IRIE
International Review of Information Ethics Vol. 32 (11/2022)

Authors: Daniel Pimienta, Gilvan Müller de Oliveira

Cyber-geography of languages. Part 2: the demographic factor and the growth of Asian languages and Arabic

Abstract:

Part 1 of this study explains the methodology, sources, biases, and results of the recent study on the presence on the Internet of the 330 languages with more than one million L1 speakers, and the reality of the place of English on the Internet. This part analyzes the results in terms of cyber-geography for the other languages. It appears that the languages of Europe, especially English, are still dominating the Internet but that the languages of Asia and the Arabic world are in a strong progression and will take the lead, in terms of connected speakers. The languages of Africa are suffering from the difficulty of many African countries to overcome the digital divide: however, demographic long-term future clearly will favor in the long run African languages and the European languages with a notable presence in Africa (by order of importance: English, French and Portuguese).

Keywords: Bias, Cultural Diversity, Cyber-geographic, Demolinguistic, Disinformation, Linguistic Diversity, Multilingualism, Cyber-Geography of Languages

Agenda:

Authors:

Daniel Pimienta

- Observatory of Linguistic & Cultural Diversity on the Internet, Email: pimienta@funredes.org, Website: funredes.org/lc

Gilvan Müller de Oliveira

- UNESCO Chair on Language Policies for Multilingualism, Federal University of Santa Catarina (UFSC), Brazil, Email: gimioliz@gmail.com, Website: https://www.unescochairlpm.org/uclpm/

ISSN 1614-1687

Introduction

The analysis of the results (see Pimienta, Müller de Oliveira, 2022) of the Observatory of linguistic and cultural diversity on the Internet[1] are considered in terms of cyber-geography.

Analysis

What do the 2021 study results (Pimienta, 2021) tell us compared to 2017 and earlier[2]? Let us look at the most powerful languages first using the bias-corrected data from the Observatory.

Power ranking

LANGUAGE	POWER
English	25,0%
Chinese	15,0%
Spanish	7,0%
French	3,5%
Hindi	3,5%
Portuguese	3,0%
Russian	3,0%
Arabic	2,5%
German	2,5%
Japanese	2,5%
Malay	1,8%
Italian	1,4%
Turkish	1,2%
Korean	1,2%
Bengali	1,2%
Vietnamese	0,7%
REMAIN	25,0%
TOTAL	100%

The three factors that will determine future evolution are, in order of importance: demographics, overcoming the digital divide, and the ability to create content. Demographics favor Hindi, which is likely to quickly overtake French and probably Spanish in the medium term. Demographics will end up favoring Arabic over other nearby languages, including Portuguese, which could, in turn, strengthen its position against Russian.

We can see that the center of gravity of the Internet is moving rapidly from the western world, where it was born and flourished in its early stages, towards Asian languages and Arabic. Demographics should favor the languages of the African continent in the long term, and indirectly the European languages with a strong presence in African (English, French and Portuguese), but the African digital divide is still very marked and slower to narrow compared to the global growth of the Internet. This table, made with the latest results from 330 languages, clearly presents this situation:

[1] https://funredes.org/lc

[2] The Observatory has conducted measurement campaigns, with another methodology, between 1998 and 2007, being obliged to interrupt them due to that the evolution of search engines (losing reliability as a scientific tool) has made the method obsolete. The results are still available at https://funredes.org/lc.

Cyber-Geography of language families

	African languages	Languages of the Americas	Arabic as a macro-language	Asian languages	Languages of Europe	Rest of languages
Internaut %	28.6%	59.7%	60.2%	46.6%	81.1%	54.2%
Power	2.55%	0.19%	3.11%	35.71%	54.93%	3.47%
Capacity	0.24	0.63	0.88	0.57	1.61	0.45
Gradient	0.45	0.59	0.80	0.65	1.07	0.55
Population L1 + L2	9.15%	0.31%	3.53%	48.21%	30.91%	7.81%
Connected population	5.18%	0.32%	3.89%	44.60%	39.51%	6.45%
Number of languages with L1> 1M	**138**	**8**	**1**	**135**	**47**	**0**

There are 138 African languages covered in the study; On average, 28.6% of its speakers are connected to them; the group represents 9.15% of the world's total of L1+L2 speakers. However, together they only represent 2.55% of the total weight of the Internet and 5.18% of the L1+L2 connected population.

