



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/>

Rapport sur le multilinguisme de la Toile #1 : Approximation du taux de multilinguisme de la Toile

CONTEXTE

DataProvider.com gère une base de données rassemblant des informations relatives à un ensemble de sites web représentant près de la totalité de la Toile (80 % si l'on compare au chiffre donné par Netcraft : 872 920 299 sur 1 101 431 853). Parmi les nombreuses informations conservées pour chaque site web, on trouve la langue principale unique du site (souvent tirée de la page d'accueil) et la présence de balises hreflang, qu'une partie des sites utilise pour indiquer toutes les versions linguistiques et qui utilise le format hreflang= ll ; ll - pp;... où ll est le code langue selon la norme ISO639-2 et pp un code pays à 2 chiffres¹.

DataProvider.com a offert un accès de courtoisie à sa base de données, afin de permettre à l'OBDILCI de développer des statistiques utiles sur les langues en ligne, conformément à sa mission à but non lucratif. Ce rapport est le premier d'une série de rapports sur les résultats obtenus grâce à la base de données de DataProvider.com.

La base de données de DataProvider.com est extrêmement impressionnante, à la fois par sa portée extraordinaire et peu commune, une grande proportion de l'ensemble des sites du WWW, et par sa facilité d'utilisation, la convivialité et la puissance de l'interface qui permet des recherches en utilisant près de 100 paramètres différents avec des combinaisons illimitées et une réponse immédiate.

PORTÉE DE L'ÉTUDE

Un indicateur clé, encore méconnu, de la diversité linguistique dans l'Internet est le **taux de multilinguisme (TM)**, défini comme le rapport entre le nombre total de versions linguistiques et le nombre total de sites web (autrement dit, le nombre moyen de langues par site web). Les deux indicateurs suivants permettent de calculer le TM selon la formule suivante :

$$TM = 1 + \text{MULTI}\% \times \text{MOY-ML}$$

- Pourcentage de sites Web ayant plus d'une version linguistique (MULTI%)
- Nombre moyen de versions linguistiques par site Web multilingue (MOY-ML)

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

ANTÉCÉDENTS

L'étude du département des arts visuels de l'université Ionienne sur la prévalence de l'anglais dans les sites Web des domaines de premier niveau (ccTLD) des pays de l'Union européenne (Giannakoulopoulos, A et al., 2020) a produit les résultats suivants pour l'espace spécifique ciblé et pour 2019 :

MULTI% = **10,99%** (chiffre calculé précisément pour chaque ccTLD et pondéré avec le nombre total de sites par pays pour obtenir le chiffre global)

MOY-ML = **4,91** (idem)

TM = **1,136** (idem)

La nouvelle étude utilisant DataProvider.com présente de nouveaux chiffres pour 2025, 6 ans plus tard :

MULTI% = **17,00 %** (chiffre calculé précisément pour chaque ccTLD et pondéré par le nombre total de sites par pays pour obtenir le chiffre global)

MOY-ML = **4,91** (chiffre inconnu tiré par défaut de l'étude précédente comme valeur conservatrice)

TM = **1,834** (idem que MULTI%).

OBDILCI a utilisé les résultats de l'étude de l'université Ionienne, y compris les calculs bruts, pour étayer une étude sur le biais résultant de la non-prise en compte du fait que de nombreux sites web ont plusieurs versions linguistiques (Pimienta D., 2024). L'échantillon de sites web utilisé pour mener cette étude comprenait 102 149 sites. Les principaux résultats de l'étude de l'université Ionienne, en ce qui concerne les langues en ligne, étaient qu'en 2019, **28,4 %** des sites web avaient une version anglaise, le taux de multilinguisme (TM) des ccTLD des pays de l'UE (avant le Brexit) était de 1,14 avec une moyenne de 4,9 langues différentes par site web multilingue.

Ce rapport calculera les mêmes données en 2025 en utilisant une portée beaucoup plus large de sites Web.

Attention : en raison du changement de méthode, la comparaison doit être prise avec prudence, car il est indécidable de savoir si les changements dans les résultats sont totalement dus à l'évolution du monde du Web au cours des 6 années écoulées ou bien s'ils reflètent également des **biais de sélection inévitables** dans les deux études.

