

The State of Cultural Diversity in Arabic Wikipedia: Challenges and Opportunities

Marc Miquel, PhD

{marcmiquel@gmail.com}

Username:marcmiquel

Pompeu Fabra University, Barcelona, Catalonia

Catalan Wikipedia, Amical Wikimedia

Marrakesh, 4-6 October 2019



My name is Marc Miquel, I am a diversity **researcher**.

In 2010, I got interested in Wikipedia, especially when I realized that each language has different content.

So I decided to study it.



How is my culture represented in Wikipedia?

...

How is Arabic culture represented in Wikipedia?

Wikipedia project does not reflect enough the world's cultural diversity.



- Some languages, contexts and their concepts are not in Wikipedia.
(**Representation**)
- Some are in Wikipedia but remain exclusive to some language editions.
(**Sharing**)

				ሂ					
			᠋ᠪᠢ	᠋ᠰᠢ		᠋ᠤᠮᠤ	᠋ᠣᠨ		
᠋ᠰᠢ	᠋ᠪᠢ	᠋ᠶ	᠋ᠡ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ
B	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ
᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ
᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ
				᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ			
				᠋ᠮᠤ	᠋ᠮᠤ	᠋ᠮᠤ			

→ Missing pieces

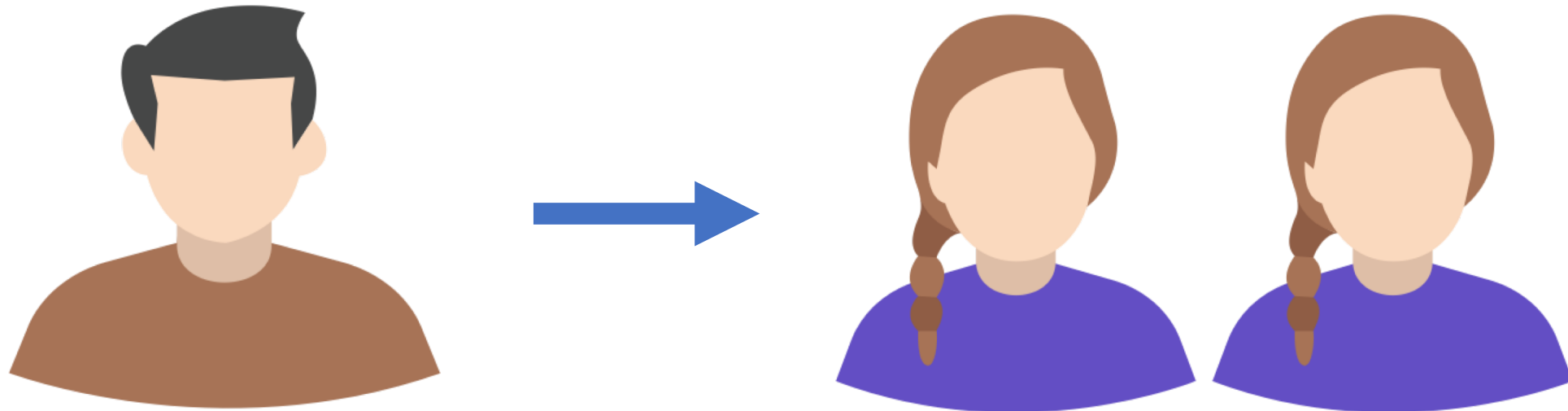
Wikipedia Cultural Diversity Observatory (WCDO).

“a joint space for **researchers** and **activists** to study and **fight against the knowledge gaps** and promote knowledge equity.”

From **Raising Awareness** to Proposing **Solutions**



We cannot do like content gender gap.



It is a bit more more difficult. We need a **cartography** of local content in every language in order to know what is **represented** and what is **shared**.



For each Wikipedia language edition, we select the **Cultural Context Content (CCC)**, i.e. traditions, language, politics, agriculture, biographies, places, events, etcetera, related to the territories where the language is spoken.

What is in Cultural Context Content?

أبو نواس

هذه المقالة عن أبو نواس، الشاعر العباسي. لتصفح عناوين مشابهة، انظر أبو نواس (توضيح).

أبو نواس أو **الحسن بن هاشم الحكمي شاهر** عربي من أشهر شعراء عصر الدولة العباسية. ويكنى بأبي علي وأبي نواس والواسي. وعرف أبو نواس بشاعر الخمر. ولكنه تآب عما كان فيه واتجه إلى الزهد وقد أشهد عدد من الأئمة التي تدل على ذلك.^[1]

محتويات [إخفاء]

- مولده
- حيته
- وفاته
- قصة رؤية أصحابه له في المنام
- أسلوبه
- ديوانه
- آراء بعض الرواة في شعر أبي نواس
- مختارات من شعر أبي نواس
- 1 الفهر
- 2 الفهر

الصفحة الرئيسية
الأحداث الجارية
أحدث التغييرات
أحدث التغييرات الأساسية

نصيح
التراسع
أبجد
بوابات
قائمة صفحات
تصفح بدون إشارات

مشاركة
تواصل مع ويكيبيديا
مساعدة
الديوان
تدريج

معمول

المعمول أو **كحك** **بعجوة** أو **كحك العيد**، هو أحد أهم مظاهر الاحتفال والفرح المصاحبة لعيد الفطر المبارك، ويتم تحضيره عادة خلال الأيام الأخيرة من شهر رمضان، وقد تضرّبت الآراء واختلفت حول سميات المعمول أو «كحك العيد»، فهو يعرف باسم «الكحك» لدى عامة المصريين، فيما يسميه الفلاحون في ريف مصر «كحك»، ويسميه الخليجيون «المعمول».

والمعمول له تاريخ يمتد إلى عهد القراة، فالقوشات التي تحمطها حبة المعمول في ذلك الوقت ما هي إلا رسة للشمس، التي عرفت عند القراة بالاله «واتن»، أحد آلهة القراة، كما تشير كتب التاريخ، حيث كان القراة يضعون المعمول مع الموش داخل المقابر، وقد استمر المصريون في صنعه وتقديمه في الأعياد أو عند زيارة الموتى والقبور إلى أيامنا هذه.

وهناك من يقول إن كلمة «كحك» فارسية، وإن أول من صنعه هو العثمانيون، بينما يقر الجميع أن صناعة الكحك شهدت فترة ازدهارها في ظل الحكم العثماني، حيث خصص الفاطميون إدارة حكومية عرفت باسم «دار الفطرية» كانت تهتم بتجهيز الكحك وتوزيعه، ويضم متحف الفن الإسلامي بعض الأثار لما كان يكتب على الكحك بواسطة القوالب التي ينشكل منها، وكانت أبرزها عبارات مثل: «كل هنيئا كل واشكر، ياكثر تودم التبع».

وفي عصرنا الحالي تطور المعمول وتبرعت أشكاله ليشمل أنواعاً وأشكالاً جديدة، فهناك المعمول التمر التقليدي، وهو الأكثر شيوعاً ورواجاً، إضافة إلى معقول المكسرات والسيد والجوز والفتق، كذلك معقول التفاح، وهو آخر مستحدثات سوق المعمول. وفي حين كان المعمول يباع في المنزل، حيث كانت تجتمع النساء لصفه، ثم مع تطور المجتمع والحصار لنطاق المعارف، إضافة إلى ضيق الوقت اعتمد كثير من الأسر على شراءه من السوق، إلا أن طعم المعمول المنزلي يختلف حيث له نكهة خاصة تميزه، ما دفع الكثير من ربات البيوت الماهرات إلى صنعه وبيعهم من بيوتهن للترابحات، وقد لقيت هذه الطريقة رواجاً وإقبالاً كبيرين، خاصة مع اقتراب شهر رمضان.^[1]

المعمول نوع من الحلويات الشعبية المشهورة في المنطقة العربية، وخاصة في لبنان وسوريا والأردن والسعودية وفلسطين والعراق. يتم عادة حشوها بالتمر.^[2] أو المكسرات كالجوز^[3] أو الفستق الحلبي.

المصادر [إعلاء]

- ↑ "المعمول" يأخذ من الشمس شكله ومن التمر حالته". *archive.aawsat.com*. اطلع عليه بتاريخ 31 مايو 2018.
- ↑ معمول بالتمر موقع معلم نسخة محفوظة 14 أغسطس 2016 على موقع واي باك مشين.
- ↑ معمول بالجزر موقع معلم نسخة محفوظة 28 سبتمبر 2015 على موقع واي باك مشين.

ويكيبيديا

الصفحة الرئيسية
الأحداث الجارية
أحدث التغييرات
أحدث التغييرات الأساسية

نصيح
التراسع
أبجد
بوابات
قائمة صفحات
تصفح بدون إشارات

مشاركة
تواصل مع ويكيبيديا
مساعدة
الديوان
تدريج

أحداث

مقالة هنا
تغييرات ذات علاقة
الصفحات الخاصة
وسمات القالب
معلومات الصفحة
تخصص ويكيبيديا

صومعة حسان

هذه المقالة عن صومعة حسان. لتصفح عناوين مشابهة، انظر صومعة (توضيح).

صومعة حسان من المباني التاريخية المشهورة بالعاصمة المغربية الرباط.^[1] والتي شيدت في عصر دولة الموحدين. تم تأسيس «جامع حسان» بناءً على أمر بطرب المنصور سنة 593 هـ (1197-1198 م).^[2] أضادت منظمة اليونسكو للتراث العالمي هذا الموقع لقائمة أروية 1 يوليو 1995 في فئة **الثقافة**.^[3]

محتويات [إخفاء]

- لمحة تاريخية
- مواد البناء
- مكتبات المسجد
- 1 بيت الصلاة
- 2 أعمدة المسجد
- 3 محراب المسجد
- 4 أبواب المسجد
- 5 صحن المسجد

الصفحة الرئيسية
الأحداث الجارية
أحدث التغييرات
أحدث التغييرات الأساسية

نصيح
التراسع
أبجد
بوابات
قائمة صفحات
تصفح بدون إشارات

مشاركة
تواصل مع ويكيبيديا
مساعدة
الديوان
تدريج

الأمن الوطني (المغرب)

لمعانٍ أخرى، انظر الأمن الوطني (توضيح).

الشرطة المغربية أو **الأمن الوطني** هو الجهاز الأول المسؤول عن الأمن في المغرب.^{[1][2][3]}تأسس على يد الملك محمد الخامس في 16 ماي 1956، ويهدف الجهاز إلى الدفاع عن مقدرات الدولة وإرساء الأمن بالبلد والحرس على تطبيق القانون، بالإضافة إلى التدخل في حالات الكوارث الطبيعية لإغاثة وإجلاء المتضررين وحماية ممتلكاتهم.

كما يحرص الأمن الوطني المغربي، على حسن سير مختلف التجمعات الكبرى، المسيرات، الملتقيات الرياضية، الأنشطة العمومية، تأمين مباريات كرة القدم والحد من التخلف.

