specification 1	Specification 2	Specification 3
socio-economic development	**0.342** **(2.382)**	**0.347** **(2.228)**	**0.293*** **(1.647)**
ln (pop share)	**0.580*** **(8.247)**	**0.616*** **(7.950)**	**0.654*** **(6.949)**
Bible age	**0.013*** **(3.989)**	**0.015*** **(4.012)**	**0.014*** **(3.260)**
missions per 100km²	2.642 (0.710)	2.817 (0.717)	2.482 (0.736)
British	0.247 (0.579)	0.322 (0.717)	0.428 (0.748)
French	0.082 (0.182)	-0.154 (-0.321)	-0.424 (-0.666)
Belgian	-0.315 (-0.482)	-0.166 (-0.240)	0.445 (0.503)
Portuguese	-0.049 (-0.074)	-0.050 (-0.070)	0.357 (0.381)
ecological diversity	-0.705 (-1.236)	-0.791 (-1.218)	-1.075 (-1.279)
mean altitude	0.001 (1.414)	0.001 (0.763)	0.001 (0.950)
std of altitude	-0.000 (-0.098)	-0.000 (-0.129)	-0.001 (-0.123)
mean precipitation	**-0.006** **(-2.487)**	**-0.006** **(-2.510)**	**-0.006** **(-2.487)**
std of precipitation	0.024 (1.282)	0.020 (1.087)	0.011 (0.571)
mean temperature	**0.022** **(2.237)**	0.015 (1.518)	0.007 (0.624)
std of mean temperature	0.007 (0.137)	0.005 (0.094)	-0.010 (-0.141)
ln(distance to equator)	**0.506*** **(2.947)**	**0.506*** **(2.947)**	**0.506*** **(2.947)**
ln(distance to coast)	0.022 (0.147)	0.079 (0.447)	-0.322 (-0.518)
ln(distance to inland water)	-0.086 (-0.942)	-0.009 (-0.088)	-0.182 (-1.416)
islam	-0.286 (-0.566)	-0.644 (-1.079)	-0.744 (-1.100)
number of observations	343	282	195
Pseudo-R²	0.228	0.256	0.289

Note: Robust t-statistics are in parentheses. *, **, *** label significance at 10%, 5%, and 1% level respectively. Estimated cuts in specification 1 are 1.931, 5.506, 8.731, 9.286, 10.491 and 11.238. Estimated cuts in specification 2 are 0.758, 4.982, 8.331, 8.911, 10.171 and 10.787. Estimated cuts in specification 3 are -3.516, 0.902, 4.772, 5.32, 6.734 and 7.496.

Table 4 supports three out of the five hypotheses. Higher local socio-economic development before the European dominance, higher relative group share within the country population and early Bible translation are indeed found to be positively associated with the current status of languages. The concentration of missionary activities and the nationality of the colonizer do not seem to matter. (However, their indirect role in explaining language status is discussed in the following sub-sections.) There are only two significant geographical control variables. Languages located in areas with lower mean precipitation and further away from the Equator

are likely to exhibit higher status. The positive coefficient of the mean temperature is not robust across specifications.

The remaining of the empirical section aims to go beyond these relatively easily identifiable linkages and reveal the factors that influence the dependent variable indirectly, i.e. via a third factor.

4.2. The development of African societies and the contact with Europeans

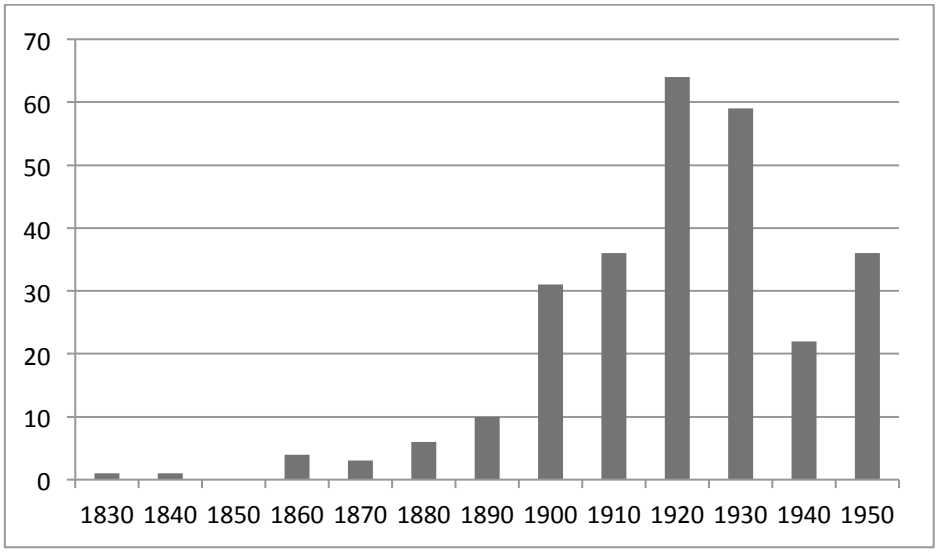
The development of traditional African societies is measured from the Ethnographic Atlas (Murdock 1967), which intends to reflect the characteristics of societies as they were in the 19th century just before the Scramble for Africa. However, since the majority of information is obtained from the first half of the 20th century (variable 102 in the Corrected Ethnographic Atlas shown in Figure 3), we argue that the Atlas is very likely to encounter precolonial and early colonial European influence to some extent.[15] We use the number of missions on the area of each group per 100 km^2 to proxy the pre-colonial and early colonial European influence. Data are obtained from the Roome's map (1924) which provides a snapshot on the geographical concentration of missionary activities in the early 1920s. The OLS regression model presented in Table 5 reinforces our expectations: the number of missions per 100 km^2 is positively associated with the socio-economic development measures (introduced in Section 3.3) even after controlling for the decade of the survey and some geographical and climatic variables that are assumed to be important determinants of the location of missions, the early contact with Europeans and the societal organization of traditional societies.

Although until now we have concentrated on the potential effects of European contact on indigenous societies, historical studies argue that the characteristics of local groups also determined the spread of missions (Johnson 1967). If simultaneity is indeed present in the regression of the estimated level of development on the number of missions per 100km^2, then the observed positive relationship between the intensity of European contact and the development levels in Table 5 are biased. We use instrumental variable technique (two-stage least squares (2SLS)) to identify the causal effect of European contact on the development of indigenous societies (Table 6a and 6b).

[15] The only study that uses the date of observations provided by Murdock (1967) in a similar way than this paper does is Henderson and Whatley (2014). By computing the number of years between the date when an ethnic group was colonized and observed, they provide evidence that the duration of colonization contributed to the shift in gender roles, the increasing relative position of men in lineage and inheritance systems, but also contribute to the reduced importance of polygyny in African societies.

Figure 3

The number of observed groups in the Ethnographic Atlas per decade from which information is obtained (variable 102)



Note: The figure represents the number of groups without taking partitioning between countries into account. Number of groups is 273.

Table 5

The relationship between the early European contacts measured with the intensity of missionary activities and the socio-economic development of African societies (OLS estimates)

	dep. var: socio-economic development1	dep. var: socio-economic development2	dep. var: socio-economic development3
missions per 100km^2	**2.976****	**3.325****	**3.838****
	(2.141)	**(2.311)**	**(2.585)**
decade of survey	**-0.006***	-0.005	**-0.008***
	(-1.877)	(-1.373)	**(-1.763)**
ecological diversity	**0.509***	**0.785****	0.596
	(1.666)	**(2.429)**	(1.554)
mean altitude	0.000	0.000	0.000
	(0.818)	(0.637)	(0535)
std of altitude	-0.001	-0.002	-0.002
	(-0.754)	(-1.380)	(-0.800)
mean precipitation	0.002	0.001	0.001
	(1.571)	(1.117)	(0.339)
std of precipitation	**0.013***	**0.015****	0.012
	(1.883)	**(2.290)**	(1.608)
mean temperature	-0.004	-0.004	-0.007
	(-1.020)	(-0.934)	(-1.440)
std of mean temperature	0.017	0.029	0.011
	(0.683)	(1.041)	(0.347)
ln(distance to equator)	**0.145****	**0.167*****	**0.174***
	(2.180)	**(2.563)**	**(1.834)**
ln(distance to inland water)	**-0.108***	**-0.140****	**-0.200*****
	(-1.840)	**(-2.242)**	**(-2.708)**
islam	**0.723*****	**0.716****	**1.009*****
	(3.010)	**(2.399)**	**(1.859)**
constant	**11.665***	9.877	**15.153***
	(1.739)	(1.267)	**(1.724)**
number of observations	236	193	135
R^2	0.229	0.246	0.273

Note: Robust t-statistics are in parentheses. *, **, *** label significance at 10%, 5%, and 1% level respectively.