PRINCIPAUX RÉSULTATS :

La proportion d'anglais est désormais réduite à **19,76 %** (soit une baisse de 30 %), le taux de multilinguisme de l'échantillon augmente de 50 % pour atteindre **1,83**. Le nombre moyen de versions linguistiques n'est pas calculable dans l'état de la base de données et a été maintenu identique pour permettre le calcul du taux de multilinguisme. Le nouvel échantillon comprend 15 779 643 sites web.

Tableau 1: Principaux résultats

	UNIV. IONIENNE	DATAPROVIDER.COM /OBDILCI
DATE	2019	2025
ÉCHANTILLON (nombre de sites)	102 149	15 779 643
TM	1.136	1.834
MOY-ML	4,91	?
ANGLAIS	28,4%	19,76%
MULTI%	10,96%	17,00%

Remarques :

Pour l'université Ionienne, les résultats moyens varient selon la méthode de comptage : moyenne simple, pondération par la population, pondération par le nombre de sites explorés dans chaque pays. La technique qui a été choisie, la pondération par le nombre total de sites existant dans le pays, est celle qui donne les chiffres les plus bas pour MULTI% et MOY-ML et les plus élevés pour le pourcentage de sites en anglais.

Pour DataProvider.com, il s'agit d'une élaboration réalisée par OBDILCI à partir de données brutes, plus quelques hypothèses de calcul décrites ci-après.

MÉTHODE

L'échantillon de données collectées à partir de la base de données DataProvider.com est réalisé de la manière suivante :

- 1) en sélectionnant l'option Réponse = disponible (cela réduit le total des sites de 872 920 299 à 166 058 954, en supprimant les sites Web qui ne sont pas disponibles²)
- 2) en sélectionnant ensuite l'option « sites développés », évitant ainsi les sites présentant des erreurs et les sites en attente (par exemple, réservés pour une utilisation ultérieure). Le chiffre descend à 78 847 894, ce qui représente l'échantillon de travail.

Le processus est le suivant :

Compléter le tableau existant créé pour l'étude précédente (Pimienta, 2024) et insérer de nouvelles données pour chaque ccTLD correspondant, après avoir collecté les informations suivantes :

- dans la base de DataProvider.com, le nombre total de sites (T), le nombre de sites en anglais et le pourcentage de sites ayant une balise Hreflang présente dans le code HTML (Nh) ;
- à partir de <https://domainnamestat.com/statistics/tld/others>, indiquer le nombre de sites enregistrés, à des fins de vérification croisée.

Hreflang étant utilisé pour une proportion moyenne Pm des sites multilingues les indicateurs sont calculés avec les opérations suivantes :

$$\text{MULTI\%} = \text{Nh} \times 1/(1-\text{Pm})$$

$$\text{TM} = 1 + \text{MULTI\%} \times \text{MOY-ML}$$

²À noter que le chiffre équivalent de Netcraft est de 192 282 517, la couverture serait alors de 86%.

À partir de l'expérience manuelle réalisée dans (Pimienta, 2024), confirmée par des requêtes sur ChatGPT, Pm a été fixé à 40 %. MOY-ML ne peut être calculé à ce stade ; sa valeur a été reprise de l'étude de l'université Ionienne, à titre de donnée conservatrice. Afin de comprendre la sensibilité des données en fonction des paramètres, quelques exécutions avec d'autres valeurs de MOY-ML et Pm sont présentées :

Tableau 2: Sensibilité aux facteurs

MOY-ML	Pm	TM
5	35%	1,97
6	30%	2.36
4	40%	1,68
4.9	40%	1.834
6	40%	2.02
4	50%	1,54
4	60%	1,45
3	60%	1.34

Le TM est un facteur clé de l'évolution du multilinguisme dans l'Internet. Il convient de noter que l'utilisation croissante de Google Translate intégré aux sites web pourrait facilement faire grimper ces chiffres, sachant que seulement 1 % de cette utilisation par les sites web augmenterait TM de 2,5 (car cela ajoute 250 langues). Quoi qu'il en soit, TM est un paramètre fondamental pour corriger les biais de la mesure de l'anglais réalisée par W3techs, comme le démontre (Pimienta, 2024).

