



Observatorio de la diversidad lingüística y cultural en la Internet
Observatoire de la diversité linguistique et culturelle dans l'Internet
Observatory of the Linguistic and Cultural Diversity in the Internet

<https://OBDILCI.ORG/>

**WebMultilingualism Reports #2: Exploring web presence and multilingualism of European minority languages with associated gTLDs.
D. Pimienta, OBDILCI, 25/03/2025**

Table of contents

CONTEXT	2
SCOPE OF THE STUDY	2
METHOD	3
BIASES.....	4
MAIN RESULTS.....	6
CONCLUSION.....	8
ANNEX: DETAILED RESULTS PER LANGUAGE.....	9
CORSICAN	9
BASQUE.....	9
CATALAN	11
GALICIAN	12
LUXEMBURGUISH.....	14
FRISIAN.....	15
WELSH.....	16
IRISH	18
SCOT	19
BRETON.....	20
ALSACE.....	21
Table 1: List of languages and gTLDs processed.....	2
Table 2: List of some 2 letters language iso 639 codes	4
Table 3: Language detection bias	4
Table 4: Data collected from Dataprovider.com.....	6
Table 5: Indicator built from the collected data.....	7
Table 6: Evaluation	7
Table 7: Synthesis of results for gTLDs	7

CONTEXT

Dataprovider.com maintains a database gathering information related to a series of websites close to the total WWW (80% if compared with the figure given by Netcraft: 872 920 299 over 1 101 431 853). Among the many information kept for each web site, there is the unique main language of the site (often taken from the home page) and the presence of hreflang tags, which a proportion of sites uses to list the linguistic versions of the site with the format hreflang= ll; ll-cc;... where ll is the ISO639-2 and cc a country code with 2 digits¹.

Dataprovider.com has offered a courtesy access of the database to allow OBDILCI to develop useful statistics on the subject of languages online, in coherence with its non-for-profit mission. This report is the second of a series of report of the findings of OBDILCI obtained thanks to Dataprovider.com database.

The data base of Dataprovider.com is extremely impressive, both by its uncommon extraordinary reach, a large proportion of the whole WWW, close to completion, and the ease of use, friendliness and powerfulness of the interface which allow searches using close to 100 different parameters with unlimited combinations and immediate response.

SCOPE OF THE STUDY

A number of minority languages from European countries have followed the drive of the Catalans who pioneer in 2012 the creation of a Generic Top-Level Domain dedicated to their language, with the objective to foster content creation in Catalan. Some of the initiatives focus the language, others focus the associated region (geographic TLD) and others have been targeting both the region and the language. The list of the languages and associated domains this study will focus is described in the following table:

Table 1: List of languages and gTLDs processed

gTLD DOMAIN	LANGUAGE	FOCUS
.corsica	Corsican	Geographic+linguistic
.eus	Basque	Linguistic
.cat	Catalan	Linguistic
.gal	Galician	Linguistic
.lu	Luxemburgish	ccTLD
.frl	Frisian	Geographic+linguistic
.gales	Welsh	Geographic
.cymru	Welsh	Linguistic
.irish	Irish	Geographic+linguistic
.scots	Scots	Geographic+linguistic
.bzh	<i>Breton</i>	Geographic+linguistic
.alsace	Alsatian (German Swiss)	Geographic

¹ See https://en.wikipedia.org/wiki/List_of_ISO_639_language_codes

METHOD

The Dataprovider.com language detection algorithm allows recognition of 77 languages including all the mentioned minority languages, with the exception of Alsatian and Breton. The database allows queries for all existing TLDs. The use of the collection of `hreflang=` instructions allows the detection of all second languages listed, including Breton and Alsatian. The method consists in collecting as much as possible information from the database concerning the mentioned languages and gTLDs. The following data will be collected (in parenthesis the corresponding label in table 3 and table 4 below):

- Number of **web sites with the given language** as main language (first line # Sites in first table, W@L in the second table).
- **Repartition of languages** in the multilingual web sites with main language selected (1st line Second Lang.)
- **Percentage of selected language** in the multilingual web sites (%2L)
- **Rate of multilingualism** for each main language web sites (1st line Multi%). This figure is approximated by counting the percentage of web sites using the instruction `hreflang=` and dividing this figure by 0.40, so to extrapolate the Multi% on the assumption that 40% of multilingual web sites use that technic to specify the different second languages.
- Number of **web sites in the given gTLD** (Second line # Sites in first table, W@D in the second table).
- **Repartition of languages** in the web sites in the given gTLD (Main Lang.).
- **Rate of multilingualism** for each gTLDs (2nd line Multi%). This figure is approximated by counting the percentage of web sites of each gTLD using `hreflang` and dividing the figure by 0.40.
- **Repartition of languages** from `hreflang` tags for the gTLD and focus on the treated language (2nd line Second Lang. in first table).