We use the natural logarithm of the distance to the coast and the identity of the colonizer as additional instruments. These variables serve as proper instruments since they are exogenous, and they all affected the probability that missionaries settled in an area, while they are not influenced by the socio-economic development level of an ethnic group. The distance to the coast is assumed to capture the part of European influence that is associated with trading before the colonial era. It has been shown by a number of studies that Europeans came into contact with ethnic groups satiated near the coastal regions first, and these groups were the easiest and most logical destination for early missionaries as well. We also argue that the socio-economic development is associated with the distance from the coast only via trade and contact with Europeans. According to the Ethnographic Atlas, in the majority of societies fishing was not the main economic sector in the 19th century which supports our assumption that the main function of the ocean was to facilitate trade with Europeans. Hence, the distance to coast should be exogenous to the level of development, especially with the proximity to fresh water, the share of Islam as the proxy to trans-Saharan trade and ecological diversity as the proxy for potential reward from trading activities being included in the specifications separately. The colonizer dummies capture the part of the spread of missions that is associated with colonial policies. These variables serve as exogenous instruments since the partition of Africa was mostly a result of the play among European powers, and a region was claimed by a colonizer not because of its relative degree of development but rather by its availability for colonization. Different colonizers lent different support for missionaries as discussed in Section 2.2. Table 6a (the first stage results of the 2SLS estimates) reveals that the geographical concentration of missionaries was higher near the cost and in former British and Belgian colonies.

Table 6a

The causal effect of European contact measured with the intensity of missions on the development of African societies (first-stage)

	dep. var: missions per 100km^2	dep. var: missions per 100km^2	dep. var: missions per 100km^2
ln(distance to coast)	**-0.019*** (-3.649)**	**-0.021*** (-3.690)**	**-0.025*** (3.600)**
British	**0.022*** (3.054)**	**0.026*** (3.059)**	**0.038*** (3.069)**
French	0.005 (0.668)	0.005 (0.556)	0.005 (1.088)
Belgian	**0.022*** (1.997)**	**0.027*** (2.090)**	**0.042*** (2.409)**
Portuguese	-0.010 (-0.714)	-0.007 (-0.550)	0.013 (0.960)
year of survey	0.000 (0.738)	0.000 (0.922)	0.000 (0.700)
ecological diversity	0.015 (1.394)	0.016 (1.347)	**0.026* (1.705)**
mean altitude	0.000 (0.593)	0.000 (0.628)	0.000 (0.976)

std of altitude	**-0.00009*** **(-1.838)**	**-0.0001*** **(-1.891)**	**-0.0001** **(-1.259)**
mean precipitation	0.000 (0.908)	0.00 (0.228)	-0.000 (-0.006)
std of precipitation	-0.000 (-0.572)	0.000 (0.163)	-0.000 (-0.339)
mean temperature	0.000 (0.263)	0.000 (0.353)	0.000 (0.696)
std of mean temperature	**0.002**** **(2.163)**	**0.022**** **(2.237)**	**0.005*** **(1.715)**
ln(distance to equator)	**-0.006**** **(-2.210)**	**-0.005*** **(-1.728)**	0.001 (0.276)
ln(distance to inland water)	-0.000 (-0.349)	-0.001 (-0.638)	0.001 (0.276)
islam	0.000 (0.033)	0.001 (0.129)	0.006 (0.469)
constant	-0.008 (-0.040)	-0.087 (-0.365)	-0.118 (-0.335)
Cragg-Donald test of weak identification	13.987 (rel. bias<10%)	12.010 (rel. bias<10%)	10.151 (rel. bias<20%)
number of observations	235	192	135
R^2	0.315	0.329	0.395

Note: Robust t-statistics are in parentheses. *, **, *** label significance at 10%, 5%, and 1% level respectively. The H0 of the Cragg-Donald test is that the additional instruments are weak and lead to a bias. (rel. bias<10%) means that the bias of the 2SLS coefficients relative to the OLS coefficients is less than 10%.

Table 6b

The causal effect of European contact measured with the intensity of missions on the socio-economic development of African societies (IV estimates)

	dep. var: socio-economic development1	dep. var: socio-economic development2	dep. var: socio-economic development3
missions per 100km^2	**11.886***** **(3.100)**	**12.248***** **(3.176)**	**11.953***** **(3.169)**
year of survey	**-0.007*** **(-1.940)**	-0.006 (-1.434)	**-0.008*** **(-1.927)**
ecological diversity	0.370 (1.189)	**0.617*** **(1.855)**	0.273 (0.675)
mean altitude	**0.0002*** **(1.689)**	0.000 (1.509)	0.000 (1.340)
std of altitude	-0.001 (-0.851)	-0.002 (-1.436)	-0.002 (-0.870)
mean precipitation	0.002 (1.416)	0.001 (1.140)	0.000 (0.173)
std of precipitation	**0.013*** **(1.932)**	**0.014**** **(2.186)**	0.013 (1.576)
mean temperature	-0.003 (-0.714)	-0.003 (-0.620)	-0.007 (-1.537)
std of mean temperature	0.012 (0.511)	0.022 (0.825)	0.005 (0.170)
ln(distance to equator)	**0.197***** **(2.872)**	**0.212***** **(3.123)**	**0.266***** **(2.848)**
ln(distance to inland water)	**-0.129**** **(-2.243)**	**-0.162***** **(-2.619)**	**-0.232***** **(-3.310)**
islam	**0.878***** **(3.674)**	**0.915***** **(3.101)**	**1.116***** **(3.458)**
constant	**11.822*** **(1.718)**	10.039 (1.248)	**16.36*** **(1.816)**
Hansen J-test (df=4, p-value)	0.532	0.679	0.298
number of observations	235	192	135
R^2	0.126	0.128	0.156

Note: Robust t-statistics are in parentheses. *, **, *** label significance at 10%, 5%, and 1% level respectively. The H0 of the Hansen test is that the additional instruments are exogenous and valid.

The second-stage results from the 2SLS specifications (Table 6b) are similar to those of the OLS estimates in Table 5, however, we find a much larger causal effect of the European contact measured with the intensity of missionary activities on socio-economic development than with the OLS method. This may be not only because of removing the effect of simultaneity but also because of removing the effect of measurement errors. Namely, if there are random measurement errors in the explanatory variable (number of missions per 100km^2) then the coefficient estimated by OLS will be biased toward zero. Since our instruments are not correlated with these measurement errors, instrumentation will reduce the bias as well.

The positive significant coefficient of ecological diversity in column 2 is in line with Fenske (2014) who finds that due to potential higher rewards from trade, societies located at the border of ecological zones exhibit higher development (measured with variable 33 (jurisdictional hierarchy beyond community level) of the Corrected Ethnographic Atlas (Gray 1999).

It is interesting that some geographical factors that are previously documented to determine ethnolinguistic diversity are also found to be associated with socio-economic development. Higher variation in precipitation and distance to the Equator not only supports low diversity (Michalopoulos and Papaioannou 2012, Cashdan 2001) but seem to be positively associated with socio-economic development. This finding is in accordance with the anthropological argumentation proposed by Nettle (1999), that instability in climatic condition encourages cooperation which might be translated to higher societal complexity.

The finding that societies located closer to lakes and rivers (inland water) exhibit higher development suggests that the proximity of fresh water as a scarce resource might have encouraged more complex social organization. Issues related to the management of scarce resources such as water have inspired many scientific fields from economic and social history (Adams and Anderson 1988) through collective action studies (Garrido 2011) to contemporary politics (Peters 1994).