Ci-dessous, le tableau réduit présente les données principales. Les totaux sont calculés en pondérant les données par le nombre de domaines afin de maintenir l'importance relative de chaque pays. Les colonnes en vert correspondent aux données de l'université Ionienne et les colonnes en marron aux données collectées auprès de DataProvider.com. Les chiffres en rouge semblent erronés ou sot manquants dans DomainStat et ont été trouvées dans d'autres sources.

Tableau 3: Mesure comparaison

TLD	Pays	Sites	Moy-ML	TM	Anglais	DomainStat	Domaine	Anglais	HREF
at	Autriche	4 063	3,50	1.131	12,0%	2 197 489	424 985	24 780	32 655
be	Belgique	4 063	3.34	1.227	20,7%	2 454 152	470 801	60 604	60 244
bg	Bulgarie	3 178	3,70	1.346	25,7%	277 449	38 899	4 206	6 822
hr	Croatie	3 489	3,85	1.363	25,0%	167 347	66 460	9 623	9 044
cy	Chypre	828	5.11	1.402	59,9%	16 076	6 771	4 628	1 387
cz	Tchéquie	4 084	3,76	1.156	12,5%	2 011 885	565 083	18 629	29 803
dk	Danemark	4 067	3.19	1.158	17,4%	1 827 738	347 506	32 247	18 418
ee	Estonie	3 556	3,50	1.471	21,9%	215 250	75 525	8 875	13 361
fi	Finlande	3 992	3,70	1.220	17,0%	716 183	201 999	19 906	22 854
fr	France	4 125	6.42	1.127	8,9%	8 667 628	1 379 860	58 341	82 105
de	Allemagne	4 150	4.10	1.098	9,7%	27 083 237	4 184 798	178 166	222 110
gr	Grèce	3 953	3,70	1.357	30,2%	658 538	237 111	51 295	50 686
hu	Hongrie	3 993	3.30	1.172	14,1%	1 194 260	322 815	17 655	19 225
ie	Irlande	3 825	3,50	1.014	98,3%	479 441	110 428	105 853	3 539
it	Italie	4 123	4.10	1.220	15,3%	5 859 250	1 172 271	88 050	123 552

lv	Lettonie	3 406	3,25	1.527	25,7%	189 691	44 085	6 031	9 922
lt	Lituanie	3 773	3,38	1.339	21,1%	319 107	75 028	6 330	11 405
lu	Luxembourg	2 876	3,72	1.362	25,9%	149 793	22 128	5 400	5 009
mt	Malte	444	6,60	1.095	89,5%	19 637	4 117	3 780	290
nl	Pays-Bas	4 133	3,39	1.140	14,2%	8 667 628	1 475 505	168 840	117 035
pl	Pologne	4 110	3,64	1.120	10,2%	4 359 431	788 235	36 006	51 792
pt	Portugal	4 084	3,80	1.163	18,3%	545 400	145 896	15 823	20 296
ro	Roumanie	3 975	3,57	1.314	27,9%	1 042 223	272 838	38 508	23 093
sk	Slovaquie	3 943	3,34	1.190	14,3%	740 532	203 334	7 429	14 148
si	Slovénie	3 619	3,60	1.286	20,6%	49 558	65 413	5 497	9 198
es	Espagne	4 088	3,44	1.162	11,9%	2 958 675	515 874	39 221	53 953
se	Suède	4 084	3,00	1.146	16,7%	3 036 275	487 922	53 437	21 728
uk	Royaume-Uni	4 125	7,83	1.013	98,4%	23 764 012	2 073 956	2 049 102	39 566
	Total/Moy.	102 149	4,91	1.136	28,4%	99 667 885	15 779 643		

En 2019, les pays présentant le taux de multilinguisme le plus élevé sur le web étaient, par ordre d'importance : la Lettonie, l'Estonie, Chypre, la Croatie, le Luxembourg, la Grèce, la Bulgarie, la Lituanie et la Roumanie. Les pays présentant le taux le plus faible étaient : le Royaume-Uni, l'Irlande, Malte, l'Allemagne, la Pologne et la France.