From this set of data, completed by the number of L1+L2 speakers by language (source Ethnologue), a series of indicators is built and presented in table 4:

L Virt. Pres. = Virtual Presence of the language, the number of web sites (W@L) divided per the value of L1+L2 speakers and multiplied by 100. The higher the value the best virtual presence of the language.

D Virt. Pres. = Virtual presence of the domain, the number of domains below the gTLD (W@D) divided per the value of L1+L2 speakers and multiplied by 100. The highest the value the most successful the domain.

Multi% = The average value between the 2 values of rate of multilingualism (one for the language, one for the gTLD). For the case of the languages without data (Alsatian and Breton), the unique value is taken.

L-Pres. = The average between the 3 collected values of percentage of presence of the selected language in different context (a unique value for Alsatian and Breton).

Ling. PERFORM = A global indicator of the linguistic performance² of the language computed as: $100 \times \text{Mean (L Virt. Pres. and D Virt. Pres.)} \times \text{Multi\%} \times \text{L-Pres.}$

As a convention to highlight some results, in each table the best values are in red, and bold for the best one, the lowest values are in blue, in bold for the lowest, black means average value. Bold is also used to highlight the selected language when a list of languages is exposed, either by its English name, either by its ISO639 2 letters code.

Table 2: List of some 2 letters language iso 639 codes

BR	Breton
CA	Catalan
CO	Corsican
CY	Welsh
DE	German
EL	Greek
EN	English
ES	Spanish
EU	Basque
FR	French
FY	Frisian
GA	Irish
GL	Galician
GD	Scots
IT	Italian
JP	Japanese
LA	Latin
LB	Luxemburgish
NL	Dutch
PT	Portuguese
VE	Vietnamese

BIASES

This work is not exempt of biases and they need to be understood prior to use the figures.

- Language detection biases.

We have manually checked a random sample of some 20 websites to audit the results of the detection algorithm. The results are summarized in the table below.

Table 3: Language detection bias

LANGUAGE	DETECTION BIAS
Alsatian	NOT DETECTED
Basque	MEDIUM LOW (<20%) Confusion with Dutch. Test under .eus excellent
Breton	NOT DETECTED
Catalan	VERY LOW (< 10%)

² In no way this indicator weights the quality of the websites or its capacity to promote the geographic location, it exclusively focuses the linguistic characteristics.

Corsican	EXTREME Confusion with FR,IT,RO,RU. However, the test under .corsica rated excellent.
Frisian	EXTREME however excellent under .frl
Galician	MEDIUM (<30%) confusion with Spanish and Portuguese. Excellent under .gal
Irish	MEDIUM but only 2 of the 5 detected Irish under .irish are Irish
Scots	HIGH mainly confuse English however excellent under .scot
Luxemburgish	EXTREME confusion with Dutch and English However excellent under .lu
Welsh	EXTREME Confusion with EN, HE, EL. However, the test below .cymru rated excellent

Given the exposed biases, the language data for Corsican, Frisian, Luxemburgish and Welsh will not be retained for the study and those languages will be processed the same way as Breton and Alsatian. Note that this bias in no way affect the language data for the domains which relies in another technic.

- Hreflang bias.

The assumption that the proportion of websites using the instruction hreflang= as the way to specify the list of languages used, is around 40% of the total, and from there extrapolating the total maybe true in average, yet this figure may change sensibly in different countries due to local computing culture and therefore may provoke a bias in the extrapolation in some cases.

- Hreflang errors

Some web designers using hreflang are not respecting the ISO639 recommendation³ and are not using the right codification. As an example, some Luxemburgish sites use LU instead of LB. The data base report a number of uncommon codes under the label Other. When the percentage of “other” is very high this is the reflect of possible errors and it has as a consequence to penalize the percentages we are collecting. This is mentioned in the table with the symbol O, when the O is bold the figure is medium when it is red it is high and when it is red and bold the percentage could be higher than 75%.