The positive significant coefficient of the current share of Muslims within ethnic groups (a proxy for Arab influence) can be considered as the reinforcement of the special importance of the trans-Sahara trade in the emergence of centralized kingdoms.

Table 6b suggests that although the concentration of missionaries does not affect the status of languages directly, it still has indirect positive effect through promoting the socio-economic development level of local groups which turned out to be significant in Table 4. The impact of the colonizer dummies can be interpreted similarly. Since the British and Belgian colonial policy fostered the spread of missions to a greater extent compared to independent areas and territories ruled by the French and Portuguese, the nationality of the former colonizer is found

to have a persistent indirect effect on language status. Geography and climate seem to play a more important role in determining the intensity of missions and the development of local societies (Table 6a and 6b) than in enhancing language development (Table 4).

4.3. The share of linguistic groups within the country population

The relative size of linguistic groups is one of the key factors which are assumed to determine the status of languages. Since historical data are not available, this study is based on the Joshua Project's database. However, since it shows the current distribution of linguistic groups, the effects of geography and history are likely to be encountered in the data. Table 7 provides evidence that the current share of linguistic groups is dependent on the historical socio-economic development, the Islamic influence and certain geographical factors. And since the socio-economic development of indigenous groups is found to be dependent on the intensity of missionary activities (shown in the previous section), the share of language groups within current country borders are indirectly influenced by it as well.

Table 7

Determinants of population share (OLS)

	Specification 1	Specification 2	Specification 3
socio-economic development	**0.631*** (5.567)**	**0.712*** (5.649)**	**0.643*** (4.051)**
islam	**0.986*** (2.665)**	**0.900** (2.050)**	0.529 (0.979)
British	-0.187 (-0.463)	-0.398 (-0.877)	-0.315 (-0.572)
French	0.112 (0.265)	-0.154 (-0.315)	-0.498 (-0.858)
Belgian	-0.857 (-1.552)	**-1.211** (-1.975)**	**-1.468* (-1.877)**
Portuguese	-0.626 (-0.884)	-0.776 (-1.061)	-1.166 (-0.686)
ecological diversity	-0.415 (-0.863)	-0.190 (-0.348)	-0.071 (-0.102)
mean altitude	**0.001** (1.981)**	**0.001* (1.779)**	0.001 (1.081)
std of altitude	-0.003 (-1.154)	-0.001 (-0.435)	-0.004 (-1.145)
mean precipitation	0.002 (1.056)	0.003 (1.253)	0.002 (0.645)
std of precipitation	**0.029** (2.113)**	**0.024* (1.719)**	**0.031* (1.911)**
mean temperature	**0.013** (1.983)**	**0.014* (1.909)**	0.012 (1.311)
std of mean temperature	0.043 (1.021)	0.019 (0.400)	0.071 (1.104)
ln(distance to equator)	**-0.182* (-1.883)**	**-0.205* (-1.948)**	-0.162 (1.101)
ln(distance to coast)	**-0.413*** (-3.871)**	**-0.402*** (-3.445)**	**-0.329* (-1.968)**
ln(distance to inland water)	-0.081 (-0.832)	-0.082 (-0.728)	-0.173 (-1.454)
constant	**-4.495** (-2.100)**	**-4.558* (-1.836)**	-3.834 (-1.245)

number of observations	343	282	195
R^2	0.236	0.248	0.241

Note: Robust t-statistics are in parentheses. *, **, *** label significance at 10%, 5%, and 1% level respectively.

4.4. Counterfactual analyses

Having identified the factors which have directly or/and indirectly determined the current status of languages, we conduct counterfactual analysis to find out if the language status pattern in Sub-Saharan Africa would be different if there was no European influence. This technique is designed to overcome the problem that we cannot carry out historical experiments. Counterfactual analysis has been used in several economic studies to measure the effects of historical events. Nunn and Qian (2011) reveal that the introduction of the potato is responsible for about one-quarter of the population growth and urbanization in the Old World between 1700 and 1900. Fernihough and O'Rourke (2014) show that the introduction of coal-using technologies explains about 60% of the increase in the European city population between 1750 and 1900.

Our goal is similar to these works. We aim to compare two situations: the pattern of the current status of African languages with and without European influence. Using the results presented in Tables 4 to 7, we can reconstruct the language situation without missionary activities, Bible translation and colonization by following the steps detailed below.

The socio-economic development of indigenous societies measured from the Ethnographic Atlas and the share of language groups are positively related to current the status of languages (Table 4). However, the intensity of missionary activities affected the development of indigenous groups (Table 6b) which has contributed to higher group share within the country (Table 7). Moreover, the nationality of the colonizer is found to exert a significant influence on the relative size of language groups in certain cases (Table 7). In order to see how the status of languages would be without the European influence, these direct and indirect effects should be filtered out.

The indigenous socio-economic development and the ln(pop share) variables without European influence are computed as shown in Formula (1) and (2), respectively. Coefficients are taken from Table 6b and Table 7. The formulas are presented in the case of the first type indigenous socio-economic development variable (specification 1 of Table 1) introduced in Section 3.3.

Let us consider the case of the Ambo in Angola (former Portuguese colony) where the number of missions per 100 km^2 is 0.007. The first type socio-economic development estimated from the Ethnographic Atlas value is 0.919 which reduces to 0.837 (0.919-11.886*0.007) if we filter the effect of missions out (Formula (1)). In other words, the presence of missions contributed to 0.082 (0.919-0.837) higher early socio-economic development in the case of the

Ambo group. According to Formula (2), the natural logarithm of the Ambo group share within Angola would be -3.381 (-3.956-0.631*0.082+0.626) instead of -3.956 if there was no European influence. The distribution of the socio-economic development with and without European influence is presented in Figures A.3a to A.3c. The order of the ten most developed indigenous societies with and without European influence is shown in Table A.6 in the Appendix.

Formula (1)

Indigenous socio-economic development without European influence

$$socioeconomic\_development1(without) = socioeconomic\_development1 - 11.886 * missions / 100km^2$$

Formula (2)

The current share of linguistic groups without European influence

$$\begin{aligned} &\ln(pop\_share)without = \ln(pop\_share) \\ &-0.631 * (socioeconomic\_development1 - socioeconomic\_development1(without)) \\ &+0.187 * British - 0.112 * French + 0.857 * Belgian + 0.626 * Portuguese \end{aligned}$$

The language status without European influence is computed according to Formula (3). Coefficients are taken from Table 4. The formula is shown only in the case of the first type socio-economic development variable. Since language status is a discrete dependent variable which is estimated with an ordered logit model, the computation of the language status without European influence is not straightforward. Instead of the observed EGIDS, our starting point is the predicted status which is a continuous variable. Instead of removing the effect of Bible translation (bible age) completely, we consider the difference between the actual and the average bible age.

Formula (3)

Current language status without European influence

$$\begin{aligned} &language\_status(without) = language\_status(predicted) \\ &-0.342 * (socioeconomic\_development1 - socioeconomic\_development1(without)) \\ &-0.580 * (\ln(pop\_share) - \ln(pop\_share)without) \\ &-0.013 * (bible\_age - 73) - 0.247 * British - 0.082 * French \\ &+0.315 * Belgian + 0.049 * Portuguese \end{aligned}$$

Again, we take Ambo in Angola to illustrate the use of Formula (3) in practice. Ambo is a language of wider communication (reversed EGIDS is 5) which has had a Bible translation since 1878. The predicted (continuous) language status is 9.069. Without European influence, this value would be 8.558 (9.069-0.342*0.082-0.580*(-0.585)-0.013*(136-73)+0.049) which equals

3 on the reversed EGIDS according to the estimated cut values listed under Table 4. This result can be interpreted that Ambo in Angola would have a lower status today without European influence. Or, in other words, Ambo has benefitted from missionary activities and colonization.