En 2025, les plus élevés : **Luxembourg, Lettonie, Chypre et Estonie, Bulgarie, Slovénie et Portugal** et les plus bas : **Royaume-Uni, Irlande, Malte, Suède, Allemagne, Danemark, Tchèque et France.**

En 2019, les taux d'anglais les plus élevés étaient observés au Royaume-Uni, en Irlande, à Chypre et en Grèce ; les plus bas étaient observés en France, en Allemagne, en Pologne, en Espagne, en Autriche et en Tchéquie.

Le taux d'anglais le plus élevé en 2025 : **Royaume-Uni, Irlande, Malte, Chypre et Grèce ; le plus bas : Tchèque, Slovaquie, France, Allemagne, Hongrie, Pologne et Autriche.**

Tableau 4: Comparaison des résultats

ccTLD	%Anglais	%HREF	ANG2019 vs. ANG2025		TM	MULTI%
			ANG2019	ANG2025		
at	5,83%	7,68%	-51,30%	1,94	19,21%	
be	12,87%	12,80%	-37,71%	2,57	31,99%	
bg	10,81%	17,54%	-58,01%	3,15	43,84%	
hr	14,48%	13,61%	-41,98%	2,67	34,02%	
cy	68,35%	20,48%	14,18%	3,51	51,21%	
cz	3,30%	5,27%	-73,56%	1,65	13,19%	
dk	9,28%	5,30%	-46,76%	1,65	13,25%	
ee	11,75%	17,69%	-46,27%	3,17	44,23%	
fi	9,85%	11,31%	-42,09%	2,39	28,28%	
fr	4,23%	5,95%	-52,30%	1,73	14,88%	
de	4,26%	5,31%	-56,10%	1,65	13,27%	
gr	21,63%	21,38%	-28,38%	3,62	53,44%	
hu	5,47%	5,96%	-61,29%	1,73	14,89%	

ie	95,86%	3,20%	-2,51%	1,39	8,01%
it	7,51%	10,54%	-51,05%	2,29	26,35%
lv	13,68%	22,51%	-46,73%	3,76	56,27%
lt	8,44%	15,20%	-59,95%	2,86	38,00%
lu	24,40%	22,64%	-5,92%	3,78	56,59%
mt	91,81%	7,04%	2,58%	1,86	17,61%
nl	11,44%	7,93%	-19,40%	1,97	19,83%
pl	4,57%	6,57%	-55,07%	1,81	16,43%
pt	10,85%	13,91%	-40,88%	2,71	34,78%
ro	14,11%	8,46%	-49,49%	2,04	21,16%
sk	3,65%	6,96%	-74,53%	1,85	17,40%
si	8,40%	14,06%	-59,29%	2,73	35,15%
es	7,60%	10,46%	-35,96%	2,28	26,15%
se	10,95%	4,45%	-34,43%	1,55	11,13%
uk	98,80%	1,93%	0,39%	1,23	4,77%
MOY	19,76%	6,80%	-30,46%	1,83	17,00%

À partir de la base de données DataProvider.com, les ccTLD des pays de l'Union européenne, plus .uk, affichent une proportion moyenne de sites en anglais de **19,8 %** et une proportion moyenne de sites multilingues de **17 %**. Si le nombre moyen de langues par site multilingue resté de **4,9**, alors le taux de multilinguisme (nombre de versions linguistiques divisé par le nombre de sites) serait de **1,83**, supérieur au taux de multilinguisme humain (1,44 selon Ethnologue). Ces chiffres montrent une diminution pour l'anglais et une croissance pour les paramètres de multilinguisme par rapport à l'étude de 2019. Cependant, la comparaison doit être observée avec prudence car il est impossible de déterminer si la différence est due au biais de sélection ou à l'évolution dans le temps.

TAUX DE MULTILINGUISME DE LA TOILE ENTIÈRE

Est-il possible de calculer les mêmes paramètres pour l'ensemble du WWW en utilisant la collecte de données hreflang de DataProvider.com ?