محتويات [إخفاء]

- الوحدات
 - 1.1 الشرطة القضائية
 - 1.2 شرطة المرور
 - 1.3 مكافحة المخدرات
 - 1.4 الشرطة القضائية والحماية
 - 1.5 فرق التدخل السريع
 - 1.6 الصفر
 - 1.7 فرقة الخيالة
 - 1.8 الحماية المقربة
 - 1.9 فرقة الكلاب البوليسية
 - 1.10 فرقة الأخلاق العامة
 - 1.11 فرقة تفكيك المتفجرات
- المتدربين
- الأسطول

ويكيبيديا

الصفحة الرئيسية
الأحداث الجارية
أحدث التغييرات
أحدث التغييرات الأساسية

نصيح
التراسع
أبجد
بوابات
قائمة صفحات
تصفح بدون إشارات

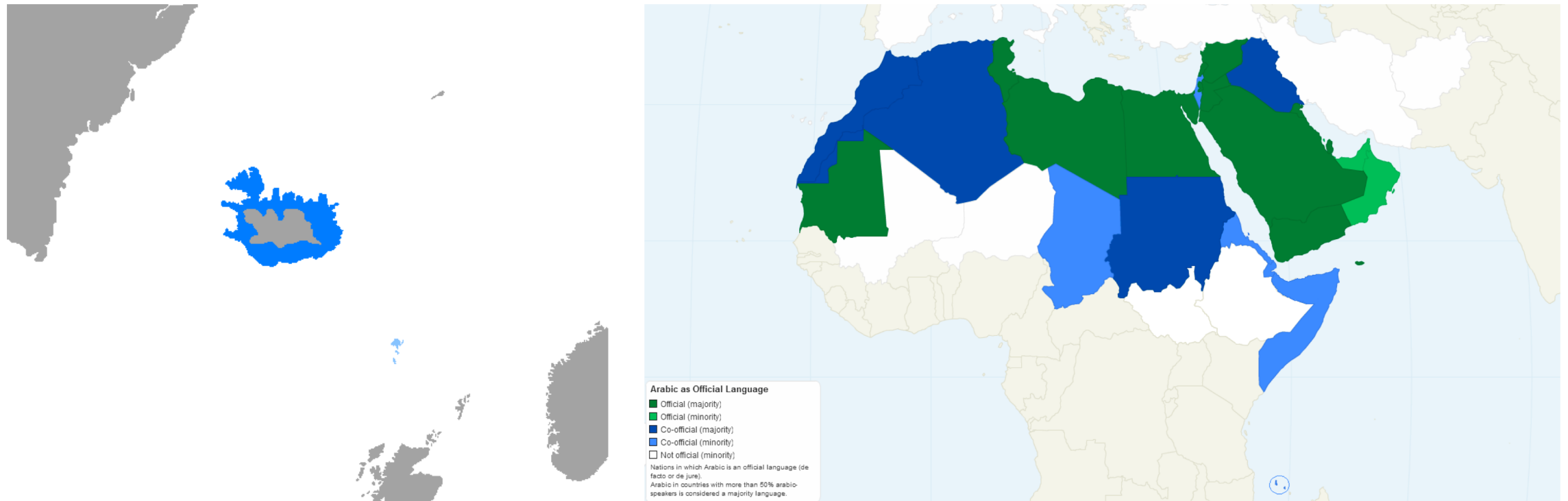
مشاركة
تواصل مع ويكيبيديا
مساعدة
الديوان
تدريج

أحداث

مقالة هنا
تغييرات ذات علاقة
الصفحات الخاصة
وسمات القالب
معلومات الصفحة
تخصص ويكيبيديا

It can be an article about a poet, a tower, a cake or the police.

Language-Territories Mapping



Icelandic CCC only relates to concepts on Icelandic context.

Arabic CCC includes articles about everything related to Algeria, Bahrain, Chad, Comoros, Djibouti, Egypt, Eritrea, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, and a long etcetera of 31 countries.

Features to assess the relatedness to a cultural context

- **Geolocation**
- **Category Graph**
- **Keywords in title**
- **Some Wikidata properties**

Location, Country, Language, Is Part of, Affiliation, etc.

- **Link features:**

Number and percentage of Inlinks/Outlinks (incoming/outgoing links).

Machine Learning Classifier -> Cultural Context Content Dataset



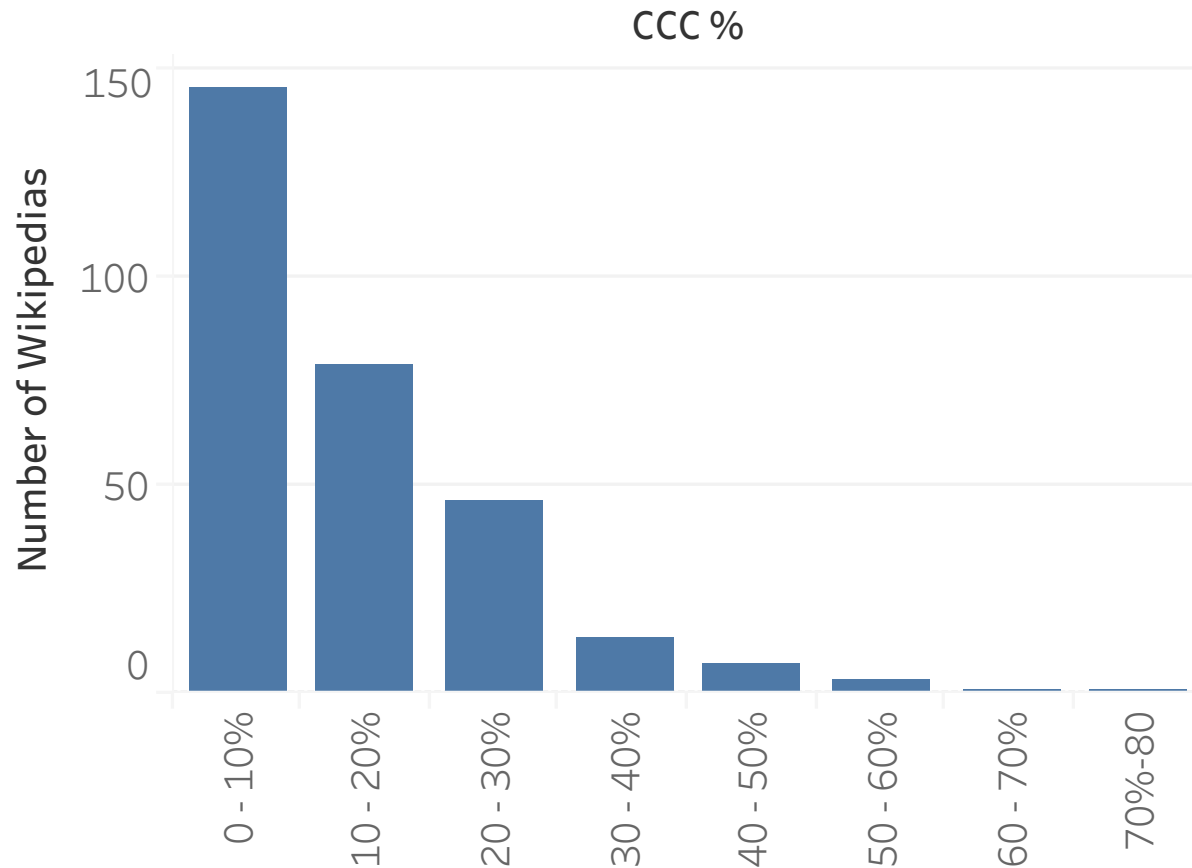
For each Wikipedia language edition, we obtain the **Cultural Context Content (CCC)**, i.e. traditions, language, politics, agriculture, biographies, places, events, etcetera, related to the territories where the language is spoken.

Taking into account **the largest 40 Wikipedia** language editions, CCC is in average about a quarter of each Wikipedia



[https://wcdo.wmflabs.org/list_of_wikipedias_by_cultural_context_content]

We have many Wikipedias with an underdeveloped CCC



145 Wikipedias CCC is below 10% of their content.

92 Wikipedias with < 100 art. geolocated in their territories.

Problem of representation

Quick Analysis of Cultural Diversity in Arabic Wikipedia

- The **representation** of Arabic Cultural Context in Arabic Wikipedia
- The degree of **coverage** of all other cultural contexts by Arabic Wikipedia, and the degree of **spread** of Arabic Cultural Context in all Wikipedias.





wikiARABia **Month?**

The **representation** of Arabic Cultural Context in Arabic Wikipedia



CCC Extent



ويكيبيديا
الموسوعة الحرة

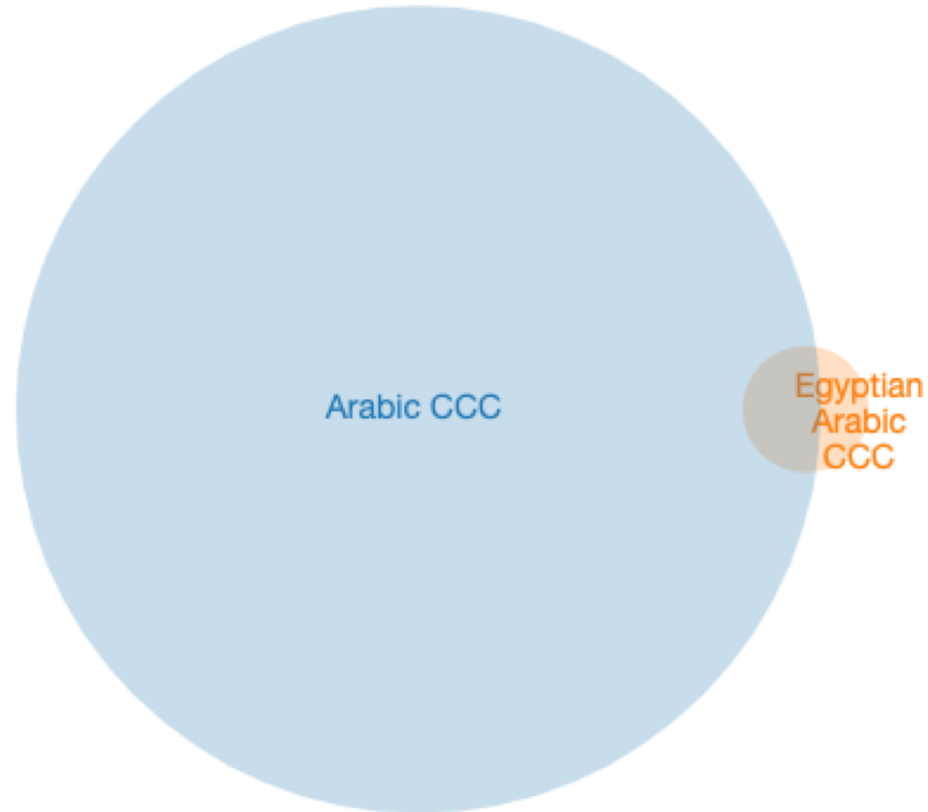
**25.15% of Arabic
Wikipedia is CCC.**

of 830,582 articles (July 2019)

Egyptian Arabic (31.1%) of 20,197 articles.