The European colonizing process in the world after the year 1500 means that almost no languages or only very few originating from the Americas (8 languages) or Oceania (0 languages) are in the group with more than 1 million speakers. On the other hand, the colonizing process was less lasting and more superficial in Asia and Africa, which results in this significant number of populations speaking Asian and African languages, as shown in the table, and with greater possibilities for growth.

Languages of European origin continue to lead above average on the Internet. Still, the recent push of Asian languages and Arabic place them in a better position, with a large room for progress in terms of connected people, and their weight on the Internet is growing rapidly. Remembering that the measuring instrument is significantly biased against Asian languages, it is likely that the difference in terms of power is considerably less than indicated. In any case, as the connectivity rate of Asian countries increases and approaches the exceptional average rate of more than 80% of European languages, they will also gain first place in terms of power (the main macro-indicator produced by the model, see (Pimienta, Müller de Oliveira, 2022).

Let us now look at the languages that follow English. Chinese is in the second position in terms of power and content, but it is already the first language in terms of people connected to the Net and, unlike Western countries, where many are above 90%, there is room for progress. The following table shows data whose biases are minimal and are obtained from ITU data of people connected to the Internet and Ethnologue demo-linguistic data, weighing the former with the latter[3]. The presence of Asian languages and Arabic (marked in yellow) is remarkable.

Top languages in terms of percentage of Internet users

[3] And again, focused on L1 + L2.

	WORLD % INTERNAUTS	WORLD % SPEAKERS	% SPEAKERS CONNECTED
Chinese	17.5%	14.6%	65.59%
English	15.2%	12.9%	64.35%
Spanish	7.0%	5.2%	73.04%
Hindi	4.2%	5.8%	40.18%
Arab	3.9%	3.5%	60.25%
Russian	3.5%	2.5%	77.21%
Portuguese	3.0%	2.5%	66.96%
French	3.0%	2.6%	63.33%
German	2.2%	1.4%	87.68%
Malay	2.2%	23%	51.01%
Japanese	2.0%	1.2%	92.62%
Turkish	1.3%	0.9%	77.95%
Bengali	1.1%	2.6%	24.16%
Urdu	1.0%	2.2%	24.13%
Persian	0.9%	0.8%	63.99%
Vietnamese	0.9%	0.7%	69.00%
Korean	0.9%	0.8%	64.73%
Italian	0.9%	0.7%	75.66%

Today, the reality is that the center of gravity of the Internet is rapidly shifting from European languages to Asian languages and Arabic as many European languages in Europe and North America are approaching maximum levels of connectivity rate, where a strong progression is no longer possible. African languages are lagging in this development for now.

In the conception of (Calvet, 2002), languages are 'markets. It is to be expected that there will be growing pressure to overcome the deficit of connectivity and production of content for African languages of large populations, which will not be the case for lesser-spoken languages, deepening the inequality between the two language groups.

