

Wikipedias in African languages 2017

An evaluation based on Wikimedia statistics

Ingo Koll a.k.a. mtumiaji:kipala on sw.wikipedia.org

Greetings from swwiki

Salamu kutoka Wikipedia ya Kiswahili

- Swwiki has been developed by many, in the core there is a team of 5 volunteers who have remained since 10 years
- After wikiindaba 2014 we had a first workshop in Tanzania, thanks to WMF support
- First time our longtime editors could meet
- Important for our motivation to continue
- Phantastic response from teachers and students at Morogoro
- Several 100 edits out of it, some continue



Impression from the Morogoro workshop

22/03/2015

Africa – the continent with many languages

2000 – 3000 languages (of 5-7000 worldwide)

But they are threatened

50-90 % estimated to die out until 2100

Any point to start a wikipedia with a “small language” ?

12-15 languages understood by 80% of all Africans in interethnic communication

Concentrate efforts here ?

African Languages are important

- The education systems in foreign languages limit severely the reach of education and spread of knowledge in society
- Children start their way into learning best in mother tongue (first language)
- Education systems have historically worked with external languages with fair results but limited reach (Latin in Europe, Arabic in Middle East, etc) – but education for all and development came after switch to local languages
- Countries in Africa have to develop their own languages , but some will not manage because of too much linguistic diversity
- Wikipedia and open media is among the major chances towards building a knowledge resource in African languages – (under some conditions)

Wikipedia: 295 languages

ca. 40 African language wikipedias

A look at all wikipedias by size

Huge Wikipedias (more than 1 Mio articles)	> 13
<i>Large</i> Wikipedias (100,000 to 1 Mio articles)	> 45
Medium size Wikipedias (10,000 to less than 100k)	> 77
Small Wikipedias (1,000 to less than 10 k articles)	>112
Tiny Wikipedias (100 to less than 1,000 articles)	> 38
<i>“Micropedias”</i> (less than 100 articles)	> 10

Wikipedias in African languages 2017:

„African languages“ – languages native to Africa = including Afrikaans + Masri
= excluding English, French, Arabic