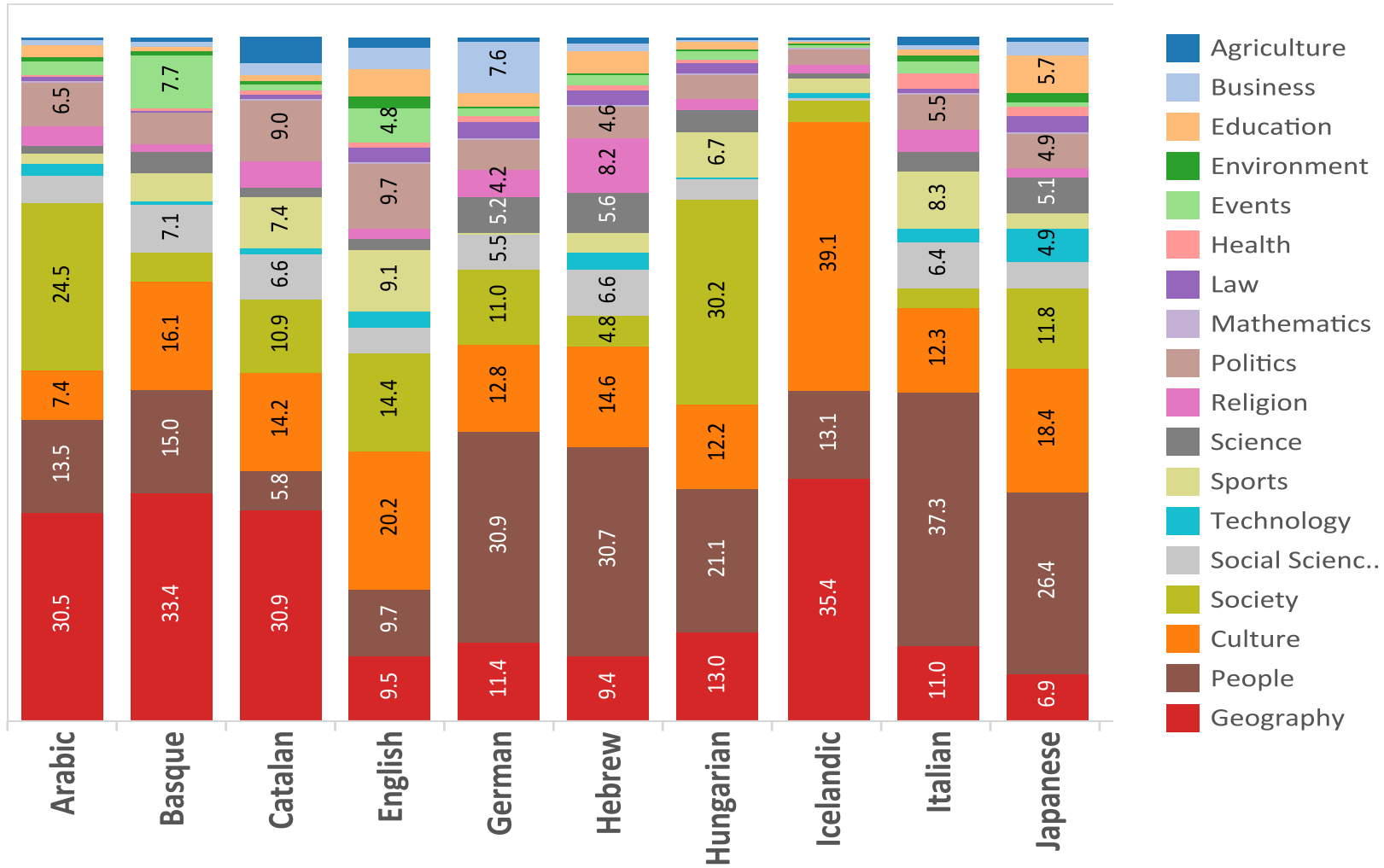
Other languages with a large number of speakers: Chinese (11.2%), English (44.2%), Spanish (29.9%), Hindi (56.3%) and Russian (32.2%).

Egyptian Arabic Wikipedia CCC



- **Egyptian Arabic CCC is 31.1% of 20,197 articles.**
- **70.7% of Egyptian Arabic CCC (4,057 articles) are covered by Arabic Wikipedia.**

Topics



Arabic CCC has 30% geography, 24.5% society, 13.5% people, 4% religion.

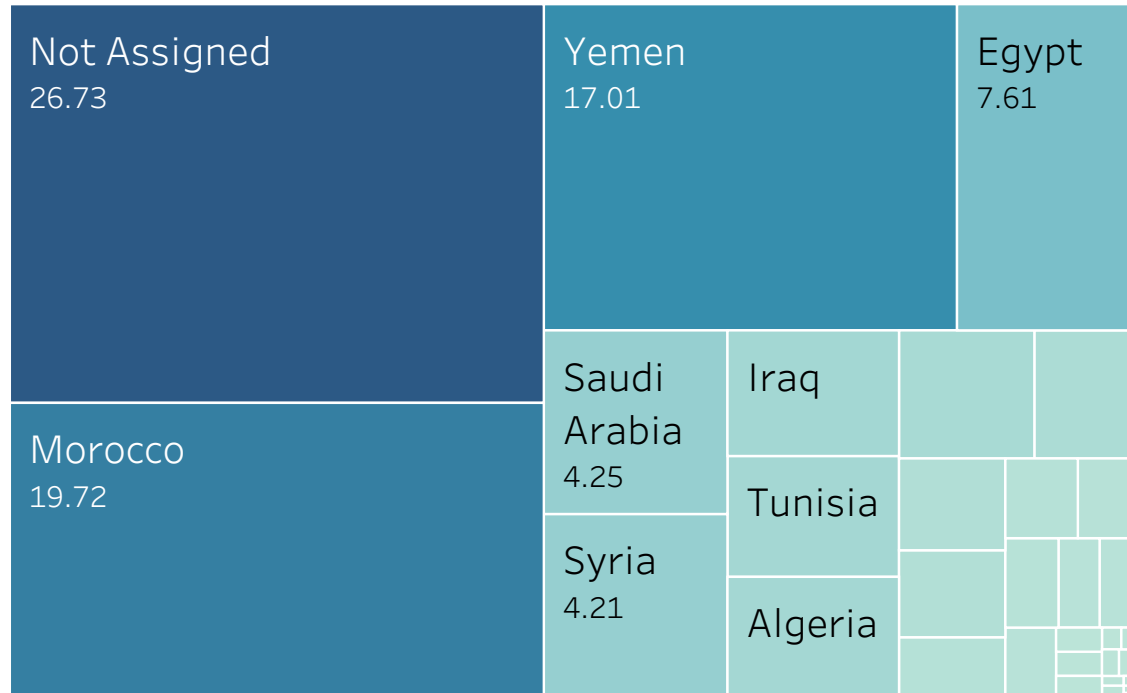
Gender gap

What is the women % in biographies?

	Wikipedia	CCC
Arabic:	16.2	14.5
Egyptian Arabic:	41.4	19.3
English:	17.4	17.8
Chinese:	18.8	36
Spanish:	20	26.6
Hindi:	21.9	24.2
Russian:	14.3	12.5

Gender Gap is more evident in Arabic CCC than in the whole Wikipedia.

List of Territories Extent (%) inside Arabic CCC



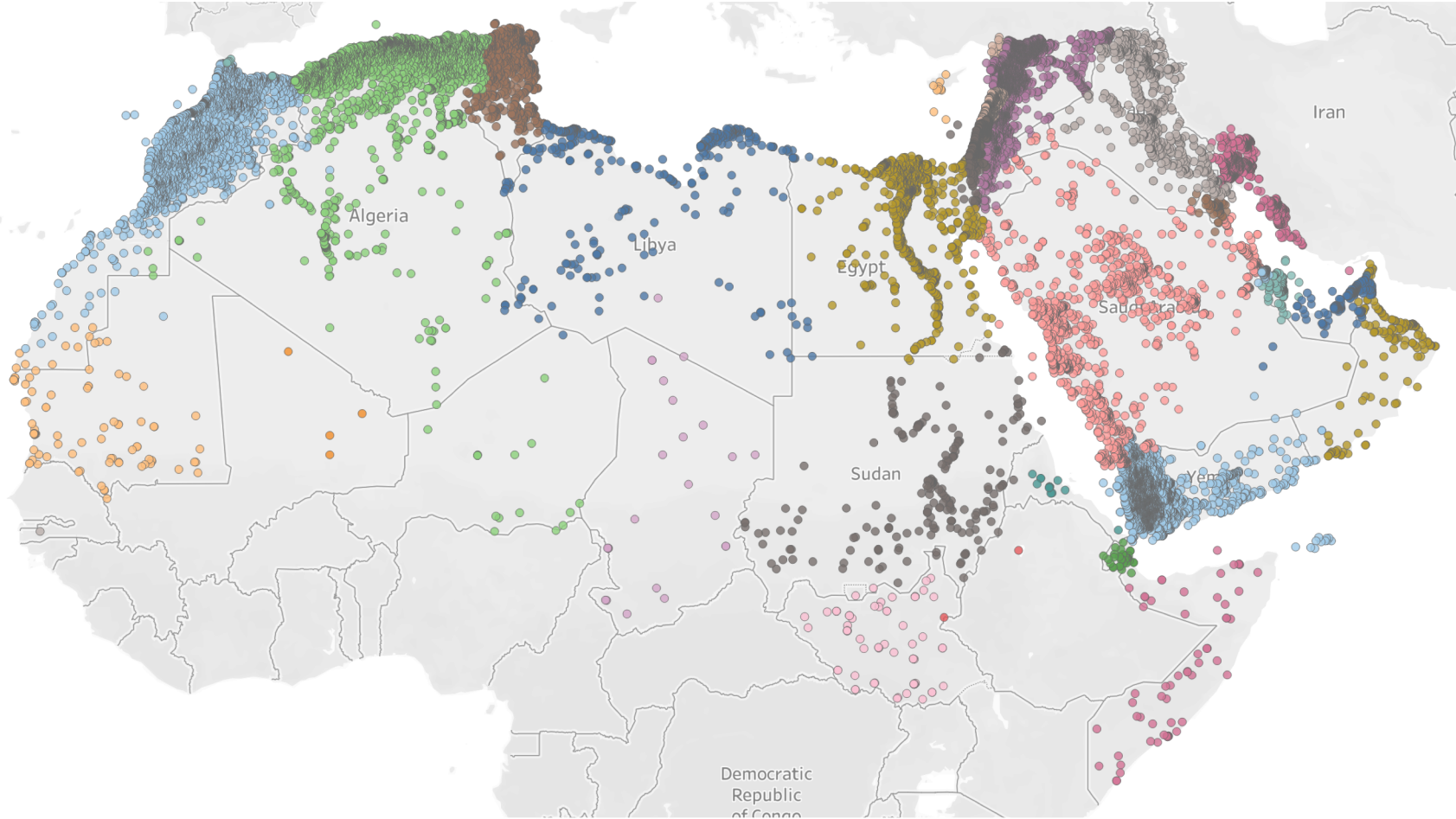
- Unassigned 55,277 (26.73%)
- Morocco 40,770 (19.72%)
- Yemen 35,167 (17.01%)
- Egypt 15,735 (7.6%)
- Saudi Arabia 8,787 (4.25%)
- Syria 8,698 (4.21%)
- Tunisia 5,359 (2.59%)
- Algeria 5,352 (2.59%)
- Iraq 5,664 (2.74%)
- Oman 3,593 (1.74%)
- ...