- Hreflang collection.

Hreflang are specified as a series like for example: EN, FR-FR, FR-CA, ES, where the second optional factor is the country code. For language studies, the second term is useless and, in the example, we would only keep EN,FR,ES to avoid counting twice FR. The database keeps the

³ https://en.wikipedia.org/wiki/List_of_ISO_639_language_codes

total of occurrence of each couple so the information we need to focus exclusively the language part is lost. It is possible to cumulate all the occurrences of the same language on the final count and we have a macro Excel doing this. However, the result does not reflect the reality that we look for but a probable overestimation of the number of websites in the strongest language as some web designer does not hesitate for a single site to repeat en-us, en-uk, en-au.... many times.

MAIN RESULTS

The results are summarized in the following 3 tables:

- ✓ Data collected from Dataprovider.com
- ✓ Indicators built from the collected data
- ✓ Synthesis of results for gTLDs

Table 4: Data collected from Dataprovider.com

MAIN LANG /TLD	# Sites	MULTI%	MAIN Lang.	% LANG	Second Lang.	% 2L
Corsican						
.corsica	836	25,7%	FR,EN,CO	2,5%	FR,EN,IT,CO	8,0%
Basque	7 321	71,3%			O,EU,ES,EN,FR	24,3%
.eus	2 580	100,0%	ES,EU,EN,FR	24%	EU,ES,EN,FR	40,4%
Catalan	75 463	55,9%			CA,ES,EN,O,FR	33,8%
.cat	32 365	50,5%	CA,ES,EN,FR	71,8%	CA,ES,EN,FR	36,6%
Galician	22 793	21,7%			O,ES,EN,GL,FR,PT	6,0%
.gal	2 542	59,5%	GL,ES,EN,PT	65,7%	GL,ES,EN,PT,CA	39,3%
Luxemburgish						
.lu	22 144	56,8%	FR,EN,DE,LB	5,0%	EN,FR,DE,NL,ES,LB	2,8%
Frisian						
.frl	1 323	15,5%	NL,FY,EN,FR,DE	1,2%	NL,EN,FY,DE,FR	7,8%
Welsh						
.gales	2 696	27,9%	EN,CY	2,4%	EN,CY	40,0%
.cymru	1 877	63,3%	CY,EN	44,8%	CY,EN	28,7%
Irish	2 052	33,0%			O, EN,GA,ES,FR,DE	2,1%
.irish	362	9,7%	EN,VE,GA,NL,FR,JP	1,4%	EN,DE,NL,IT	0,0%
Scottish gaelic						
.scots	3 235	5,3%	EN,GD,FR	0,5%	EN,DE,ES,FR,GD,NL	5,7%
.bzh	4 902	17,5%	FR,EN,LA,LB	NA	FR,EN,BR,DE,ES,IT,NL	9,3%
.alsace	620	23,8%	FR,EN,DE	NA	FR,EN,DE,ES,NL	0,0%

Table 5: Indicator built from the collected data

	Speakers L1+L2 ⁴	W@L	W@D	L- VirtPres.	D- Virt.Pres.	Multi%	L-Pres	LINGU. PERFORM
.corsica	121 000		836		0,69	15,4%	5,7%	0,93
.eus	1 086 860	7 321	2 580	0,67	0,24	85,6%	38,0%	14,84
.cat	9 299 570	75 463	32 365	0,81	0,35	53,2%	47,4%	14,63
.gal	3 378 300	22 793	2 542	0,67	0,08	40,6%	37,0%	5,63
.lu	529 300		22 144		4,18	32,9%	5,1%	9,27
.frl	720 840		1 323		0,18	13,4%	6,8%	0,13
.gales	548 490		2 696		0,49	32,2%	14,4%	2,91
.cymru	548 490		1 877		0,34	49,9%	24,8%	7,96
.irish	1 199 280	2 052	362	0,17	0,03	21,4%	1,2%	0,03
.scot	1 599 200		3 235		0,20	6,4%	7,1%	0,03
.bzh	206 000		4 902		2,38	17,5%	9,3%	3,89
.alsace	6 463 630		620		0,01	23,8%	0,0%	0