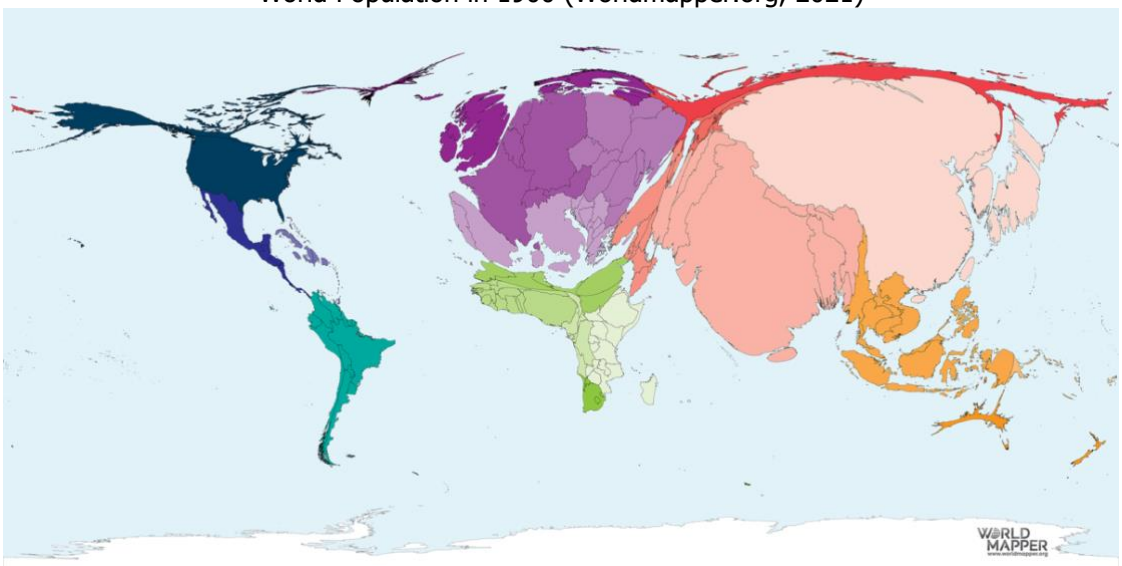
We will demonstrate below that this shift to a predominance of Asian and African languages on the Internet (provided the access gap is corrected) is not faster because of the rapid growth of European language speakers present in Asia, such as English, but especially in Africa, like English, French and Portuguese.

The macro-indicator power, by definition, as we know, favors languages with the highest number of speakers. Transforming country figures into language figures permits not only significant research on multilingualism in cyberspace but also integrating demolinguistic projections as a possibility to understand the dynamics of linguistic uses on the Internet.

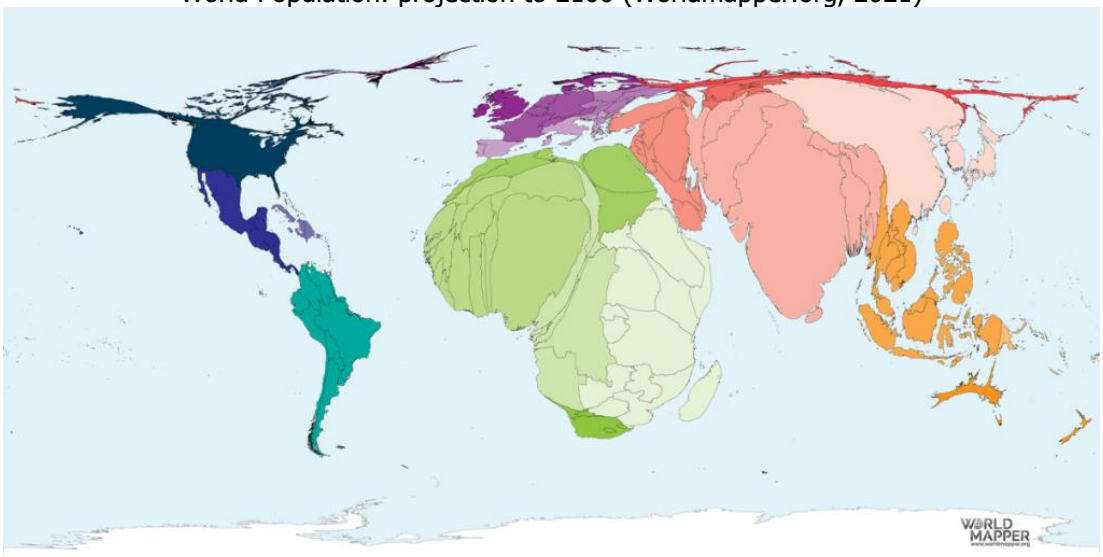
Demographic and demolinguistic trends

The next two maps compare the proportional demographic distribution by continent in 1900, the period of European colonial predominance, with the projection for 2100.