Can you explain these proportions? Population, size, GDP...? None correlates.

Some countries cultural context are underrepresented in Arabic CCC

[https://wcdo.wmflabs.org/list_of_language_territories_by_cultural_context_content/]

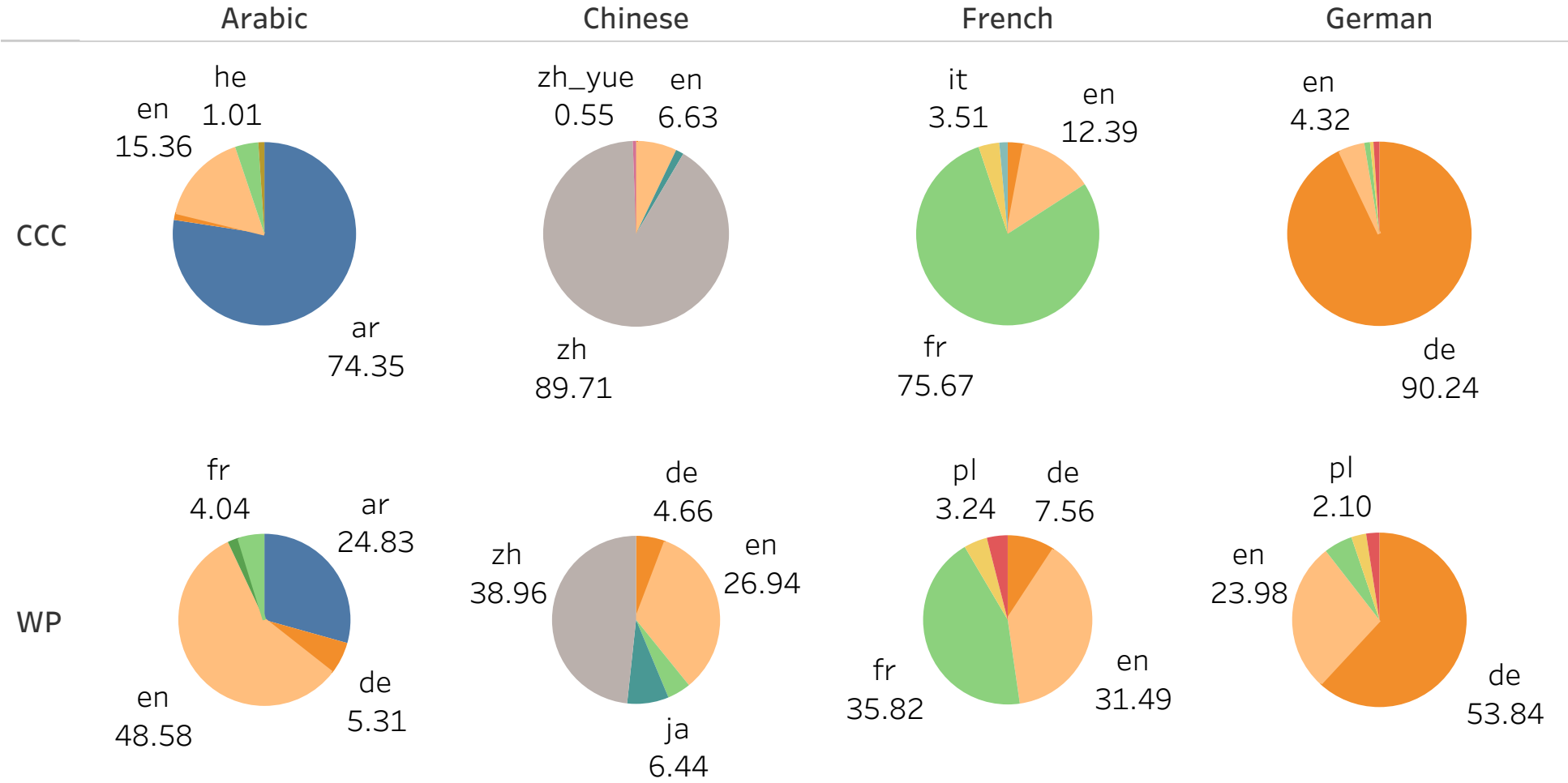
Arabic context geolocated articles



- Syria 3,977 (17,2%)
- Yemen 3,099 (13,4%)
- Algeria 2,759 (11,9%)
- Morocco 2,132 (9,2%)
- Saudi Arabia 1,439 (6,2%)
- Egypt 1,364 (5,9%)
- Tunisia 1,215 (5,2%)
- Iraq 972 (4,2%)
- Israel 917 (3,9%)
- ...

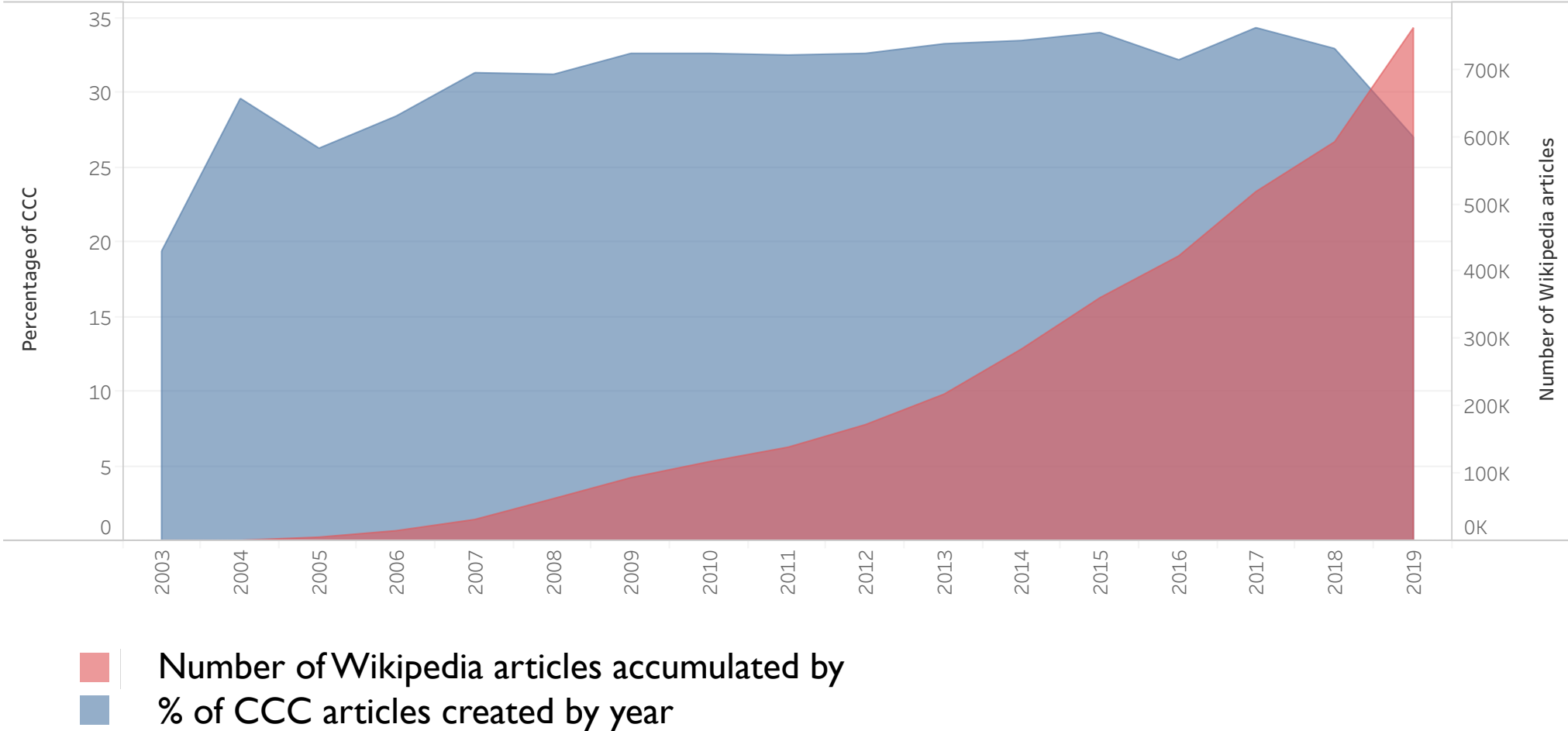
23,007 Arabic Wikipedia Geolocated articles in Arabic speaking countries

Originality Percentage of articles with the first timestamp by Language



Arabic Wikipedia tends to create articles that have previously been created in English Wikipedia more than other languages.

Creation of CCC over time



CCC is continually created over time – in the last year it decreased around 10%

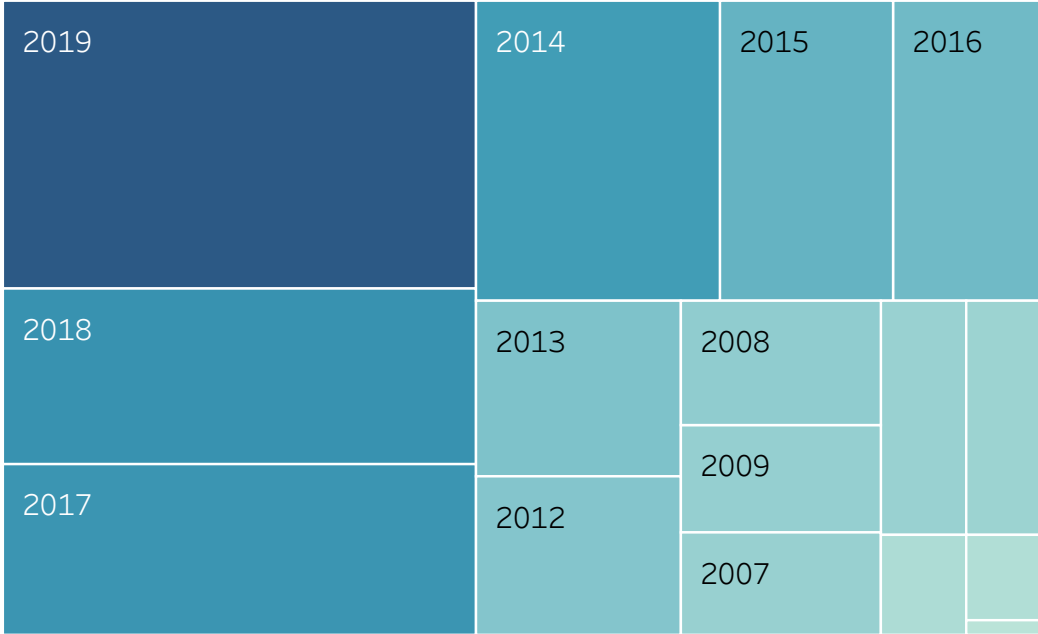
Should we continue creating Cultural Context Content?