Table 6: Evaluation

	L- VirtPres.	D- Virt.Pres.	Multi%	L-Pres	LING. PERFORM
.corsica		FAIR	FAIR	LOW	LOW
.eus	FAIR	LOW	CHAMP.	HIGH	CHAMP.
.cat	MEDIUM	MEDIUM	VERY HIGH	CHAMP	CHAMP.
.gal	FAIR	VERY LOW	VERY HIGH	HIGH	MEDIUM-H
.lu		CHAMP	HIGH	LOW	HIGH
.frl		VERY LOW	LOW	LOW	VERY LOW
.gales		MEDIUM	FAIR	MEDIUM-L	MEDIUM
.cymru		MEDIUM	VERY HIGH	MEDIUM-H	MEDIUM-H
.irish	VERY LOW	VERY LOW	MEDIUM	VERY LOW	NULL
.scot		LOW	VERY LOW	LOW	NULL
.bzh		VERY HIGH	MEDIUM	LOW	MEDIUM-L
.alsace		VERY LOW	MEDIUM-H	NULL	NULL

Table 7: Synthesis of results for gTLDs

SYNTHESIS	
.corsica	Strong geographic domain - Weak linguistic domain
.eus	Top linguistic domain - top multilingual domain - low presence
.cat	Top linguistic and multilingual domain - room for presence growth
.gal	Strong linguistic domain - top multilingual domain - very low presence
.lu	ccTLD domain with solid multilingualism - weak linguistic presence
.frl	Geographic domain with low presence - weak linguistic presence and multilingualism
.gales	Solid geographic domain
.cymru	Good linguistic domain with solid bilingualism
.irish	Weak geographic domain in every factor except multilingualism
.scot	Weak geographic domain - weak in every factor
.bzh	Successful geographic domain - Weak linguistic domain - Low multilingualism
.alsace	Pure geographic domain, very weak presence, absence of Alsatian, fair multilingualism

⁴ Source Ethnologue 2024.

CONCLUSION

Two gTLDs show **excellent linguistic performances**, both in terms of presence of the minority language and in terms of multilingualism: **.cat** and **.eus**; both, specially **.eu**, have room to increase their presence.

Two gTLD, **.cymru** and **.gal**, show **fair linguistic performance**, together with **.lu**. The case of **.lu**, the ccTLD of Luxemburg, is to be treated separately, the study highlighting the low presence of the national language in a context of strong multilingualism.

Three gTLDs show medium results with promises in some of the factors and difficulties in others: **.gal**, **.gales** and **.bzh**.

Two gTLDs, **.corsica** and **.frl** has not reached the point of benefiting the local language nor multilingualism, and seems to remain in a stage of geographic TLD.

Finally, **.alsace**, **.irish** and **.scot** does not show linguistic performance and remain geographic domain with very limited penetration and linguistic impact at this stage.

ANNEX: DETAILED RESULTS PER LANGUAGE

CORSICAN

The two first tables concern sites with Corsican as main languages, the two last tables concern the sites belonging to .corsica.

Results .corsica		
Sites 836		
Name	Count	Percentage
French	652	90,6%
English	45	6,3%
Corsican	18	2,5%
Latin	2	0,3%
German	1	0,1%
Irish	1	0,1%
Dutch	1	0,1%

			MULTI%
href	86	10,29%	25,72%
Hrefgrouped			
Name	Count	Percentage	
fr	106	44,5%	
en	63	26,5%	
it	24	10,1%	
co	19	8,0%	
de	14	5,9%	
es	6	2,5%	
nl	3	1,3%	
ca	1	0,4%	
pt	1	0,4%	
ru	1	0,4%	

Under .corsica domain, there is 836 sites, more than 90% in French, and only 2.5% in Corsican. The % of multilingual sites is a fair 26%, with Corsican in fourth position behind French, English and Italian and before German and Spanish. The rate of multilingualism is on the low range. The manager of the .corsica domain could stand in a fair presence to develop initiatives to foster contents in Corsican and promote multilingualism.