The effect of early European contact and colonialism on the distribution of language status in the case of the first socio-economic development variable is presented in Figure 4. The same figures for the second and third type socio-economic variable are found in Table A.4a and A.4b in the Appendix. According to our models, the distribution of the status of languages would be similar to the current distribution in terms of the mode: category 3 which is EGIDS 5 (developing) in the original coding is the most populous. However, none of the languages belong to category 7 (EGIDS 1 (national)) and category 1 (combined group from EGIDS 6b, 7 and 8a (threatened)) anymore when socio-economic development is measured with the first two types. The only language that falls into the highest development category 7 when socio-economic development of traditional societies is measured with the third type variable is the Sotho (Lesotho and South Africa). Thus, Figure 4 (and Figures A.4a and A4.b) suggests that the colonial rule and missionary activities altogether increased polarization in the distribution of the status of languages in Sub-Saharan Africa which can be traced in the smaller kurtosis and a relative heavy-tailedness of the distribution when European influence is not filtered out.

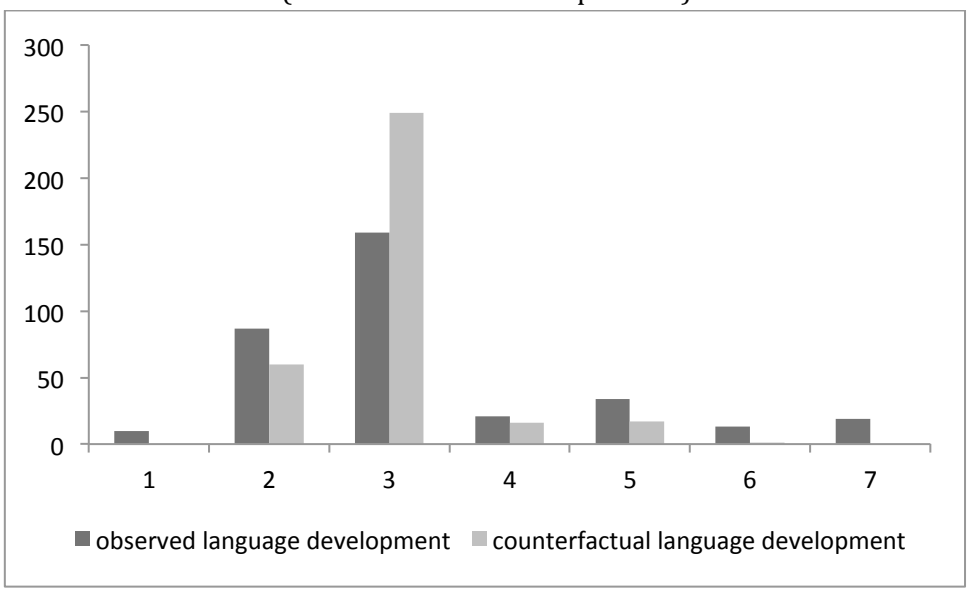
In order to indicate how much European influence affected the distribution of the status of languages, we apply the 1-Herfindahl index[16], which is often used in ethnolinguistic diversity measurement. In our case, 1-Herfindahl index can be interpreted as the probability that two randomly selected languages in our dataset have different status. The value of the 1-Herfindahl index related to the distribution of the current status of languages in Figure 4 is 0.702 (dark grey columns), and 0.515 in the case of the counterfactual situation (light grey columns). The difference between the two values suggests that the European influence increased the probability that two languages are of different status by about 20 percentage points. (The computed values related to Figure A.4a and A.4b are similar.)

Since language status determine the potential of its speakers for participating in the labor market and political decision making, our results are broadly related to the issue of individual well-being, economic development and inequality. What we have found can be interpreted that early contact with Europeans, the selective Bible translation strategies of missions and colonial policy contributed to higher inequality across groups through influencing their language development paths. The counterfactual analysis also suggests that if the current status of

[16] The Herfindahl-index (Herfindahl 1950) is computed as $\sum_{j=1}^{7} g_j^2$ where g_j is the share of languages within the sample with status j.

languages was dependent only on their traditional socio-economic characteristics, the gaps between the least and most linguistically developed groups would be smaller today.

Figure 4
The distribution of observed and counterfactual language status
(socio-economic development 1)



5. Conclusion

This paper aims to reveal the historical determinants of the current status of languages in Sub-Saharan Africa. The development of languages is measured with the Expanded Graded Intergenerational Disruption Scale (EGIDS) (Lewis and Simons (2010)) from Ethnologue (Lewis et al. 2014) which evaluates the vitality of languages according to five key concepts. The sources include the map by Murdock (1959) displaying the geographical location of indigenous societies in the 19th century, the Ethnographic Atlas (Murdock 1967) on the socio-economic characteristics of ethnic groups, the Roome map (1924) that contains the location of early missions, different sources on the year of the first Bible translation, the Joshua Project on the share of ethnic groups within country borders, the nationality of colonizer before independence, and the share of Islam within each group. The empirical models also control for geographical and climatic factors (distance to equator, coast and inland water, the mean and standard deviation of temperature, altitude and precipitation, and ecological diversity).

Although the Ethnographic Atlas is a widely utilized anthropological source in development studies, the way we apply this source is different from the traditional approaches in two aspects. First, most economic studies that show the persistence and the long term impact of past social traits on current development focus on variable 33 'jurisdictional hierarchy beyond the community level'. Using a general structural equation modeling technique, this paper utilizes

several variables related to economic and social organization from the Ethnographic Atlas to estimate the (assumed) underlying latent socio-economic development of indigenous societies, which is, unlike the original ordered variables, measured on a continuous scale. We estimate three latent socio-economic development variables with three different observed variables sets which, due to the high share of missing values in the Ethnographic Atlas, results in observation numbers of 343, 282, and 195. Second, we also argue that the Ethnographic Atlas does not purely provides information on the 'aboriginal' traits of African societies, but incorporates the effects of contacts with Europeans (earlier via trade and missions and later through colonial practices and policies) to some extent. Using an instrumental variable technique (two-stage least squares) we separate the effect of European influence measured with the intensity of missionary activities on the development of indigenous societies (estimated with the latent variable model).

When analyzing the main determinants of the current status of languages, we find that the socio-economic development of traditional societies, the age of Bible translation, the share of ethnic groups within the country are positively associated with the dependent variable. However, the number of missions per 100 km^2 and the colonizer dummies do not yield significant coefficients. Since the colonizer dummy is proved to be an important determinant of the intensity of missionary activity, we argue that colonial history has indirect effect on language status through determining the spread of Christian missions.

Comparing the distribution of the current status of languages with the counterfactual distribution if no European influence had taken place suggests that missionary activities and colonialism have a persistent linguistic effect that contributed to higher polarization in the language status distribution, thus higher socio-economic inequality among linguistic groups in Sub-Saharan Africa.

References

Acemoglu, D., Johnson, S., and Robinson, J. A., 2001. The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369-1401.

Adams, W. M. and Anderson, D. M., 1988. Irrigation before development: Indigenous and induced change in agricultural water management in East Africa. *African Affairs*, 87(349), 519-535.

Adegbija, E. E., 1994. Language attitudes in Sub-Saharan Africa: A sociolinguistic overview. Clevedon: Multilingual Matters.

Ajayi, J. F. A., 1965. Christian missions in Nigeria, 1841-1891: The making of a new elite. London: Longman.

Albaugh, E. A., 2014. State-building and multilingual education in Africa. New York: Cambridge University Press.

Angeles, L. and Neanidis, K. C., 2015. The persistent effect of colonialism on corruption. *Economica*, 82(326), 319-349.

Ansre, G., 1974. Language standardization in Sub-Saharan Africa. In: Fishman, J. A., ed., Advances in language planning. The Hague: Walter de Gruyter, 369-390.

Asafo, D. R., 1997. Social class conversion: Socioeconomic status of early Christian converts in Africa. *Nordic Journal of Africa Studies*, 6(1), 81-96.

Atkinson, D., 1987. The mother tongue in the classroom: a neglected source? *ELT Journal*, 41(4), 241-247.

Bandyopadhyay, S., and Green, E., 2012. Pre-colonial political centralization and contemporary development in Uganda. *Afrobarometer Working Paper No. 141*.

Bates, R., 1983. Essays on the political economy of rural Africa. Cambridge: Cambridge University Press.

Batibo, H. M., 2005. Language decline and death in Africa: Causes, consequences and challenges. Clevedon: Multilingual Matters.

Berman, E. H., 1974. African responses to Christian mission education. *African Studies Review*, 17(3), 527-540.