Well, there are countries to cover better...

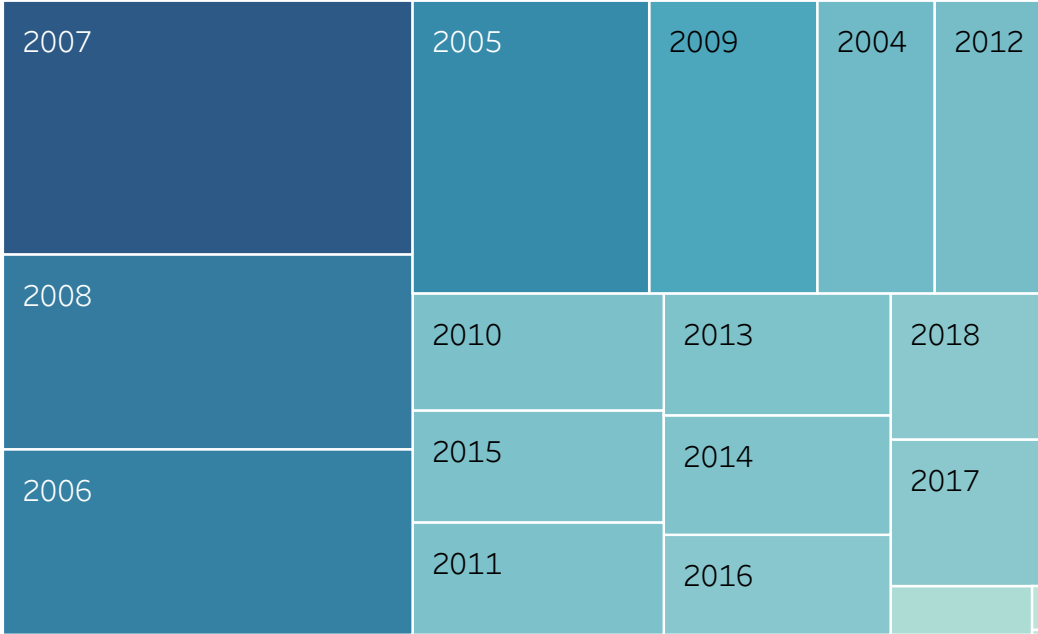


CCC articles are continuously created and new ones are read to a certain extent.

Arabic CCC articles by Year of Creation



Sum of Pageviews in Arabic CCC by Year of Creation



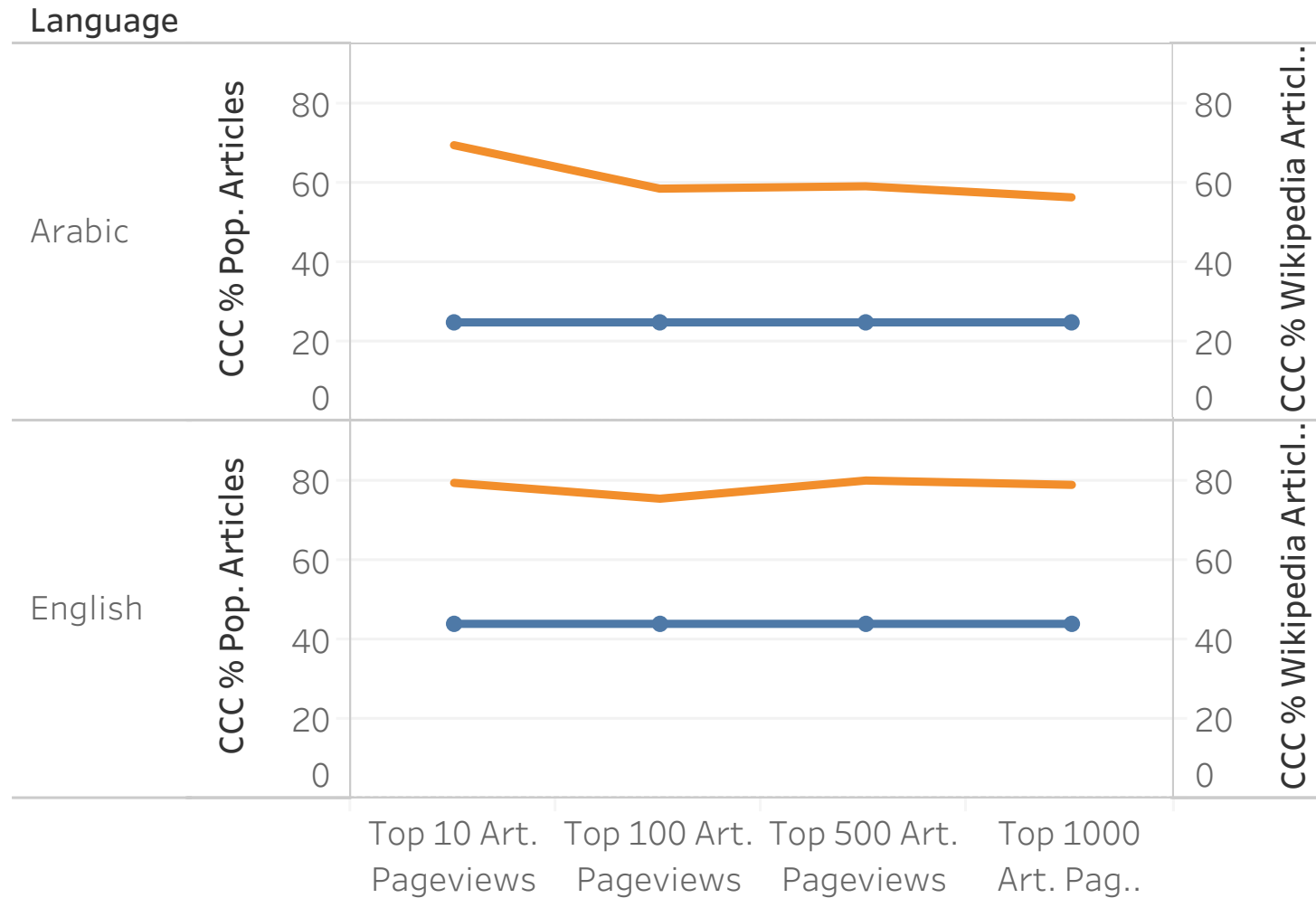
Most popular articles in Arabic Wikipedia during August 2019

Rank	Page	Edits	Editors	Pageviews
1.	قصب السكر (جنس)	8	5	352,785
2.	يوم عرفة	8	7	243,553
3.	ياسمين عبد العزيز	5	4	222,421
4.	مديحة كامل	16	11	174,192
5.	بيت من ورق (مسلسل)	2	13	148,172
6.	فيلم إباحي	3	3	126,661
7.	بطولة اتحاد غرب آسيا لكرة القدم 2019	130	34	122,029
8.	محمد بن سلمان آل سعود	3	2	120,135
9.	قط (جنس)	2	2	117,198
10.	عيد الأضحى	29	17	111,610

7/10 are clearly CCC articles

[<https://tools.wmflabs.org/topviews>]

* Exceptions are sex and cats. Not surprising.



Percentage of CCC in the Top 10, 100, 500 and 1000 most viewed articles (left axis and blue) higher than the **percentage of CCC in all Wikipedia articles** (right axis and orange).

Sometimes other Wikipedia language editions cover your language cultural context too.

We need to take advantage of it to stimulate representation



Select the Wikipedia



Target Language: Arabic (ar) × ▾
Target Country (Optional): Morocco × ▾
Target Region (Optional): Select a region (optional) ▾

Filter by content

Search Type: Type of gap ▾
CCC Segment: Geolocated × ▾
Topic: Select a topic ▾

Choose source of content

Source language(s): French (fr) × ▾
Order by feature: Number of Interwiki × ▾
Limit the results: 100
QUERY RESULTS!

N°	Language	Title	Number of Editors	Number of Pageviews	Number of Interwiki	Number of Bytes	Label (Lang)	Qitem
1	fr	Archimedes (cratère)	36	0	19	0.0k		Q634366
2	fr	Vénus de Tan-Tan	7	0	16	0.0k		Q856764
3	fr	Circuit Moulay El Hassan	22	0	14	0.0k	en Circuit International Automobile Moulay El Hassan	Q1092834
4	fr	Montes Spitzbergen	8	0	12	0.0k		Q501100
5	fr	Skoura	62	0	11	0.0k	ar واحة سكورة	Q986666
6	fr	Gorges du Dadès	41	0	11	0.0k	en Dadès Gorges	Q901124
7	fr	Tournoi de tennis du Maroc (ATP 2013)	14	0	10	0.0k	ar بطولة الدار البيضاء للتنس 2013	Q8661147
	fr	Mausolée Mohammed-V	56	0	10	0.0k	ar moselée	Q219028
	fr	Tournoi de tennis du Maroc (ATP 2014)	10	0	9	0.0k	ar بطولة الدار البيضاء للتنس 2014	Q15981145
10	fr	Cap Chaunar	15	0	9	0.0k	en Cape Chaunar	Q788817

Solution: Missing CCC

For Arabic, it works well with English, French,...

The degree of **coverage** of all other cultural contexts by Arabic Wikipedia, and the degree of **spread** of Arabic Cultural Context in all Wikipedias.