World Population in 1900 (Worldmapper.org, 2021)

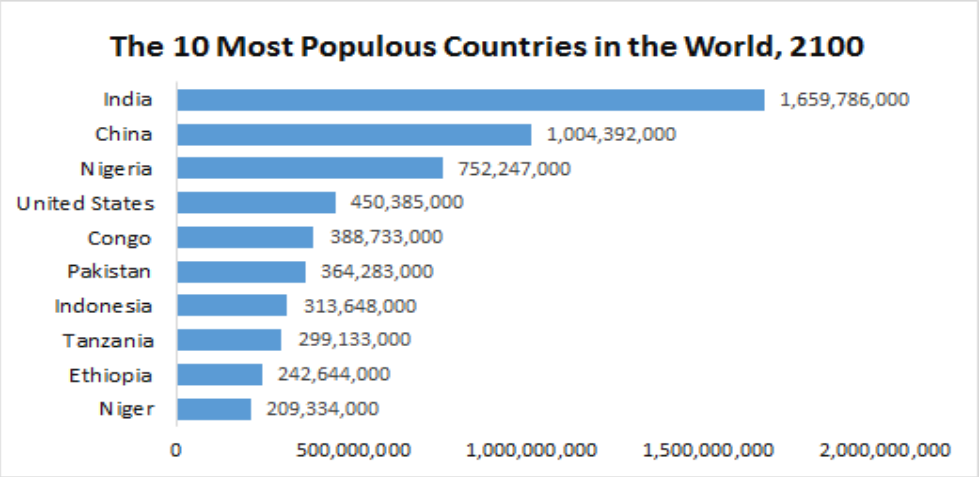


World Population: projection to 2100 (Worldmapper.org, 2021)



The maps show the relative gain in demographic importance of Asia and Africa, including Arabic-speaking regions, such as L1 (North Africa and the Arabian Peninsula) or L2 (the rest of the Islamic world where Arabic is a reference language). Projections show an internal differentiation in Asia, with a relative loss of importance in China and the north of the continent and accelerated growth in the Indian sub-continent and, to a lesser extent, in Southeast Asia. As for the African continent, the data point to a more intense growth that is well distributed across its various sub-regions.

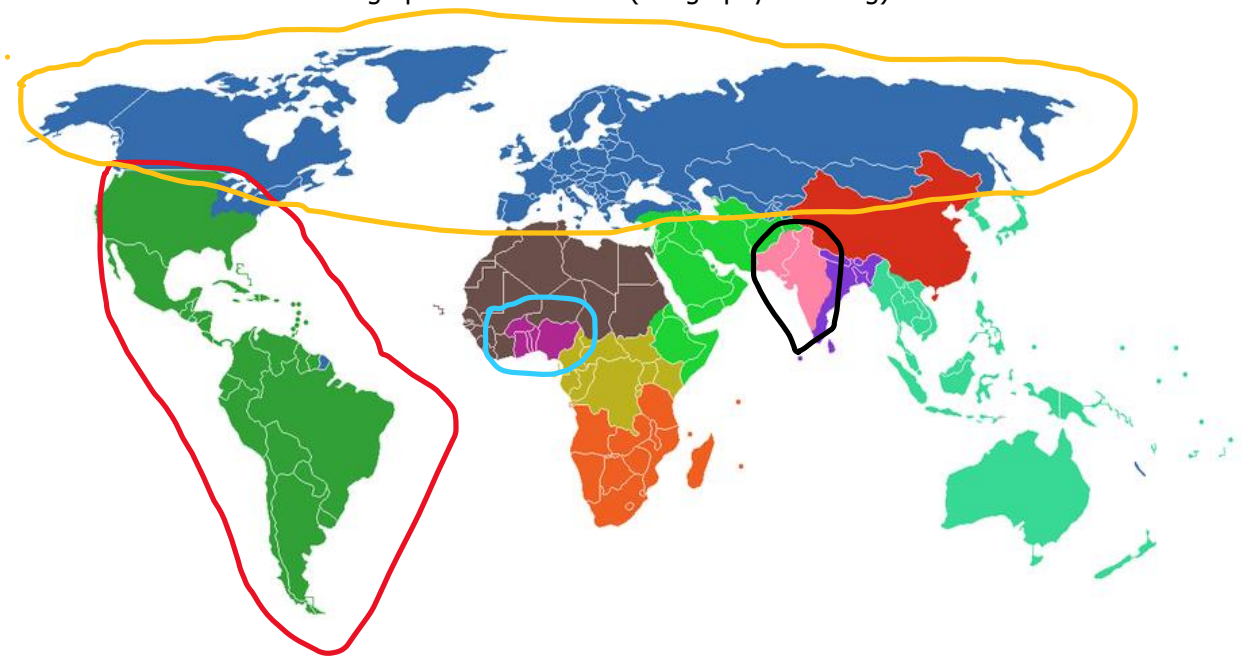
The demolinguistic consequences of this shift can be seen in a more refined way when we look at the list of the most populous countries in the world in the projection for 2100, as can be seen in the table below. In this list are 4 Asian countries with 3.3 billion inhabitants, 5 African countries with 1.9 billion inhabitants, and one American country with 450 million inhabitants:



Projections - Institut National d'Études Démographiques: https://www.ined.fr/en/everything_about_population/data/world-projections/projections-by-countries/

The map below allows us to verify the extreme inequality of demographic concentration globally in its projection for 2100. Each of the regions indicated in different colors will have approximately the same population. Relatively small areas, such as the western or eastern part of the Indian sub-continent or the Gulf of Nigeria region, concentrate as many people as the whole of Latin America plus almost the whole of the United States, or that of all northern Eurasia, including Western Europe and Canada.

Demographic concentration (Geographyrealm.org)



These maps and previous considerations are intended to briefly present the general lines of the demographic shift for Asia and Africa as an essential criterion for the presence of Asian and African languages on the Internet in the coming decades.

European languages in Africa and Asia: the case of Portuguese

Let's return to the issue of the impact of population growth in Asia and Africa to understand the change in the geographic base of some of the most widely spoken and used European languages on the Internet, such as English, French, or Portuguese. Portuguese will be taken as an example.

English is spoken (L1+L2) by 236,808,000 people in Africa, including 111 million in Nigeria (all L2), whereas 109,245,420 speak French, and 32,531,250 Portuguese (Ethnologue, 2021), both L1+L2 also.

This number of Portuguese speakers in the PALOP (Portuguese-speaking African Countries: Angola, Cape Verde, Guinea Bissau, Equatorial Guinea, Mozambique, and São Tomé and Príncipe) is about three times the number of speakers in Europe, but only around 15% of the number of speakers of the language in Brazil (Ethnologue, 2021). The table below, using relatively conservative data from the UN Population Division, shows numbers of each Portuguese-speaking country and the proportions of the total number of speakers of Portuguese in 2015:

Population of the Portuguese-speaking countries in 2015 (in millions). Population Division, UN (2015)

Angola	22,820	8.52%
Brazil	203,657	76.03%
Cape Verde	508	0.19%
Guinea Bissau	1,788	0.66%
Equatorial Guinea	0	0%
Mozambique	27,122	10.12%
Portugal	10,610	3.96%
São Tome And Principe	203	0.08%
East Timor	1,173	0.44%
Total	**267,881**	**100%**

In 2015, the population of Portuguese-speaking countries in Africa was 52,441,000, or 19.57% of the total population of Portuguese-speaking countries. Considering the internal multilingualism of those countries and a realistic estimate that only 50% are Portuguese speakers, we would reach an approximate number of 26 million Portuguese speakers on the continent (2015), or about 10% of the total number of speakers.