~~Huge Wikipedias (1mio+)~~ ~~Large Wikipedias (100k+)~~ **None**

Medium size Wikipedias (10k +) **6**

Small Wikipedias (1k+) **10 (+3)**

Tiny Wikipedias (100+) **16 (-7)**

Closed Micropedias **4**

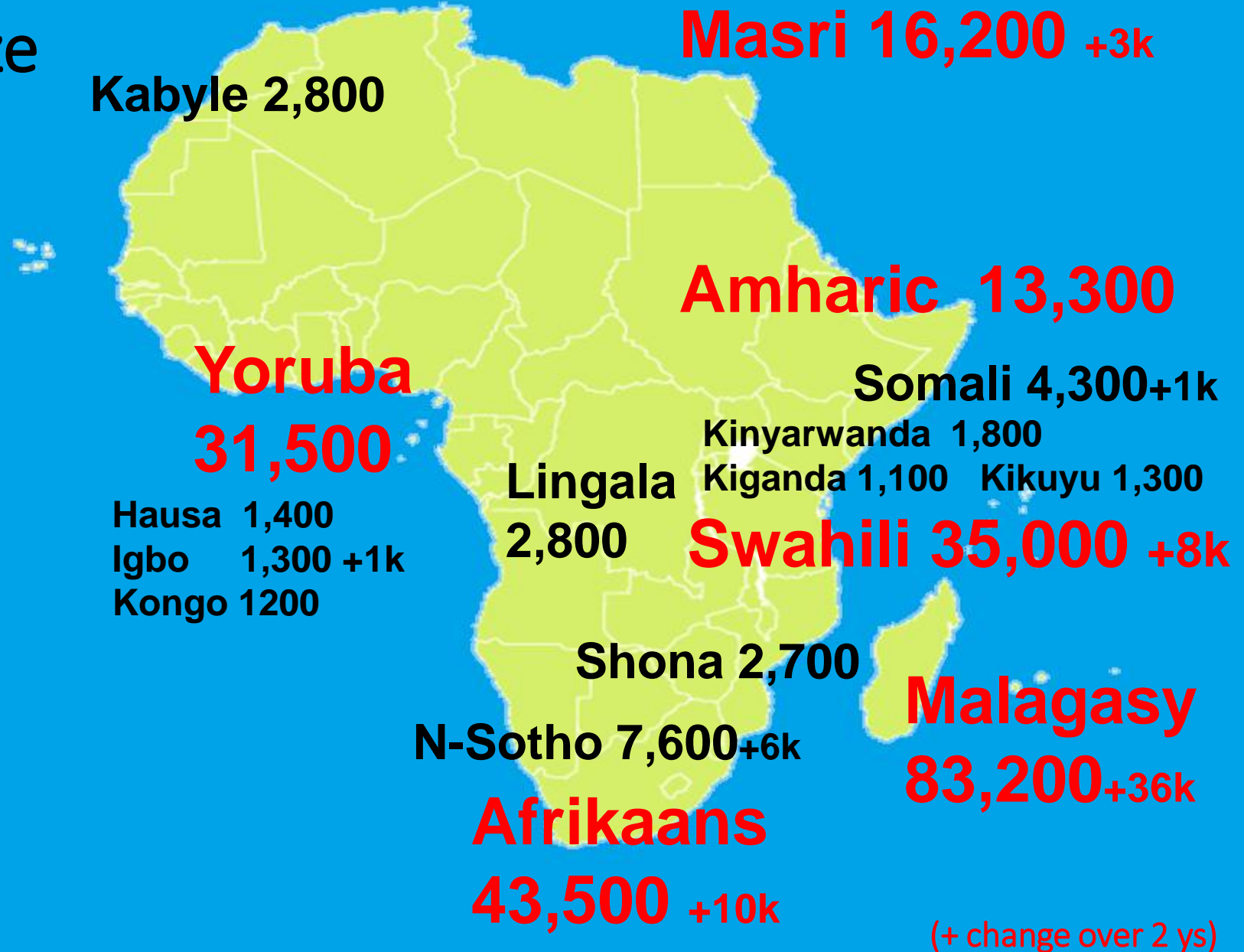
(change over 2 ys)

The African Language Wikipedias

Medium & small size

**Minipedias under
1,000 articles**

Zulu
Oromo
Tswana
Xhosa
Twi
Tumbuka
Kirundi
Swati
Bambara
Tsonga
Sesotho
Chichewa
Ewe
Akan
Venda
Tigrinya



Large community
Large numbers Pageviews: readers

Example
Egyptian Arabic **Masri**
76 Mio speakers
±2.5 Mio in Dec 2016

Smaller communities
Lower numbers

Afrikaans
12 Mio speakers
1.4 Mio views in
Dec 2017



not necessarily
viewed by many
in the community
**Masri views in
Egypt 1.7% of
all wikipedia
views there**

**Afrikaans views
reach more
than 10% in the
language
community**

Abolute pageviews in ALWs (December 2016)

All ALWs have very small view numbers compared to the huge Wikipedias like English (9,868 Mio) or French (741Mio)

Egyptian Masri 2.5Mio

Afrikaans 1.4Mio

Swahili 1.1Mio

Somali 448 k

Amharic 446 k

Yoruba 299 k

Malagasy 283 k

ALWs shares per country Dec 2016	Number	% of all Wikipedia views	
Egypt			
Egyptian Arabic Masri	709.000	1.7%	
South Africa			
Afrikaans Wp	334.000	0.9%	(but +10% amongst speakers)
Tanzania			
Swahili Wp	417.000	6.8%	(8-9% in 2016)
Kenya			
Swahili Wp	136.000	0.9%	(around 1% in 2016)
Nigeria			
Igbo Wp	363.000	1.2%	
Yoruba	??	??	not visible
Ethiopia			
Amharic Wp	108.000	2.6%	
Somalia			
Somali Wp	91.000	9.1%	small numbers ?