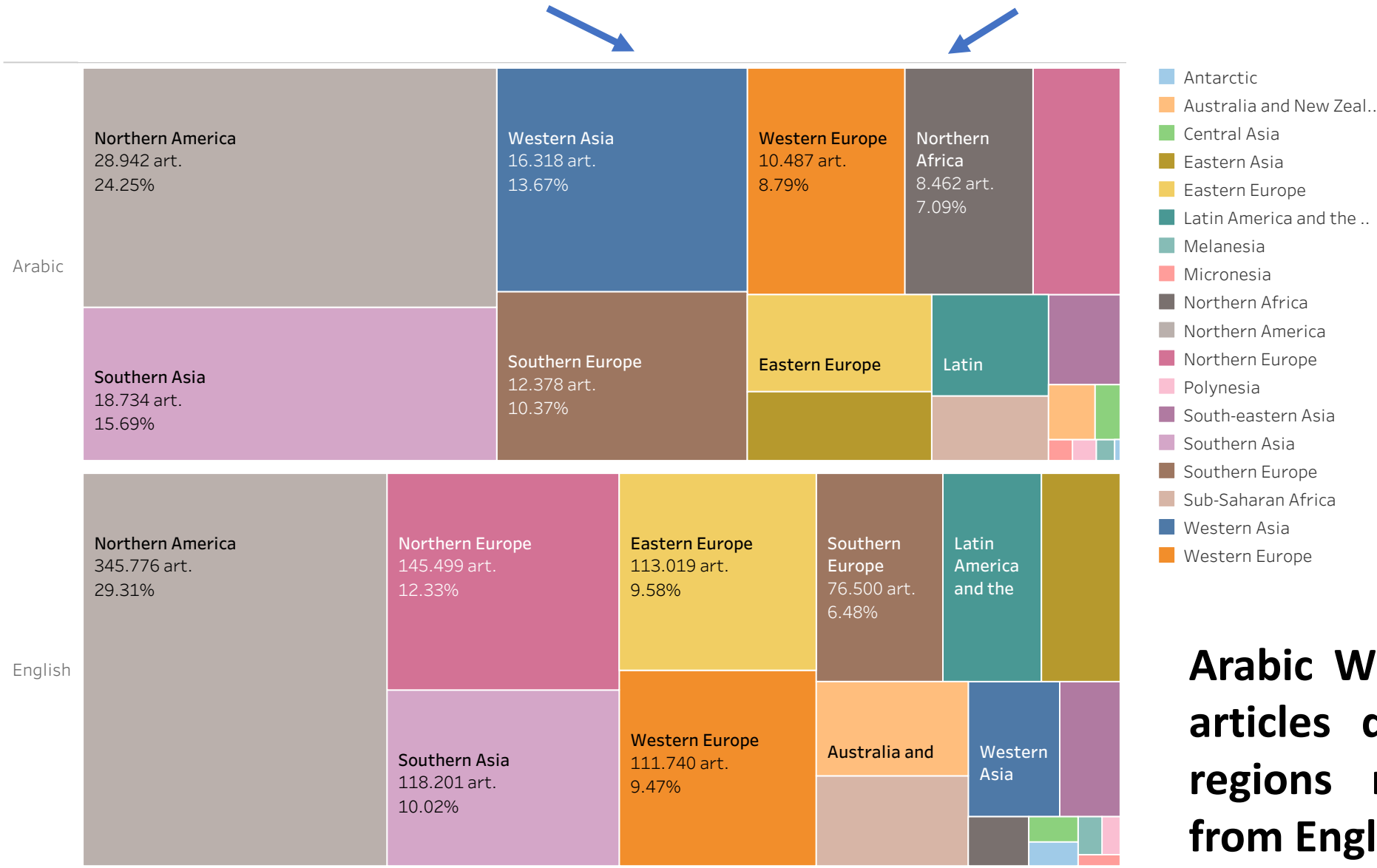
Wikipedia geolocated articles by world region



Asia is more prominent in Arabic, but Europe and North America are very present.

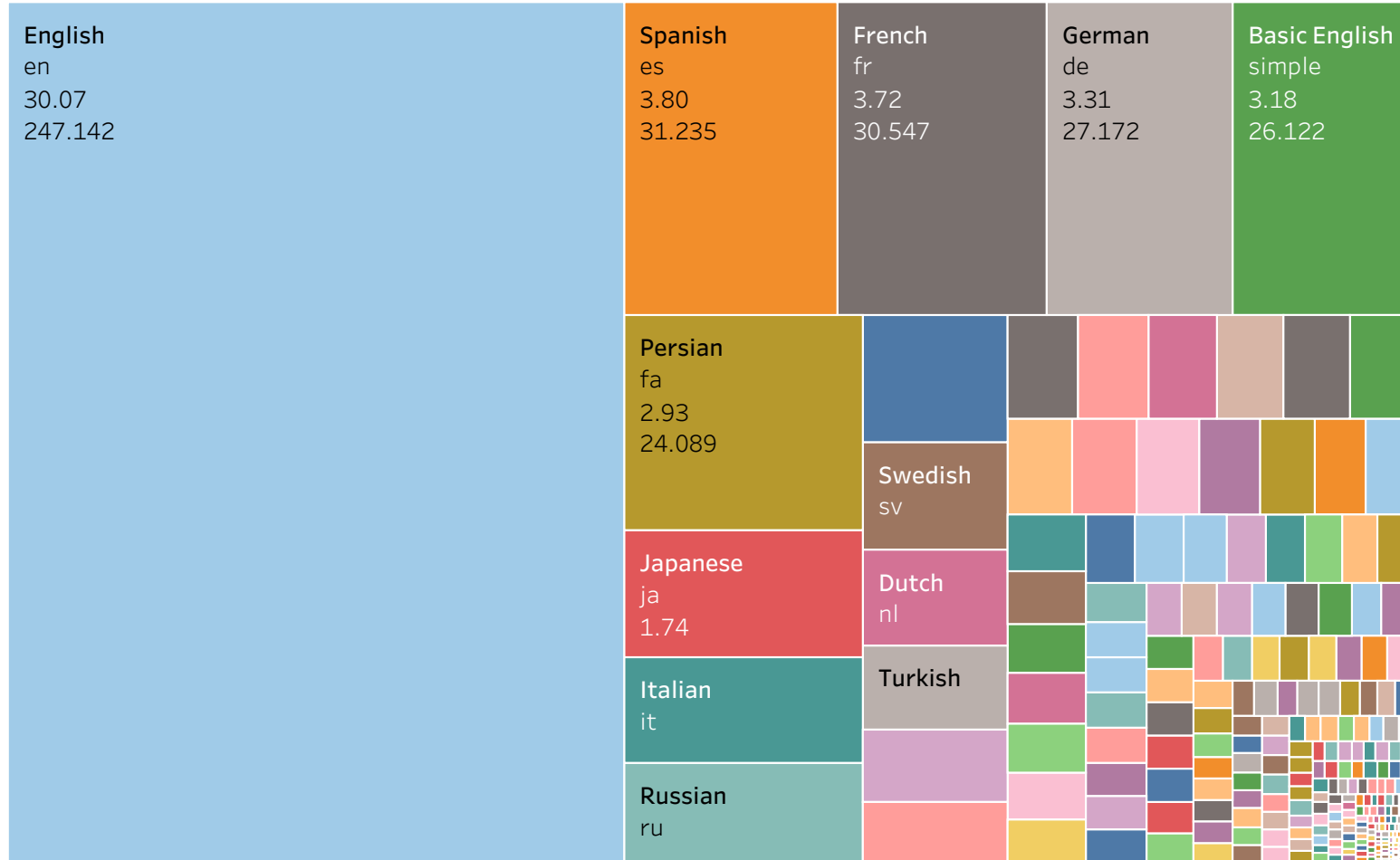
Arabic seems to follow English to create geolocated articles.

Wikipedia geolocated articles by world sub region



Arabic Wikipedia geolocated articles distribution by sub regions not that different from English.

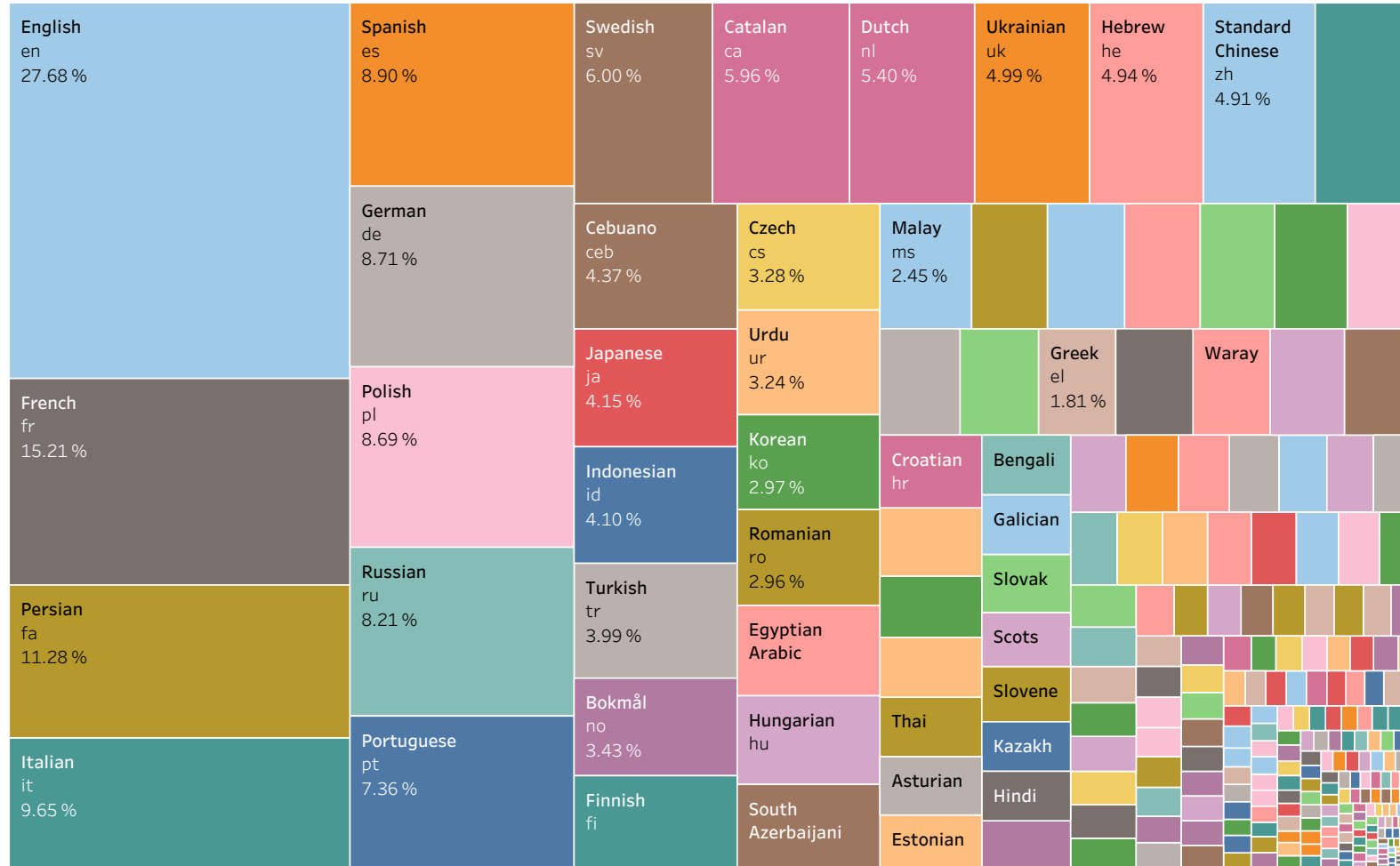
Wikipedia languages CCC extent in Arabic Wikipedia



Arabic CCC is 27%
(not in the viz).