Bertocchi, G., and Canova, F., 2002. Did colonization matter for growth?: An empirical exploration into the historical causes of Africa's underdevelopment. *European Economic Review*, 46(10), 1851-1871.

Betts, R. F., 2005. Assimilation and association in French colonial theory, 1890-1914. Lincoln: University of Nebraska Press.

Blanton, R., Mason, D. T., and Athow, B., 2001. Colonial style and post-colonial ethnic conflict in Africa. *Journal of Peace Research*, 38(4), 473-491.

Bodenhorn, H., and Ruebeck, C. S., 2003. The economics of identity and the endogeneity of race. *NBER Working Paper No. 9962.*

Bokamba, E. G., 1991. French colonial language policies in Africa and their legacies. In: Marshall, D. F.,ed., Focus on Language Planning: Essays in honor of Joshua A. Fishman, vol. 3., Amsterdam: John Benjamins, 175-213.

Bokamba, E. G. and Tlou, J. S., 1977. The consequences of the language policies of African states vis-à-vis education. In: der Houssikian, H. and Kotey, P. A., eds., Language and linguistic problems in Africa. Columbia: Hornbeam, 35-53.

Brown, D. S., 2000. Democracy, colonization, and human capital in Sub-Saharan Africa. *Studies in Comparative International Development*, 35(1), 20-40.

Bunyi, G., 1999. Rethinking the place of African indigenous languages in African education. *International Journal of Educational Development*, 19(4-5), 337-350.

Cashdan, E., 2001. Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity. *American Anthropologist*, 103(4), 968-991.

Cheeseman, N., Ford, R., 2007. Ethnicity as political cleavage. Afrobarometer Conference on 'The micro-foundations of mass politics in Africa'. East Lansing, Vol. 12, No. 15.03.2007

Chimhundu, H., 1992. Early missionaries and the ethnolinguistic factor during the 'invention of tribalism' in Zimbabwe. *The Journal of African History*, 33(1), 87-109.

Cogneau, D., 2003. Colonisation, School and development in Africa. An empirical analysis. *DIAL Document de Travail 2003/I.*

Conklin, A. L., 1997. A mission to civilize: The republican idea of empire in France and West Africa, 1895-1930. Stanford: Stanford University Press.

Crowder, M., 1968. West Africa under colonial rule. London: Hutchinson and Co.

Doke, C. M., 1931. Report on the unification of the Shona dialects: carried out under the auspices of the Government of Southern Rhodesia and the Carnegie Corporation: presented to the Legislative Assembly, 1931, The Government of Southern Rhodesia

Easterly, W., and Levine, R., 1997. Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics*, 112(4), 1203-1250.

Egger, P., 2002. An econometric view on the estimation of gravity models and the calculation of trade potentials. *The World Economy*, 25(2), 297-312.

Englebert, P., Tarango, S., and Carter, M., 2002. Dismemberment and suffocation, a contribution to the debate on African boundaries. *Comparative Political Studies*, 35(10), 1093-1118.

Fabian, J., 1983. Missions and the colonization of African languages: Developments in the former Belgian Congo. *Canadian Journal of African Studies*, 17(2), 165-187.

Fenske, J., 2014. Ecology, trade, and states in pre-colonial Africa. *Journal of the European Economic Association*, 12(3), 612-640.

Fernihough, A. and O'Rourke, K. H., 2014. Coal and the European Industrial Revolution. *NBER Working Paper No. 19802*

Fishman, J. A., 1991. Reversing language shift. Clevedon: Multilingual Matters.

Frankema, E. H. P., 2012. The origins of formal education in sub-Saharan Africa: was the British rule more benign? *European Review of Economic History*, 16(4), 335-355.

Gallego, F. A., and Woodberry, R., 2010. Christian missions and education in former African colonies: How competition mattered. *Journal of African Economies*, 19(3), 294-329.

Garrido, S., 2011. Governing scarcity. Water markets, equity and efficiency in pre-1950s eastern Spain. *International Journal of the Commons*, 5(2), 513-534.

Gennaioli, N. and Rainer, I., 2007. The modern impact of precolonial centralization in Africa. *Journal of Economic Growth*, 12(3), 185-234.

Gilmour, R., 2007. Missionaries, colonialism and language in nineteenth-century South Africa. *History Compass*, 5(6), 1761-1777.

Gray, J. P., 1999. A corrected ethnographic atlas. *World Cultures*, 10(1), 24-85.

Green, E., 2013. Explaining African ethnic diversity, *International Political Science Review*, 34(3), 235-253.

Grenier, G., 1982. Language as human capital: Theoretical framework and application to Spanish-speaking Americans. PhD thesis, Princeton University

Grier, R. M., 1999. Colonial legacies and growth. *Public Choice*, 98(3-4), 317-335.

Groves, C. P., 1964. The Planting of Christianity in Africa: 1914-1954, vol. 4., London: Lutterworth Press.

Hailey, L., 1945. An African Survey. Oxford: Oxford University Press.

Harbert, W., McConnell-Ginet, S., Miller, A., and Whitman, J., 2009. Language and poverty. Clevedon: Multilingual Matters.

Hargreaves, J. D., 1985. The making of African boundaries: Focus on West Africa. In: Asiwaju, A. I., ed., Partitioned Africans, ethnic relations across Africa's International boundaries 1884-1984. New York: St. Martin's Press.

Harris, P., 1988. The roots of ethnicity: Discourse and the politics of language construction in South-East Africa. *African Affairs*, 87(346), 25-52.

Heine, B., 1970. Status and use of African lingua francas. Afrika-Studien No. 49. Muenchen: Weltforum-Verslag GmbH.

Henderson, M. and Whately, W. C., 2014. Pacification and gender in colonial Africa: Evidence from the Ethnographic Atlas. MPRA Working Paper No. 61203. Accessed from: http://mpra.ub.uni-muenchen.de/61203/1/MPRA_paper_61203.pdf [09 March 2015]

Herfindahl, O. C., 1950. Concentration in the steel industry. Thesis (PhD), Columbia University.

Hijmans, R.J., Cameron, S.E., Parra, J. L., Jones, P. G., and Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965-1978.
Johnson, H. B., 1967. The location of Christian missions in Africa. *Geographical Review*, 57(2), 168-202.
Joshua Project, http://joshuaproject.net/ [14 March 2015]
Kaufman, E., 2015. Land, history or modernization? Explaining ethnic fractionalization. *Ethnic and Racial Studies*, 38(2), 193-210.
Konczacki, Z. A., and Konczacki, J. M., 1977. An economic history of tropical Africa, vol. 1. Abingdon: Frank Cass and Co.
Levinsohn, J., 2007. Globalization and the returns to speaking English in South Africa. In: A. Harrison, ed., Globalization and poverty. Chicago: University of Chicago Press, 629-646.
Lewis, M. P., and Simons, G. F. (2010). Assessing Endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, 55(2), 103-120.
Lewis, M. P., Simons, G. F. and Fenning, C. D. (eds.), 2014. Ethnologue: Languages of the world, 17th edition, SIL International, Dallas, Texas. Online version: http://www.ethnologue.com/17
Mansour, G., 1993. Multilingualism and nation building, Clevedon: Multilingual Matters.
Martinez-Zarzoso, I., 2003. Gravity model: An application to trade between regional blocs. *Atlantic Economic Journal*, 31(2), 174-187.
Mauro, P., 1995. Corruption and Growth. *The Quarterly Journal of Economics*, 110(3), 681-712.
Mesthrie, R., Swann, J, Deumert, A. and Leap, W. L., 2009. Introducing Sociolinguistics, 2nd edition. Edinburgh: Edinburgh University Press.
Michalopoulos, S., 2012. The origins of ethnolinguistic diversity. *American Economic Review*, 102(4), 1508-1539.
Michalopoulos, S., and Papaioannou, E., 2013. Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, 81(1), 113-152.
Michalopoulos, S., and Papaioannou, E., 2015. Further evidence on the link between pre-colonial political centralization and comparative economic development in Africa. *Economics Letters*, 126(1), 57-62.
Middleton, J., 1971. Some effects of colonial rule among the Lugbara. In: Turner, V., ed., Colonialism in Africa 1870-1960. Vol. 3.: Profiles and change: African society and colonial rule. Cambridge: Cambridge University Press, 6-48.
Moseley, C., ed., 2010. Atlas of the world's languages in danger, 3rd edition, Paris: UNESCO. Online version: http://www.unesco.org/culture/en/endangeredlanguages/atlas
Murdock, G. P., 1959. Africa: Its people and their cultural history. New York: McGraw-Hill.
Murdock, G. P., 1967. Ethnographic Atlas. Pittsburgh: University of Pittsburgh Press.
Nettle, D., 1999. Linguistic diversity. Oxford: Oxford University Press.
Nunn, N., 2008. The long-term effects of Africa's slave trade. *The Quarterly Journal of Economics*, 123(1), 139-176.
Nunn, N. and Qian, N., 2011. The potato's contribution to population and urbanization: Evidence from a historical experiment. *The Quarterly Journal of Economics*, 126(2), 593-650.
Oliver, R. 1952. The missionary factor in East Africa. London: Longmans.
Ostrom, E., 2000. Collective action and the evolution of social norms. *Journal of Economic Perspectives*. 14(3), 137-158.
Pendakur, K., and Pendakur, R., 2002. Speaking in tongues: Language as both human capital and ethnicity. *International Migration Review*, 36(1), 147-178.
Peters, P. E., 1994. Dividing the commons: Politics, policy, and culture in Botswana. Charlottesville: University Press of Virginia.
Peterson, D., 1997. Colonizing language? Missionaries and Gikuyu dictionaries, 1904 and 1914. *History in Africa*, 24, 257-272.
Posner, D., 2004. The political salience of cultural difference: why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review*, 98(4), 529-545.