Assessing relevance / usefulness of an ALW market shares in “home countries”
 (Percent of all wikipedia views in the country)

Competitiveness: Swahili compared to English on local topics

Looking at topics where we can assume few international views in English and that requests on enwiki mostly come from Tanzania. Likely assume that swwiki views come from Tanzania. Swwiki is surprisingly competitive in local views!

Pageviews for selected articles on localities in Tanzania from Jan 2016-Jan 2017

Mbozi Meteorite	en 3.014	sw 3.369
Kigoma Region ·	en 11.334	sw 8.924
Mbozi District ·	en 2.177	sw 1.336
Kakonko District	en 705	sw 586

(via <https://tools.wmflabs.org/pageviews>) Usually the relation between pageviews is 1 swwiki : 11,000 enwiki

Results of “market share”view

- Most wikipedia views in Africa are in English, French, etc
- most ALWs have no traceable views in home country (under 0.1%)
- Somali 9% and Swahili 7-9% have strongest national shares of all ALWs (Somali with small numbers overall)
- Afrikaans has small national but 10% community share!
- Large Malagasy has only 8000 = 0.3% viewers at home
- **Surprise Nigeria:** Yoruba (31,500 art) is not traceable in Nigeria, but Igbo (1,300 art) has 1% of Nigerian views
- Tanzanians open swwiki only a bit less than enwiki on local topics (if article exists already in swwiki; other topics can not be verified)

Reasons for stronger / weaker viewership?

- Probably quality!
- Malagasy & Yoruba have a large share of bot created articles (96% in Malagasy)
- Yoruba has extremely short articles (sometimes not even a sentence), Malagasy just many lists
- Yoruba may have a special problem with its orthography reducing the views via search machines
- Afrikaans and Swahili have followed recommended topics like “1,000” and “10,000 articles every Wikipedia should have”, much local content, bringing views via search machines
- both have hardly any bot-created articles

Community Interest: Views from home vs. abroad

Pageviews in ALWs, coming thru IPs from these countries

Igbo (ig)

Nigeria	85.8%
United States	6.1%
France	3.8 %
U.K.	3.5%

Swahili (sw)

Tanzania	40.4%
United States	16.4%
Kenya	13.2%
France	11.6%
Ireland	7.4%
Germany	4.6%

Afrikaans (af)

South Africa	27.0%
United States	22.8%
France	18.5%
Germany	9.4%
Canada	4.9%
Netherlands	3.9%
Other	13.5%

Somali (so)

United States	33.8%
Somalia	29.3%
Germany	17.4%
France	12.9%
Other	6.6%

Yoruba (yo)

France	40.9%
United States	34.5%
Germany	13.5%
Canada	5.6%
Other	5.5%
Other	7.6%

Malagasy (mg)

France	43.1%
United States	29.8%
Germany	14.2%
Canada	3.7%
Madagascar	3.7%
Other	5.5%

Weakness & strength: teams

Wikipedia grew as teamwork

Many ALWs have no or very small teams

In December / October 2016 there were active Wikipedians (who edited 5 times or more):

- In Malagasy **1 / 0**
- In Afrikaans **25 / 12**
- in Swahili **10 / 3**
- In Yoruba **5 / 0**
- In Masri **8 / 3**
- in Amharic **1 / 1**

Observations and assumptions

Perspectives for ALWs

- potential is ahead, internet access grows in Africa.
- Igbo ALW views show strong interest in some communities (measurable views although there is little content)
- The ALWs without team have little perspective
- Quality strategy is more important than quantity – widespread topics bring viewers via searchmachines only if the ALW has topics to offer
- Local content in local language attract views

Challenges for ALWs

- The near monopoly of European languages as medium of higher education
- Weaknesses of most educated Africans to discuss scientific etc. topics in mother tongue & “national language”
- Little faith among Africans in ability of African Languages to express a number of issues
- Reduced ability to read content in ALWs on topics people know through education in foreign languages

Closing remark

- An ALW has good chances for development if its language is medium of instruction at least at primary school level
- It is not necessary for its development to use a „large“ language (cf the example of Afrikaans)
- A wikipedia in a language which is not medium of instruction at school will likely get stuck on a lower level