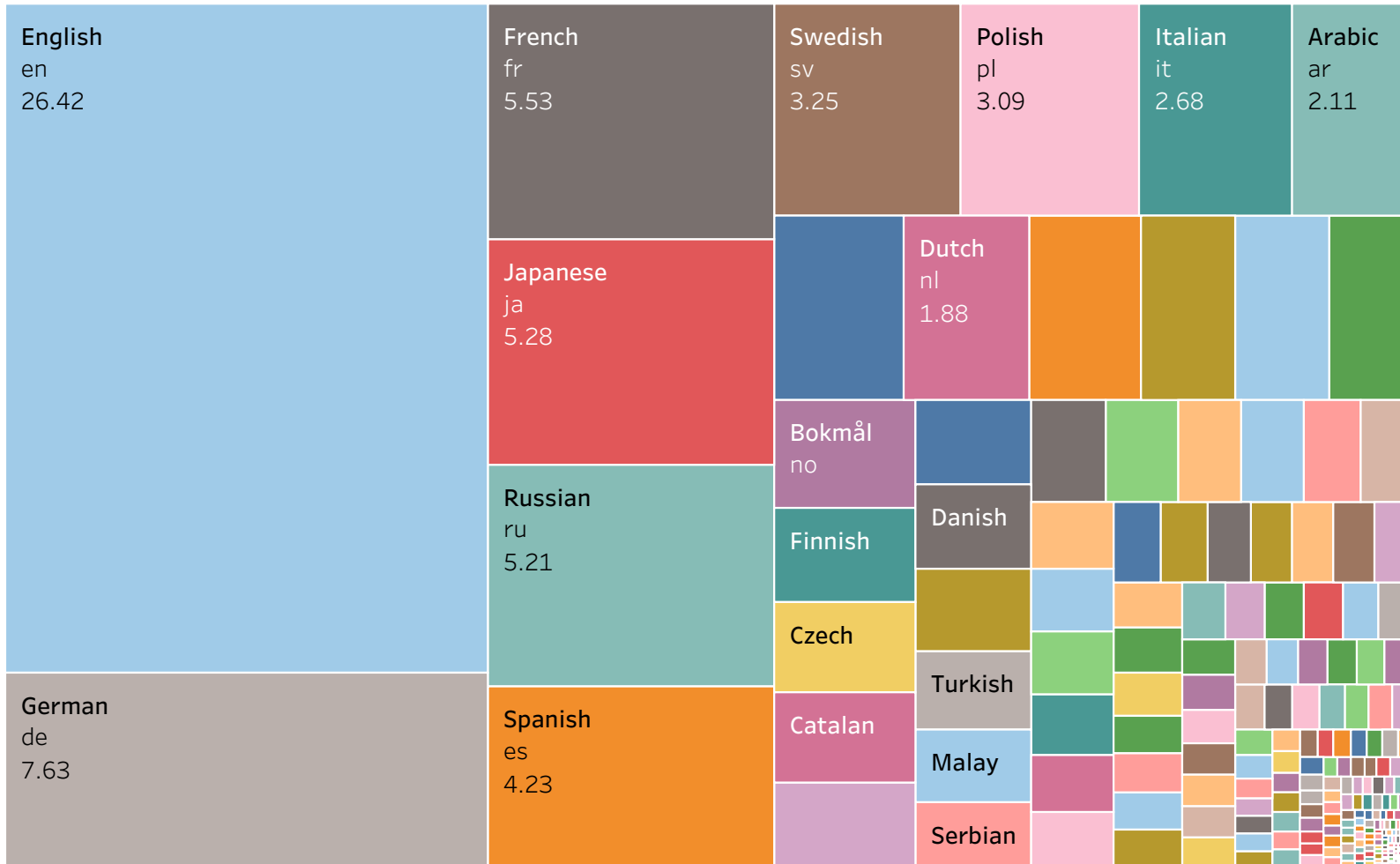
We see that Arabic dedicates many efforts to cover the English Cultural Context.

Coverage of Arabic CCC (%) by other Wikipedia language editions



English 27.68%
 French 15.21%
 Persian 11.28%
 Italian 9.65%
 Spanish 8.90%
 German 8.71%
 ...

What is the impact of Arabic CCC in all Wikipedia languages CCC?



The extent of the Arabic CCC in all Wikipedia languages CCC does not reflect what the Arabic language cultural context means.

2.11% is not enough if we compare it to other smaller contexts (Swedish, Polish, Italian, etc.).

Sum and percentage of Language edition CCC spread to all Wikipedia language editions content

Arabic CCC spread across all the Wikipedias

About a 57.5% of the Arabic CCC is not available in other language editions.

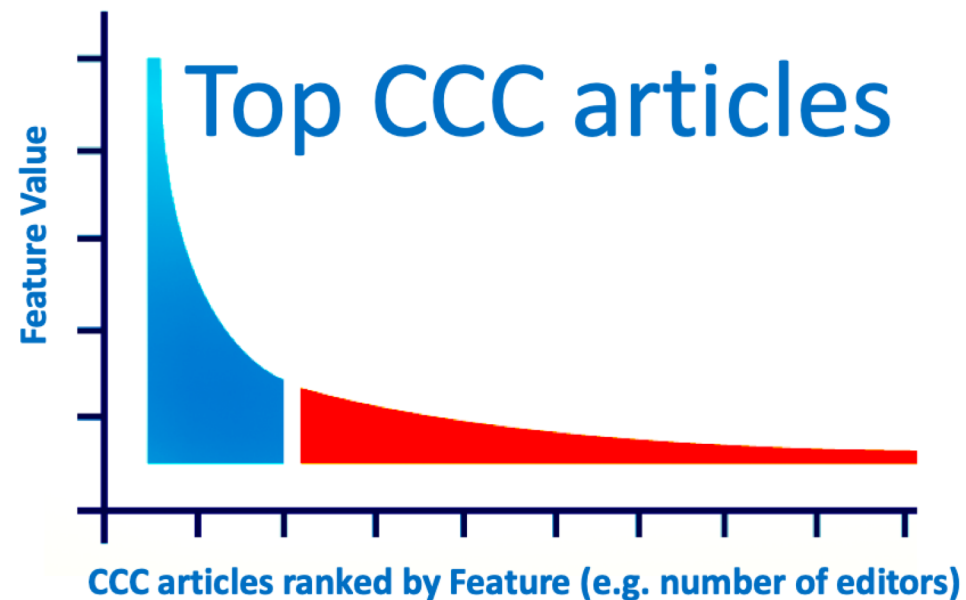
34.4% of Egyptian Arabic CCC

English 55, Chinese 57.9, Russian 34.6,
Spanish 38.7, Japanese 75.29. and Hindi 38.7.



Solution: Top CCC

Since it is impossible to bridge all the knowledge gaps between languages, we propose lists of very relevant articles from each language CCC.



Lists of 100 – 500 articles that should be in every language edition

Solution: Top CCC

From each language, those articles from their cultural context which are the **most relevant according to specific features**:

- List = [editors, featured, geolocated, keywords, women, men, created_first_three_years, created_last_year, pageviews, discussions]
- Country_origin (optional) = ISO3166 code
- Lang_origin = wikicode
- Lang_target = wikicode

http://wcd0.wmflabs.org/top_ccc_articles/?list=men&lang_origin=pl&lang_target=uk

http://wcd0.wmflabs.org/top_ccc_articles

Arabic Top CCC articles list "Women" related to Morocco and its coverage by English Wikipedia

N°	Arabic Article Title	Edits	Editors	Pageviews	Bytes	References	Wikidata Properties	Interwiki Links	Inlinks from CCC	Creation Date	Other Languages	English Article Title
1	ميساء مغربي	605	245	1082	13.0k	14	17	5	39	2008-11-21	es	Mayssa Maghrebi
2	شذى حسون	605	174	263	28.3k	18	13	7	116	2007-03-31	fr	Shatha Hassoun
3	دنيا بطمة	429	172	1046	11.2k	5	8	3	106	2012-03-04		Dounia Batma
4	منى أمرشا	375	185	277	5.3k	2	9	0	91	2009-02-19		Mona Amarcha
5	ابتسام تسكت	354	123	346	12.3k	15	17	3	92	2014-12-26	fr	Ibtissam Tiskat
6	جنات (مغنية)	347	146	881	19.5k	9	13	5	297	2008-01-01	fr	Jannat
7	ونام الححماني	269	113	604	5.2k	8	16	3	5	2012-08-09		Wiam Dahmani
8	عزيزة جلال	265	84	679	31.6k	35	10	3	141	2006-06-30	fr	Aziza Jalal
9	أمل العنبري	264	79	120	9.1k	14	7	0	12	2011-08-16		Amal Anbari
10	سلمى بناتي	225	115	1087	8.4k	11	15	24	128	2007-10-30	fr , de , it , es	Princess Lalla Salma of Morocco
11	حنان الخضر	205	70	233	11.3k	10	7	2	91	2016-02-07		Hanane El Khader
12	أسماء لمنور	187	87	318	6.6k	14	23	4	96	2012-03-09	fr	Asma Lamnawar
13	نوال المتوكل	173	67	85	7.8k	11	26	25	56	2007-05-31	fr , de , it , es	Nawal El Moutawakel
14	مريم سعيد	168	69	331	5.0k	5	5	0	23	2015-04-21		
15	أميناتو حيدر	164	76	29	15.4k	9	17	19	134	2009-12-08	fr , de , it , es	Aminatou Haidar
16	ليلى غفران	162	70	204	14.6k	4	10	3	95	2007-06-23	fr , es	Laila Ghofran
17	دنيا بوتازوت	157	73	158	6.1k	6	9	2	7	2012-07-26	fr	Dounia Boutazout
18	سلمى رشيد	155	67	411	10.9k	2	11	3	99	2014-04-22		Salma Rachid
19	أسماء المرابط	142	9	11	3.6k	10	15	2	5	2015-06-12	fr	Asma Lamrabet
20	ملیكة أوفكير	137	54	204	6.9k	15	25	10	45	2008-09-03	fr , de , it , es	Malika Oufkir

Top 500 CCC articles list "Editors" from Egyptian Arabic CCC to Arabic

In this page, you can consult Top 500 CCC articles list from any country or language CCC generated with the latest CCC dataset. It is updated on a monthly basis (between updates there may be changes and the table may not reflect them).

The following table shows the Top 500 articles list Editors from Egyptian Arabic CCC and its article availability in Arabic Wikipedia. Articles are sorted by the feature **Editors**. The rest of columns present complementary features that are explicative of the article relevance (number of editors, edits, pageviews, Bytes, Wikidata properties or Interwiki links). In particular, number of Inlinks from CCC (incoming links from the CCC group of articles) highlights the article importance in terms of how much it is required by other articles. The column named Other Language present Interwiki links to the article version when available in the four languages closer to the target language (those that cover best this language and therefore it is likely their editors consult it).

The table's last column shows the article title in its target language, in **blue** when it exists, in **red** as a proposal generated with the Wikimedia Content Translation tool or as an existing Wikidata label in the same language, and **empty** when the article does not exist or there is no title proposition available.

The available Top CCC articles lists are: list of CCC articles with most pageviews during the last month (**Pageviews**), list of CCC articles with most number of editors (**Editors**), list of CCC articles created during the first three years and with most edits (**First 3Y**), list of CCC articles created during the last year and with most edits (**Last Y**), list of CCC articles with most edits in talk pages (**Discussions**), list of CCC articles with featured article distinction (**Featured**), most bytes and references (weights: 0.8, 0.1 and 0.1 respectively), list of CCC articles with geolocation with most links coming from CCC (**Geolocated**), list of CCC articles with keywords on title with most bytes (**Keywords**), list of CCC articles categorized in Wikidata as women with most edits (**Women**) and list of CCC articles categorized in Wikidata as men with most edits (**Men**).