Romaine, S., 2009. Biodiversity, linguistic diversity and poverty: Some global patterns and missing links. In: Harbert, W., McConnell-Ginet, S., Miller, A., and Whitman, J., 2009. Language and poverty. Clevedon: Multilingual Matters, 127-
Roome (1924). Ethnographic survey of Africa, map
Smith, E. A., 2010. Communication and collective action: language and the evolution of human cooperation. *Evolution and Human Behavior*, 31(4), 231-245.
Spencer, J., 1974. Colonial language policies and their legacies in Sub-Saharan Africa. In: Fishman, J. A., ed., Advances in language planning, 163-175. The Hague: Walter de Gruyter.
Steinhauer, H., 2005. Colonial history and language policy in Insular Southeast Asia and Madagascar. In: Adelaar, K., and Himmelmann, N., eds., The Austronesian languages of Asia and Madagascar. Abingdon: Routledge.
Stock, R., 2012. Africa South of the Sahara. A geographical interpretation, 3rd ed. New York: The Guilford Press.
Transparency International, 2014. Corruption Perception Index 2014. Accessed from: http://www.transparency.org/cpi2014 [14 March 2015]
UNESCO, 2003. Language vitality and endangerment. International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages
United Nations, 2014. Human Development Indicators. Accessed from: http://hdr.undp.org/en/data [14 March 2015]
van den Bersselaar, D., 1997. Creating 'union Ibo': Missionaries and the Igbo language. *Africa*, 67(2), 273-295.
Ward, I., 1940. Foreword. In: Benzies, D. Learning our language. London: Longmans.
Warren-Rothlin, A., 2009. Script choice, politics, and Bible agencies in West Africa, *Bible Translator*, 60(1), 50-66.
Welbourn, F. B. 1971. Missionary stimulus and African responses. In: Turner, V., ed., Colonialism in Africa 1870-1960. Vol. 3.: Profiles and change: African society and colonial rule. Cambridge: Cambridge University Press, 310-345.
Welmers, W. E., 1974. Christian missions and language policies in Africa. In. Fishman, J. A. (ed.). Advances in language planning, 191-205. The Hague: Walter de Gruyter.
White, F., 1983. The vegetation of Africa: A descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. *Natural Resource Research*, 20, 1-356.
White, B., 1996. Talk about school: Education and the colonial project in French and British Africa, (1860-1969). *Comparative Education*, 32(1), 9-25.
Whiteley, W. H., 1956. The changing position of Swahili in East Africa. *Africa*, 26(4), 343-353.
Wietzke, F.-B., 2015. Long-term consequences of colonial institutions and human capital investments: Sub-national evidence from Madagascar. *World Development*, 66, 293-307.
World Bank, 2013. GDP per capita (current USD). World Development Indicators. Accessed from: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD [23 Feb 2015]
worldbibles.org, http://worldbibles.org [Accessed: 04 05 2015]

Appendix

Table A.1

The description of the dependent variable and the distribution of sample languages per status

original EGIDS scale	status	explanation	number of languages (total: 389)	percent	number of African languages in Ethnologue (total: 2735)	percent	recoded value for the empirical analysis (reversed EGIDS)
0	international	The language is widely used between nations in trade, knowledge exchange, and international policy.	0	0	0	0	-
1	national	The language is used in education, work, mass media, and government at the national level.	19	4.88	72	2.63	7
2	provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.	14	3.6	17	0.6	6
3	wider communication	The language is used in work and mass media without official status to transcend language difference across a region.	36	9.25	109	3.99	5
4	educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.	26	6.68	89	3.25	4
5	developing (written)	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.	182	46.79	756	27.64	3
6a	vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.	99	25.45	1223	44.72	2
6b	threatened	The language is used for face-to-face communication within all generations, but it is losing users.	10	3.34 (share of the combined category)	199	7.28	1
7	shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.	2		86	3.14	1
8a	moribund	The only remaining active users of the language are members of the grandparent generation and older.	1		58	2.12	1
8b	nearly extinct	The only remaining users of the language are	0	0	66	2.41	-

		members of the grandparent generation or older who have little opportunity to use the language.					
9	dormant	The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.	0	0	24	0.88	-
10	extinct	The language is no longer used and no one retains a sense of ethnic identity associated with the language.	0	0	36	1.32	-

Note: since our analysis is concerned with explaining language status instead of language endangerment we reversed the scale so that higher values refer to higher status. Thus, the interpretation of the coefficients in the empirical analysis is more straightforward.

Table A.2

Variables used for the estimation of the socio-economic development of indigenous societies

variable name	variable number in the Corrected Ethnographic Atlas	#	original coding	recoding	remark
intensity of agriculture	var 28	389	no agriculture (1), casual agriculture (2), extensive or shifting agriculture (3), intensive agriculture, using fertilization, crop rotation, or other techniques to shorten or eliminate fallow period (5), intensive irrigated agriculture (6)	1 and 2 -> 0 3 -> 1 5 and 6 -> 2	Our sample does not include category 4 (horticulture, vegetal gardens or groves, fruit trees) of v28. The aim of the recoding strategy is to distinguish between extensive and intensive agriculture, assuming that intensive agriculture indicates higher social and economic development.
jurisdictional hierarchy beyond community level	var 32 (only the second digit)	386	no levels (1), one level (2), two levels (3), there levels (4), four levels (5)	1 -> 0 2 -> 1 3 -> 2 4 and 5 -> 3	Category 4 and 5 are collapsed since there is only one observation in category 5.
class stratification	var 67	354	absence among freemen (1), wealth distinctions (2), elite (based on control of land and other resources) (3), dual (hereditary aristocracy) (4), complex (social classes) (5)	1 -> 0 2 -> 1 3 -> 2 4 -> 3 5 -> 4	
succession of the office of local headmen	var 73	319	patrilineal heir (1), matrilineal heir (2), appointment by higher authority (non-hereditary) (3), seniority or age (non-hereditary) (4), influence, wealth or social status (non-hereditary) (5), election or other formal consensus (non-hereditary) (6), informal consensus (non-hereditary) (7), absence of any such office (9)	9 -> 0 1 and 2 -> 1 3, 4, 5, 6, 7 -> 2	The recoding strategy is aimed to distinguish between hereditary and non-hereditary systems. Societies with no such office are considered as the reference category.
high god	v34	263	absent or not reported (1), not active in human affairs (2), active in human affairs but not supportive in human morality (3), supportive of human morality (4)	1->0 2->1 3-> 2 4->3	

Table A.3

Correlation coefficients between the different estimated socio-economic development variables

	dev 1	dev 2	dev 3
dev 1	1		
dev 2	0.990 (194)	1	
dev3	0.983 (136)	0.993 (136)	1

Note: the number of observations is reported in parentheses. All coefficients are significant at 1%.