BASQUE

Sites in Basque			
		7321	MULTI%
href	2087	28,51%	71,27%
Href grouped		7818	

Name	Count	Percentage
Other	2184	27,9%
es	1933	24,7%
eu	1899	24,3%
en	484	6,2%
fr	288	3,7%
de	141	1,8%
nl	60	0,8%
it	53	0,7%
pt	50	0,6%
ca	48	0,6%
pl	35	0,4%
ru	28	0,4%
da	20	0,3%
zh	20	0,3%
ar	19	0,2%
sv	18	0,2%
+ 73 other languages		6,9%

Sites in .eus 6097

Name	Count	Percentage
Basque	2846	49,0%
Spanish	2466	42,4%
English	310	5,3%
French	148	2,5%
Galician	14	0,2%
Latin	14	0,2%
German	4	0,1%
Greek	2	0,0%
Polish	2	0,0%
Portuguese	2	0,0%
Russian	2	0,0%
Afrikaans	1	0,0%
Catalan	1	0,0%
Esperanto	1	0,0%
Chinese	1	0,0%

href	Count	Percentage	MULTI%
href	2580	42,32%	105,79%
Href			
Grouped	6060		
Name	Count	Percentage	
es	2477	40,9%	
eu	2451	40,4%	
en	596	9,8%	

fr	411	6,8%
ca	43	0,7%
de	17	0,3%
ja	9	0,1%
gl	7	0,1%
it	8	0,1%
pt	8	0,1%
ru	6	0,1%
ar	4	0,1%
+ 17 others		0,4%

There are 7321 sites which main language is Basque, with a very high rate of multilingualism of 71%. For those multilingual sites the main languages are Spanish and Basque very close, followed by English and French, then Galician and Latin (sites under construction?). There is some bias which may reduce the percentage of Basque in multilingual sites due to the existence of 30% of languages classified as other. In terms of global performance, Basque is a co-winner of the gold medal, being the Champion of multilingualism, way above the rest. The managers of .eus just need to get more customers using the same multilingual and local language respectful approach, to compensate a still weak presence.

CATALAN

Sites en catalan	75463		MULTI%
href	16876	22,36%	55,91%
Href grouped			
	47022		
Name	Count	Percentage	
ca	15877	33,8%	
es	15868	33,7%	
en	5954	12,7%	
Other	2465	5,2%	
fr	2339	5,0%	
de	705	1,5%	
pt	380	0,8%	
it	340	0,7%	
nl	214	0,5%	
ru	180	0,4%	
46 others		5,7%	

Sites in .cat	33397		
	Name	Count	Percentage
Catalan	23252	71,8%	
Spanish	5685	17,6%	
English	2942	9,1%	
French	102	0,3%	
Latin	71	0,2%	

Czech	60	0,2%
German	31	0,1%
Chinese	31	0,1%
Portuguese	27	0,1%
Russian	26	0,1%
Japanese	22	0,1%
Galician	19	0,1%
Dutch	11	0,0%
Vietnamese	10	0,0%
+ 32 Other lang.		0,2%

			MULTI%
href	Count	Percentage	
href	6749	20,21%	50,52%
Name	Count	Percentage	
ca	6376	36,6%	
es	6331	36,3%	
en	2605	15,0%	
fr	861	4,9%	
de	240	1,4%	
it	118	0,7%	
pt	113	0,6%	
nl	84	0,5%	
ru	67	0,4%	
eu	55	0,3%	
gl	43	0,2%	
zh	38	0,2%	
33 others	38	2,8%	

There are 75,463 websites whose primary language is Catalan. The percentage of multilingual sites among those sites is 56%. For those multilingual sites, the primary languages are: Catalan, Spanish, English, and French. The large number of languages is notable. Under the .cat domain, there are 33,397 websites, 72% in Catalan, 18% in Spanish, followed by English and French. The multilingualism rate is 51%. For multilingual sites under .cat, Catalan is slightly above Spanish, ahead of English and French. In terms, of global performance, .cat the historical creator of the concept of linguistic gTLD, is a gold medal champion with first contents in Catalan associated with a very high level of multilingualism. Presence is fair but there is plenty of room to enhance.

GALICIAN

Sites in Galician			MULTI%
href	Count	Percentage	
href	22793	8,67%	21,67%
Href grouped	20882		
Name	Count	Percentage	
Other	15227	72,9%	
es	2105	10,1%	

en	1452	7,0%
gl	1243	6,0%
fr	175	0,8%
pt	151	0,7%
de	105	0,5%
it	72	0,3%
ca	54	0,3%
nl	42	0,2%
eu	35	0,2%
sv	24	0,1%
da	17	0,1%
ru	16	0,1%
pl	14	0,1%
ar	13	0,1%

Note: The huge proportion of “other” affect the percentage of Galician which is probably higher. It appears at the same time that a consistent proportion of sites detected in Galician are Spanish sites.