It is possible to query any list by changing the URL parameters. You need to specify the list parameter (editors, featured, geolocated, keywords, women, men, created_first_three_years, created_last_year, pageviews and discussions), the language target parameter (as lang_target and the language wikicode), the language origin (as lang_origin and the language wikicode), and, optionally to limit the scope of the selection, the country origin parameter as part of the CCC (as country_origin and the country [ISO3166 code](#)). In case no country is selected, the default is 'all'.

Select the parameters

[Download Table \(Excel\)](#)

List: Language origin: Country origin: Language target:

N°	Egyptian Arabic Title	Editors	Pageviews	Bytes	References	Wikidata Properties	Interwiki Links	Inlinks from CCC	Creation Date	Other Languages	Arabic Title
1	شخصيات مصريه مشهوره	135	225	63.3k	1	2	3	77	2009-05-16	en	قائمة أعلام المصريين
2	الاهلي	134	49	12.2k	6	23	41	48	2009-02-15	en , fr , fa , de	(النادي الأهلي (مصر)
3	مصر	106	437	38.5k	29	126	309	1437	2008-04-05	en , fr , fa , de	مصر
4	نادي الزمالك	103	37	17.1k	6	24	34	47	2008-12-01	en , fr , fa , de	نادي الزمالك
5	القاهره	98	112	12.5k	4	59	216	602	2008-07-18	en , fr , fa , de	القاهرة
6	محمد مرسى	95	35	6.9k	3	36	98	142	2012-06-03	en , fr , fa , de	محمد مرسى
7	اسكندريه	94	102	22.6k	2	60	145	337	2008-07-18	en , fr , fa , de	الإسكندرية

Language edition coverage of other languages Top CCC lists

Languages Top 100 CCC articles lists coverage by Arabic Wikipedia

This page shows some statistics that explain how well Arabic Wikipedia language edition covers the Top 100 of the Top CCC articles lists from other Wikipedia language editions. It is updated on a monthly basis (between updates there may be changes and the table may not reflect them).

These lists are created by ranking the articles according to specific features and sometimes giving them weights. These different features are based on the *article characteristics* (number of Editors) or *content type* (e.g. geolocated articles).

The Top CCC articles lists are: list of CCC articles with most pageviews during the last month (**Pageviews**), list of CCC articles with most number of editors (**Editors**), list of CCC articles with most edits in talk pages (**Discussions**), list of CCC articles created during the first three years and with most edits (**First 3Y. A. Edits**) and list of CCC articles created during the last year and with most edits (**Last Y. A. Edits**), list of CCC articles with featured article distinction (**Featured**), most bytes and references (weights: 0.8, 0.1 and 0.1 respectively), list of CCC articles with geolocation with most links coming from CCC (**Geolocated**), list of CCC articles with keywords on title with most bytes (**Keywords**), list of CCC articles categorized in Wikidata as women with most edits (**Women**) and list of CCC articles categorized in Wikidata as men with most edits (**Men**).

The following table is useful to assess how well Arabic Wikipedia covers the Top 100 CCC articles from the lists generated from all the other language editions CCC. Languages are sorted in alphabetic order by their Wikicode, and columns present the number of articles from each list covered by English language. The last two columns, **Lists Coverage Idx.** and **Sum Covered Articles** present the percentage of articles from the lists covered and the overall sum of articles from the lists covered by Arabic Wikipedia for each language. The maximum Sum Covered Articles can be 1000 (there are 100 in each of the 10 lists), although it may be lower since some lists have articles in common. When Sum Covered Articles has 100 or more articles it will turn green. In some languages, it is not possible to cover them because unfortunately the lists do not have enough articles.

The challenge is to reach 100 articles covered (Sum Covered Articles) from each language CCC!

Select the language

Language

Arabic (ar)

QUERY RESULTS!

Language	Wiki	Page views	Editors	Talk Edits	First 3Y. A. Edits	Last Y. A. Edits	Featured	Geolocated	Keywords	Women	Men	List Coverage Idx.	Sum Covered Articles	World Subregion
Abkhaz	ab	9%	8%	8%	4/9	1/20	9%	3/87	2/13	0/0	2/11	12	9	Western Asia
Acehnese	ace	11%	27%	3%	24%	0%	0%	14%	6%	0/1	2/14	9.9	32	South-eastern Asia
Adyghe	ady	4/12	4/12	4/12	2/3	0/0	4/12	0/0	2/4	0/0	0/3	25	4	Eastern Europe
Afrikaans	af	41%	79%	33%	61%	5%	25%	23%	14%	9%	22%	31.2	149	Sub-Saharan Africa
Akan language	ak	12/57	12/57	12/57	8/16	0/0	12/57	4/29	2/2	1/4	5/13	31.1	12	Sub-Saharan Africa
Alemannic	als	68%	88%	42%	77%	13%	19%	70%	19%	16%	41%	45.3	253	Western Europe
Amharic	am	13%	14%	9%	4/5	1/1	5%	6/98	2/25	0/0	1/8	24.8	16	Sub-Saharan Africa
Aragonese	an	41%	50%	18%	38%	11/61	30%	20%	15%	20%	54%	30.4	183	Southern Europe
Old English	ang	66%	66%	64%	23/37	2/2	65%	19/24	6/7	3/6	26/41	70.1	70	Northern Europe
Arabic	ar	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100	677	Western Asia
Aramaic	arc	14/17	14/17	14/17	4/5	0/0	14/17	5/5	2/5	0/0	1/1	64.9	14	Western Asia

Arabic Wikipedia covers > 100 articles from 96 other language editions Top CCC lists.

There is work to cover a minimum of every other cultural context.

https://wcd0.wmflabs.org/languages_top_ccc_articles_coverage/?lang=ar

Top CCC lists from a language edition spread across others

Arabic Wikipedia Top 100 CCC article lists spread across the rest of Wikipedias

This page shows some statistics that explain how well the first Arabic Wikipedia Top 100 CCC articles are spread across the other language editions. It is updated on a monthly basis (between updates there may be changes and the table may not reflect them).

These lists are created by ranking the articles according to specific features and sometimes giving them weights. These different features are based on the *article characteristics* (number of Editors) or *content type* (e.g. geolocated articles).

The Top CCC articles lists are: list of CCC articles with most pageviews during the last month (**Pageviews**), list of CCC articles with most number of editors (**Editors**), list of CCC articles with most edits in talk pages (**Discussions**), list of CCC articles created during the first three years and with most edits (**First 3Y. A. Edits**) and list of CCC articles created during the last year and with most edits (**Last Y. A. Edits**), list of CCC articles with featured article distinction (**Featured**), most bytes and references (weights: 0.8, 0.1 and 0.1 respectively), list of CCC articles with geolocation with most links coming from CCC (**Geolocated**), list of CCC articles with keywords on title with most bytes (**Keywords**), list of CCC articles categorized in Wikidata as women with most edits (**Women**) and list of CCC articles categorized in Wikidata as men with most edits (**Men**).

The following table is useful in order to assess how known the Top 100 CCC articles from Arabic Wikipedia language are in the entire Wikipedia project. Languages are sorted in alphabetic order by their Wikicode, and columns present the number of articles from each list covered by the language. The last column **Sum Spread Articles** the overall sum of articles from the Top 100 of each list spread across a specific language. The maximum Sum Spread Articles can be 1000 (there are 100 in each of the 10 lists), although it may be lower since some lists have articles in common. When Sum Spread Articles has 100 or more articles it will turn green.

The challenge is to reach 100 articles (Sum Spread Articles) in each other Wikipedia language edition!

Select the language

Language

Arabic (ar) x ▾

QUERY RESULTS!

Language	Wiki	Page views	Editors	Talk Edits	First 3Y. A. Edits	Last Y. A. Edits	Featured	Geolocated	Keywords	Women	Men	Sum Spread Articles	World Subregion
Afar	aa	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0	Sub-Saharan Africa
Abkhaz	ab	5%	4%	4%	4%	0%	7%	2%	3%	0%	2%	11	Western Asia
Acehnese	ace	25%	33%	26%	42%	0%	32%	26%	18%	1%	8%	57	South-eastern Asia
Adyghe	ady	19%	22%	19%	28%	0%	19%	23%	18%	0%	1%	38	Eastern Europe
Afrikaans	af	37%	47%	43%	59%	0%	50%	25%	25%	5%	27%	125	Sub-Saharan Africa
Akan language	ak	5%	5%	4%	6%	0%	3%	5%	3%	0%	1%	9	Sub-Saharan Africa
Alemannic	als	33%	40%	34%	47%	0%	43%	25%	19%	4%	16%	94	Western Europe
Amharic	am	30%	38%	32%	48%	0%	40%	26%	20%	2%	20%	86	Sub-Saharan Africa
Aragonese	an	37%	43%	35%	51%	0%	38%	25%	20%	5%	25%	98	Southern Europe
Old English	ang	25%	31%	23%	33%	0%	26%	20%	17%	3%	9%	53	Northern Europe
Arabic	ar	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	677	Western Asia
Aramaic	arc	24%	32%	29%	36%	0%	28%	22%	19%	0%	9%	57	Western Asia
Egyptian Arabic	arz	63%	87%	71%	89%	1%	61%	27%	35%	40%	79%	265	Northern Africa

Only 42 Wikipedia languages editions cover more than 100 articles from the 10 Arabic Top CCC articles lists.

There is work to do in spreading the most relevant content from Arabic CCC.

Conclusions

1. Use the tools “Top CCC” and “Missing CCC” or create the lists yourselves. Representation and Sharing need to be embedded in your daily activities.
2. Find your path to increase cultural diversity **coverage** (now it follows English). Engage in all sort of contests that improve cultural diversity.

Most important:

2. Give support to local communities in countries **underrepresented** in your CCC.
3. Organize contests to **spread** a minimum of content for every Arabic country.

Create your own model to spread your CCC.

Cross-language article exchange ([Wikimedia CEE Spring](#), [Intercultur Wikimedia España](#)).

One culture spreading to the rest of languages ([Catalan Culture Challenge](#)).

All languages covering one “culture” or context ([Asian Month](#))





wikiARABia **Month?**

Thank you very much!

Marc Miquel, PhD

{marcmiquel@gmail.com}

Username:marcmiquel

Pompeu Fabra University, Barcelona, Catalonia

Amical Wikimedia (Catalan Wikipedia)

Wikimedia WG Diversity 2030

Diversity Observatory



References (if you want to know more)

Miquel-Ribé, M., & Laniado, D. **(2019)**. Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions. *Proceedings of the 13th International AAI Conference on Web and Social Media. ICWSM. ACM.*

Miquel-Ribé, M., & Laniado, D. **(2018)**. Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions. *Frontiers in Physics, 5, 12.* (CC BY) Open Access.

Miquel-Ribé, M. **(2017)**. *Identity-based motivation in digital engagement: the influence of community and cultural identity on participation in wikipedia* (Doctoral dissertation, Universitat Pompeu Fabra).

Miquel-Ribé, M., & Laniado, D. **(2016)**. Cultural identities in wikipedias. In *Proceedings of the 7th 2016 International Conference on Social Media & Society* (p. 24). ACM.