Table A.4

Spearman rank correlation coefficients between the estimated latent socio-economic development variables and their components

	dev 1	dev 2	dev 3
agricultural intensity	0.301	0.332	0.390
jurisdictional hierarchy (beyond community level)	0.904	0.847	0.850
class stratification	0.863	0.908	0.915
authority succession	-	-	0.225
high god	-	-	0.324
Number of obs	237	194	136

Note: All coefficients are significant at 1%.

Table A.5

The ten most developed indigenous societies with and without European influence according to the three indigenous socio-economic development measures

order	socio-economic development1	socio-economic development1 without European influence	socio-economic development2	socio-economic development2 without European influence	socio-economic development3	socio-economic development3 without European influence
1	Rundi	Amhara	Amhara	Amhara	Amhara	Amhara
2	Ruanda	Songhai	Rundi	Songhai	Songhai	Songhai
3	Amhara	Rundi	Ruanda	Rundi	Rundi	Rundi
4	Songhai	Oyo Yoruba	Songhai	Oyo Yoruba	Ruanda	Tigrinya
5	Oyo Yoruba	Ruanda	Oyo Yoruba	Tigrinya	Tigrinya	Kanuri
6	Hunde	Kanuri	Tigrinya	Kanuri	Oyo Yoruba	Oyo Yoruba
7	Tigrinya	Fur	Nupe	Ruanda	Nupe	Fur
8	Lozi	Janjero (Yemsa)	Ganda	Fur	Kanuri	Janjero (Yemsa)
9	Sotho	Kafa	Kanuri	Janjero (Yemsa)	Wolof	Ruanda
10	Nyoro	Sotho	Merina (dialect of Malagasy Plateau)	Sotho	Fur	Nupe

Table A.6

Variable description

variable name	description	variable type	source
language status	The status of each language as understood by the EGIDS (Lewis and Simons 2010). The original and recoded values are presented in Table A.1. For more information consult http://www.ethnologue.com/about/language-status	categorical (ordered)	Ethnologue (Lewis et al. 2014)
British, French, Belgian, Portuguese	It labels ethnic groups that were assigned to countries ruled by the British, French, Belgian and Portuguese. If a group was partitioned between countries, it is labeled in each country accordingly.	dummy/binary	Bertocchi and Canova (2002)
socio-economic development of indigenous societies	Estimated with GSEM techniques from variables in Table 2.A as shown in Table A.4 described in Section 3.	continuous	Ethnographic Atlas (Murdock 1967)
ln(pop share)	The share of each ethnic group within country population.	continuous between 0 and 1	Joshua Project
bible age	The number of years between the year that the Bible was translated to a certain language and 2014 (present).	continuous with integer values	Groves (1964), Ethnologue (2014), worldbibles.com
missions per 100km^2	The number of missionary locations on the territory of each ethnic group presented in Murdock (1959) map.	continuous	Roome (1924), Nunn (2008 and 2010)
islam	The share of each ethnic group affiliated with the Islam religion.	continuous between 0 and 1	Joshua Project
ln(distance to equator)	The distance of the center of each ethnic group to the Equator (in km).	continuous	The shapefile (ESRI data format) containing the Equator is downloaded from the Natural Earth website. http://www.naturalearthdata.com/features/
ln(distance to coast)	The distance of the center of each ethnic group to the coast (in km).	continuous	The shapefile of African coastline is downloaded from the website of the African Marine Atlas. http://omap.africanmarineatlas.org/BASE/pages/coastline.htm
ln(distance to inland water)	The distance of the center of each ethnic group to the closest river or lake (in km).	continuous	Shapefile with rivers is downloaded from http://www.arcgis.com/home/item.html?id=fedf8e234b614ecaac65893f807344f5 The shapefile with lakes is downloaded from
mean temperature	The mean of monthly temperature on the area of each ethnic group between 1950 and 2000 (in 0.1 Celsius).	continuous	Raster data (ESRI grids format, 10 arc-minutes resolution) are from the WorldClim website. For the description of applied
std of	The standard deviation of		

temperature	monthly temperature within the territory of each group (in 0.1 Celsius).		methods see http://www.worldclim.org/methods and Hijmans, R. J. et al. (2005)
mean precipitation	Mean monthly precipitation on the territory of each ethnic group (in mm).		
std of precipitation	The standard deviation of mean precipitation within the territory of each ethnic group (in mm).		
mean altitude	The mean elevation of each ethnic group above sea level (in m).		
std of altitude	The standard deviation of elevation within the territory of each ethnic group (in m).		
ecological diversity	This variable show the probability that two geographical points selected at random within the territory of each ethnic group belong to different ecological zones defined in White (1983).	continuous between 0 and 1	The vegetation map is published by White (1983). The shapefile is available at James Fenske's website https://sites.google.com/site/jamesfenske/data

Note: original and recoded values are understood and presented in the case of categorical variables. Original values column shows only values that are present in our sample. For information on other values consult the Corrected Ethnographic Atlas (Gray 1999).

Figure A.1
The location of indigenous ethnic groups (Murdock 1959)

Figure A.2
The location of early missions (Roome 1924)

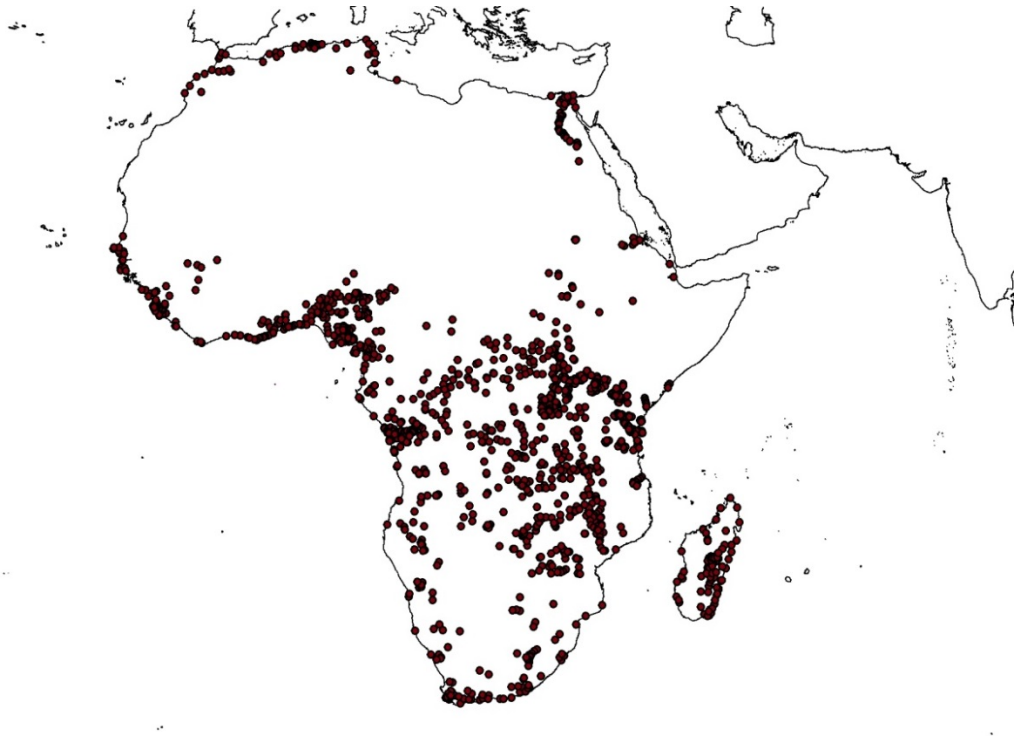


Figure A.3a

Kernel density estimates of the distribution of socio-economic development with and without European influence (development variable: socio-economic development 1)