Résultats .gal		
	2542	11,15%
Name	Count	Percentage
Galician	1596	65,7%
Spanish	688	28,3%
English	90	3,7%
Portuguese	35	1,4%
Latin	9	0,4%
German	4	0,2%
Polish	3	0,1%
Russian	2	0,1%
Bosnian	1	0,0%
Catalan	1	0,0%
French	1	0,0%
Irish	1	0,0%

MULTI%			
href	Count	Percentage	
gl	594	39,3%	
es	593	39,2%	
en	223	14,8%	
pt	30	2,0%	
ca	14	0,9%	
eu	12	0,8%	
fr	17	1,1%	
va	9	0,6%	
			59,50%

de	8	0,5%
it	4	0,3%
+ 7 lang		0,5%

There are 22793 web sites in Galician of which the percentage of multilingual websites is 22%. For multilingual websites, the main languages are: Spanish, English, Galician and French with an extremely high “other” indicator. That part of data is therefore to be taken with caution, biases are important. Under the .gal domain, there are reportedly 2 542 websites with a rate of multilingualism of 60%, more than 65% for Galician and 28% for Spanish, followed by English and Portuguese. For multilingual sites under .gal, Galician and Spanish have an equal proportion around 40%, followed by English, Portuguese, Catalan, Basque and French. The performance in terms of multilingualism and language presence is comparable to that of the winners, but the presence, specially under the .gal domain, does not match the expectation for a speaker’s population of more than 3 million.

LUXEMBURGUISH

Résultats .lu		
22144		
Name	Count	Percentage
French	11057	53,1%
English	5475	26,3%
German	2748	13,2%
Luxembourgish	1040	5,0%
Dutch	108	0,5%
Portuguese	50	0,2%
Polish	47	0,2%
Latin	36	0,2%
Italian	30	0,1%
Russian	28	0,1%
Japanese	27	0,1%
Chinese	26	0,1%
Spanish	20	0,1%
Czech	15	0,1%
+40 others		0,5%

			MULTI%
href	Count	Percentage	
	5032	22,72%	56,81%
Name			
en	5587	23,95%	
fr	5580	23,92%	
de	3627	15,55%	
Other	3417	14,65%	
nl	810	3,47%	
es	665	2,85%	
pt	415	1,78%	

it	414	1,77%
lb	386	1,65%
lu	269	1,15%
+26		9,3%

Under .lu, the percentage of multilingual sites is 57%. French and German share first place, followed by Dutch before Luxembourgish (Note the frequent use of lu, an invalid code, instead of lb the one set by ISO693). The case of .lu, the ccTLD of Luxemburg, cannot be compared with the rest of gTLDs of the study. It remains interesting to see that cyber-Luxemburg is somehow coherent with the country in the sense of being a large international community with dominance of French, English, German, Dutch and Portuguese. The space let for Luxembourgish, the national language of the country seems very narrow in regard to the fact that this is the mother tongue of more than half citizen of the country. In any case, the particular linguistic characteristics of the lu domain mirrors the fact Luxemburg is the most multilingual country of Europe.

FRISIAN

The Frisian languages are a group of West Germanic languages, spoken by about 400,000 Frisian people, who live on the southern fringes of the North Sea in the Netherlands and Germany. the language is thought to be used as a first language by all in the ethnic community. It is used as a language of instruction in education. The .frl geographic domain is targeted at Frisian people living in Friesland, a province of the Netherlands located in the country's northern part, Frisians living in other part of the Netherlands, and Frisians living abroad and wanting to identify themselves with the Frisian culture and heritage.