The table below presents demographic projections for 2100:

Population of the Portuguese-speaking countries in 2100 (in millions), Population Division, UN (2015)

Angola	97,337	22.97%
Brazil	194,533	45.91%
Cape Verde	552	0.13%
Guinea Bissau	5,628	1.33%
Equatorial Guinea	2,419	0.57%
Mozambique	112,018	26.43%
Portugal	7,457	1.76%
Sao Tome And Principe	568	0,13%

East Timor	3,265	0,77%
Total	**423,777**	**100.00%**

Demo-linguistic projections for the coming decades show an expressive growth of Portuguese speakers in the world, but a new relative distribution of speakers, with the predominance of Africans and a relative decrease in the weight of Brazil (Americas) and Portugal (Europe), a trend summarized below, calculated on the total of 423.777 million speakers, as shown in the previous table:

Portuguese speaker's projection

Africa (Angola + Mozambique Alone = 209,355/49.40%)	218,522 million	51.56%
Brazil	194,533 million	45.90%
Portugal	7,457 million	1.76%

A similar study could be carried out for the case of demolinguistic projections for English and French in Africa and Asia. The predominance of African speakers in international languages of European origin, such as Portuguese, has several consequences for the cyber-geography of languages. One of them is to displace the geopolitical center of the language from more monolingual spaces, such as Europe and mainly the Americas, with the National States organized around a national language, to more multilingual spaces, such as Africa, where this language works more like L2 than L1.

The web is transformed every day more and more into one of the places where the expression of multilingualism is the strongest. If there were really a *lingua franca* in the web, it would undoubtedly be **translation**, the indispensable crutch of multilingualism, to follow the vision of Umberto Eco who stated in 1993 that *the language of Europe is translation*.

The relationships between multilingual and monolingual solutions, in cyberspace and beyond, can be analyzed in terms of a continuum, the *Multilingualism Continuum*, which involves not only different languages but different models of communication:

Multilingualism – Translation – Lingua Franca

The more multilingual the Internet becomes, the less a lingua franca or translation is needed. The more instruments, models and translation practices evolve, the less we need one lingua franca. And, finally, with the growth of the first two factors, the possibility of several *lingua(s) franca(s)*, for specific communicative circuits, is given, or a repertoire of languages emerges, which Steyaert (2014) called 'multilingual franca'. In conclusion, the belief that the lingua franca of the Internet is English is an illusion emerging from the past: what characterizes the Internet today, and more and more in the future, is **multilingualism**[4], and the digital economy is increasingly characterized by the same multilingualism factor.

[4] Internet is the human space where multilingualism is expressed in its best and greatest way, given its borderless characteristics. The degree of multilingualism of the Internet could be higher than that of humans, be it in terms of content, traffic, uses, or interfaces.

References

Calvet, Louis-Jean. Les marchés aux langues. Les effets linguistiques de la Mondialisation. Paris, Plon, 2002.

Müller de Oliveira, Gilvan. The system of national standards and the demolinguistic evolution of Portuguese. In Muhr, Rudolf in collaboration with Eugenia Duarte, Amália Mendes, Carla Amorós Negre and Juan A. Thomas (Ed.) Pluricentric languages: non-dominant varieties worldwide: Volume 2. The Pluricentricity of Portuguese and Spanish: New concepts and descriptions. Frankfurt a.M. / Wien, Peter Lang Verlag, 2016, 31-43.

Pimienta, Daniel, Müller de Oliveira, Gilvan. Cyber-geography of languages. P1: method, results and focus on English – International Review on Information Ethics, Vol 32 (12/2022)

Pimienta, Daniel. New and improved version of an alternative approach to the production of linguistic indicators on the Internet. Observatory of linguistic and cultural diversity on the Internet – 8/2021. http://funredes.org/lc2021/ALI%20V2-EN.pdf

Steyaert, Janssens. M. Re-considering language within a cosmopolitan understanding: Toward a multilingual franca approach in international business studies. J Int Bus Stud 45, 623–639 (2014).

United Nations. Department of Economic and Social Affairs, Population Division. World Population Prospects, the 2012 Revision. (2015) http://esa.un.org/unpd/wpp/