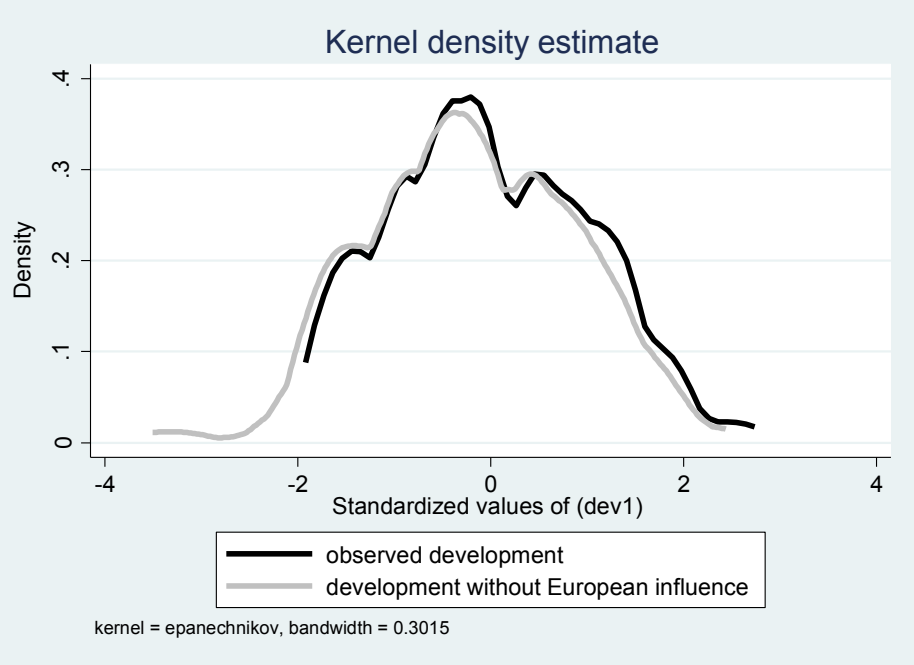


Figure A.3b

Kernel density estimates of the distribution of socio-economic development with and without European influence (development variable: socio-economic development 2)

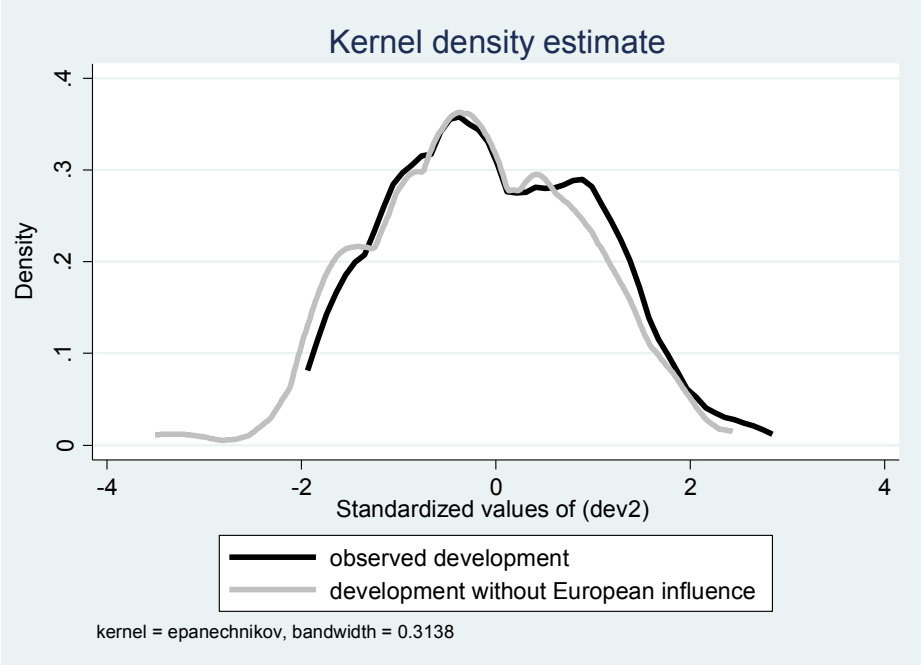


Figure A.3c

Kernel density estimates of the distribution of socio-economic development with and without European influence (development variable: socio-economic development 3)

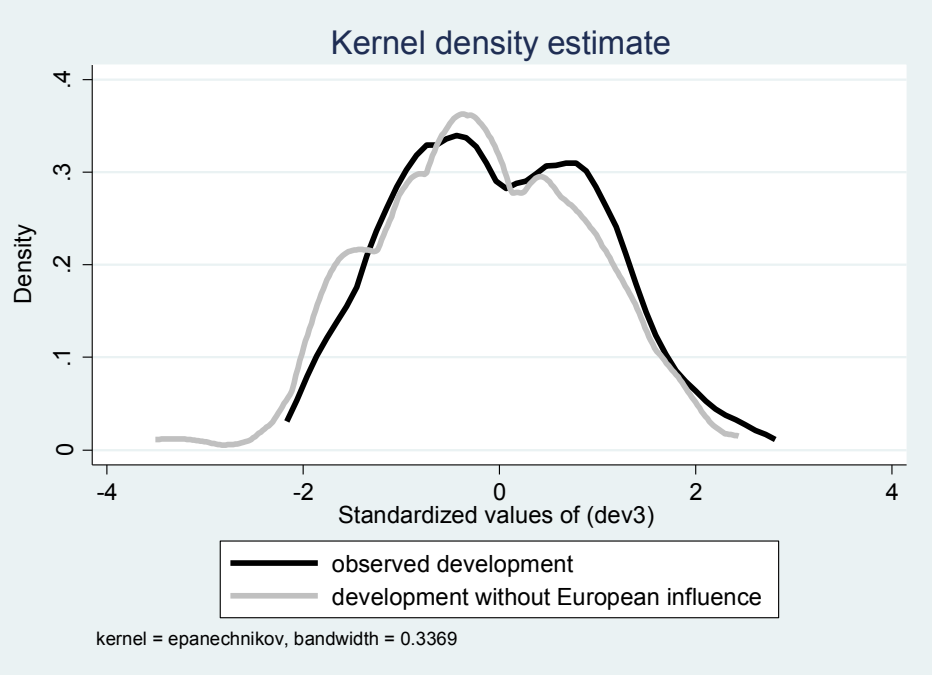


Figure A.4a

The distribution of the observed and counterfactual language status (socio-economic development 2)

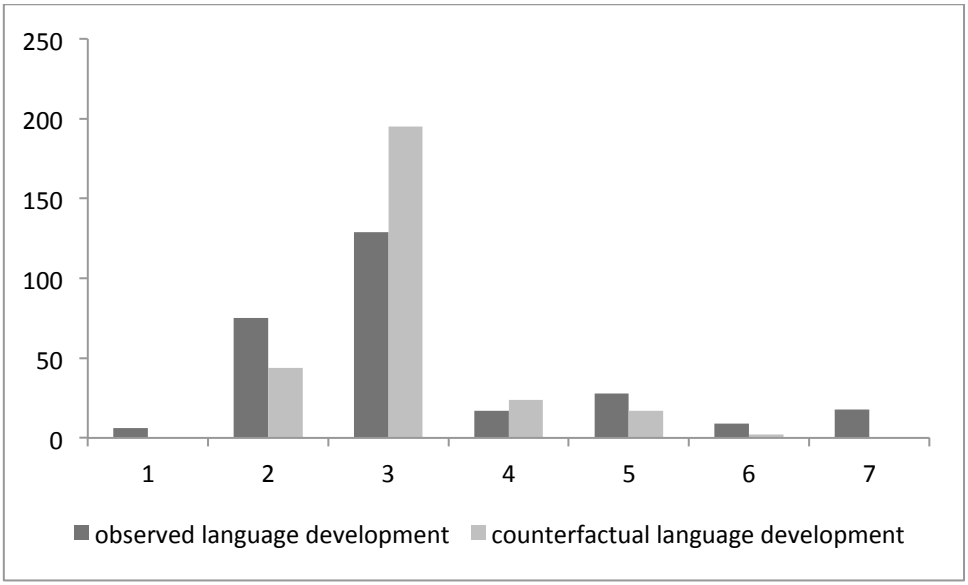


Figure A.4b

The distribution of the observed and counterfactual language status (socio-economic development 3)