Résultats .frl		
1323		
Name	Count	Percentage
Dutch	1045	83,1%
Frisian	98	7,8%
English	84	6,7%
French	15	1,2%
German	8	0,6%
Latin	2	0,2%
Afrikaans	1	0,1%
Spanish	1	0,1%
Maltese	1	0,1%
Russian	1	0,1%
Sundanese	1	0,1%
Vietnamese	1	0,1%

			MULTI%
href	82	6,20%	15,50%
Href regroupés	194		
Name	Count	Percentage	

nl	81	41,8%
en	50	25,8%
fy	22	11,3%
de	19	9,8%
fr	9	4,6%
hi	2	1,0%
+ 8 other		5,0%

Under the .fls domain, there are 1,323 websites, more than 80% in Dutch less than 8% in Frisian, followed by English, French and German. The multilingualism rate below .frl is 16%, with Frisian in third place behind Dutch and English, ahead of German and French. The results show that .frl is a geographic domain with low performance for Frisian and multilingualism and with large room for growth.

WELSH

Two domains are associated with this language spoken in Wales, one with geographic orientation, .wales, and one with linguistic orientation .cymru.

Results .wales		
2696		
Name	Count	Percentage
English	2477	95,9%
Welsh	62	2,4%
Vietnamese	17	0,7%
Indonesian	4	0,2%
Latin	4	0,2%
Polish	3	0,1%
Thai	3	0,1%
Arabic	2	0,1%
German	2	0,1%
Japanese	2	0,1%
French	1	0,0%
Italian	1	0,0%
Lithuanian	1	0,0%
Romanian	1	0,0%
Russian	1	0,0%
Chinese	1	0,0%

Wales			MULTI%
href	301	11,16%	27,91%
Href regroupés	675		
Name	Count	Percentage	
en	329	48,7%	
cy	270	40,0%	
de	11	1,6%	

es	10	1,5%
fr	9	1,3%
nl	5	0,7%
it	5	0,7%
zh	4	0,6%
ja	2	0,3%
ar	2	0,3%
pt	2	0,3%
+ 25		3,7%

Results .cymru

1877

Lang

Name	Count	Percentage
English	960	53,8%
Welsh	799	44,8%
Vietnamese	13	0,7%
Indonesian	7	0,4%
Danish	1	0,1%
German	1	0,1%
Hungarian	1	0,1%
Latin	1	0,1%
Sundanese	1	0,1%

Href

Name	Count	Percentage
grouped	2111	100,0%
Other	1000	47,4%
cy	606	28,7%
en	501	23,7%
de	3	0,1%
br	1	0,0%

Under the .wales domain, there are 2,696 sites, 96% in English, 2.4% in Welsh, followed by Vietnamese (gaming sites). The multilingualism rate is 28% for multilingual sites in .wales, with 49% in English and 40% in Welsh, followed by German, Spanish and French. The huge proportion of “others” drops the credit to those values.

Under the .cymru domain, there are 1,877 sites, 54% in English, 45% in Welsh, followed by Vietnamese. The multilingualism rate is 63% for multilingual sites in .cymru, with 29% in Welsh and 24% in English. Again, a high value of the other factor drops credibility to those results. The observation of language detection in Welsh show a large number of errors removing trust to those figures.

In any case, the results comfort the perception that .cymru has a linguistic focus, while .wales has a geographical focus and within the reservations due to strong biases, it seems that .cymru is a successful language gTLD with a fair level of bilingualism.

IRISH

Sites en irish	2052		MULTI%
Href	271	13,21%	33,02%
Href grouped			
Name	Count	Percentage	
Other	3600	70,0%	
en	1194	23,2%	
ga	110	2,1%	
es	63	1,2%	
fr	52	1,0%	
de	50	1,0%	
it	15	0,3%	
ar	12	0,2%	
nl	9	0,2%	
hu	8	0,2%	
ro	7	0,1%	
bg	6	0,1%	
cs	6	0,1%	
da	6	0,1%	
el	6	0,1%	

Results .irish			
	387		
Name	Count	Percentage	
English	323	89,2%	
Vietnamese	10	2,8%	
Irish	5	1,4%	
Indonesian	4	1,1%	
French	3	0,8%	
Japanese	3	0,8%	
German	2	0,6%	
Latin	2	0,6%	
Polish	2	0,6%	
Thai	2	0,6%	
Spanish	1	0,3%	
Persian	1	0,3%	
Italian	1	0,3%	
Norwegian	1	0,3%	
Turkish	1	0,3%	
Chinese	1	0,3%	

			MULTI%
href	15	3,88%	9,69%
Hrefgrouped	41		
Name	Count	Percentage	

en	24	58,5%
de	5	12,2%
nl	3	7,3%
it	2	4,9%
ar	1	2,4%
es	1	2,4%
fi	1	2,4%
fr	1	2,4%
nb	1	2,4%
pt	1	2,4%
zh	1	2,4%

There are 2,052 sites whose primary language is Irish. The percentage of multilingual sites in Irish is 33%. For multilingual sites, the primary languages are English, Gaelic, Spanish, and French, with 70% others. Under the Irish domain, there are 387 sites, more than 70% classified as other, with Vietnamese second to English at 89% and only 5 in Gaelic Irish (2.5%). The percentage of multilingual sites is 10%, with Gaelic absent. Vietnamese sites are games. This domain seems to have stayed in a very early stage with a presence of Iris extremely limited and a poor multilingualism. Note however that the data is strongly biased with high error rate on language detection and huge “other” parameter. The linguistic performance is very poor. This is purely a geographic domain.

SCOT

Results .scot		
3235		
Name	Count	Percentage
English	2990	98,5%
Scottish Gaelic	16	0,5%
Latin	9	0,3%
German	4	0,1%
French	4	0,1%
Polish	4	0,1%
Czech	2	0,1%
Indonesian	2	0,1%
Danish	1	0,0%
Japanese	1	0,0%
Dutch	1	0,0%
Portuguese	1	0,0%
Vietnamese	1	0,0%

			MULTI%
href	Count	Percentage	
href	68	2,10%	5,26%
Hrefgrouped	261		
en	86	33,0%	
de	23	8,8%	

es	21	8,0%
fr	16	6,1%
gd	15	5,7%
nl	7	2,7%
pl	6	2,3%
sv	6	2,3%
pt	6	2,3%
da	5	1,9%
ja	5	1,9%
it	5	1,9%
+26		23,0%

Under the Irish domain, there are 3235 sites, 99% in English, followed by Scots, Latin, German and French. The percentage of multilingual sites is very low at 5%, with Gaelic at 6%, behind English, German, Spanish and French and a balanced number of languages. Same as Irish, the linguistic performance is very poor, this is purely a geographic domain.

BRETON

Résultats .bzh		
	4902	
	4529	100,0%
Name	Count	Percentage
French	4365	96,4%
English	148	3,3%
Latin	4	0,1%
Luxembourgish	4	0,1%
Spanish	2	0,0%
Hungarian	2	0,0%
Danish	1	0,0%
German	1	0,0%
Norwegian	1	0,0%
Sundanese	1	0,0%

			MULTI%
href	Count	Percentage	17,49%
href	343	7,00%	
Hrefgrouped	856		
Name	Count	Percentage	
fr	361	42,2%	
en	204	23,8%	
br	80	9,3%	
de	59	6,9%	
es	28	3,3%	
it	20	2,3%	
nl	13	1,5%	
ru	7	0,8%	
zh	7	0,8%	

bz	6	0,7%
ja	6	0,7%
pt	6	0,7%
cs	5	0,6%
da	5	0,6%
el	5	0,6%
no	5	0,6%
sv	5	0,6%
go	5	0,6%
+26		3,4%

Under the .bzh domain there is close to 5 000 sites; the percentage of multilingual sites is 17% and none appear in Breton as main language. 96% are in French, and 3% in English. Breton appears in 9% of multilingual sites, in third position after French and English and before German, Spanish and Italian with a large number of languages behind. The linguistic performance remains poor leaving .bzh as a geographic domain with the best penetration if one excludes the ccTLD .lu.

ALSACE

Résultats .alsace		
	620	
	552	
Name	Count	Percentage
French	526	95,3%
English	21	3,8%
German	5	0,9%

			MULTI%
Href	59	9,52%	23,79%
Hrefgrouped	160		
Name	Count	Percentage	
Fr	59	36,9%	
En	42	26,3%	
De	38	23,8%	
es	7	4,4%	
nl	5	3,1%	
it	3	1,9%	
al	1	0,6%	
ja	1	0,6%	
ko	1	0,6%	
pt	1	0,6%	
ru	1	0,6%	
zh	1	0,6%	

There are 620 web sites under the .alsace with a tiny language repartition between French (95%), English and German. The rate of multilingualism is fair with 24% but there is a unique site in Alsatian. This is definitively a geographic TLD with a very low penetration.