Le paramètre Hreflang permet de déduire le paramètre Multi% avec la règle d'extrapolation qui suppose que seulement 40 % des sites multilingues utilisent cette approche. Dans l'état actuel de la base de données, il est possible d'estimer le pourcentage de sites multilingues dans le monde, mais il est difficile de déterminer le nombre moyen de langues par site multilingue, ce qui empêche de fournir un chiffre précis pour le TM.

En supposant que Hreflang soit utilisé par 40 % des sites multilingues, le pourcentage de sites multilingues pour l'ensemble du Web serait d'environ **12 %**. Si le nombre moyen de langues par site web est confirmé à 5 pour l'ensemble du Web, le TM pour l'ensemble du Web serait alors de **1,6**. À ce stade, nous ne disposons pas de moyen de calculer précisément ce chiffre pour l'ensemble du Web ; cette estimation est donc à considérer avec prudence.

PROPRIÉTÉ DE MULTILINGUISME DE LA TOILE

En utilisant la même collecte de données href de DataProvider.com et la même méthode d'inférence de Multi%, et en croisant dans la base de données avec d'autres paramètres, il est

possible de corrélérer les paramètres avec le niveau de multilinguisme, en ciblant les critères propices au multilinguisme (ceux avec Multi% supérieur à la moyenne globale de 12%), et ceux qui sont au contraire plus propices au monolinguisme.

Tableau 5: Variabilité du Multi% selon le contexte

ÉCHANTILLON	%Hreflang	Multi%
70 < Empreinte économique ³ < 79	28,3%	70,75%
60 < Empreinte économique < 69	27,4%	68,50%
Empreinte économique > 89	24,5%	61,25%
50 < Empreinte économique < 59	19,1%	47,75%
Livraison internationale	16,6%	41,50%
40 < Empreinte économique < 49	16,4%	41,00%
Siège	12,8%	32,00%
E-commerce	11,6%	28,00%
30 < Empreinte économique < 39	10,3%	25,75%
Boutique en ligne	9,8%	24,50%
WebType = E-commerce	9,8%	24,50%
E-commerce élevé	9,1%	22,75%
E-commerce Confiance max.	8,2%	20,50%
Tous les sTLD	8,1%	20,25%
ccTLD de l'UE	6,8%	17,00%
Géo. gTLD	5,8%	14,50%
Tous les ccTLD	5,6%	14,00%
Type de site Web = Entreprise	5,5%	13,75%
20 < Empreinte économique < 29	6,2%	12,50%
Tous les sites	4,9%	12,25%
Tous les continents (voir ci-dessous)	4,7%	11,75%
.com	4,3%	10,75%
Tous les gTLD	4,2%	10,50%
.org	3,7%	9,25%
.net	3,5%	8,75%
10 < Empreinte économique < 19	3,5%	8,75%
Tous les nouveaux gTLD	3,1%	7,75%
WebType = Contenu	1,7%	4,25%

³Citation du dictionnaire Dataprovider.com : L'empreinte économique est une mesure exclusive qui donne une estimation de l'impact économique d'un site web, sur une échelle de 0 (faible) à 100 (élevé). Ce score est calculé à partir de plusieurs facteurs clés, chacun offrant un aperçu des différents aspects de la présence et de l'activité commerciales du site. Il prend en compte la taille globale du site, comme le nombre de sous-domaines ou de domaines de redirection. L'activité économique est également estimée en examinant diverses fonctionnalités de commerce électronique.

Empreinte économique < 11	0,9%	2,25%
Type de site Web = Blog	0,8%	2,00%
Type de site Web = Forum	0,3%	0,75%

Ce premier tableau montre une forte corrélation positive du multilinguisme avec :

- Sites à forte empreinte économique
- Sites de commerce électronique
- Site du siège social de l'entreprise
- Domaines de premier niveau sponsorisés⁴
- ccTLD de l'Union européenne

Une corrélation positive pour :

- Géo gTLD et tous les ccTLD
- Type de Web d'entreprise.

En revanche, il montre une forte corrélation négative avec le multilinguisme avec :

- Sites Web de type Forum ou Blog
- Sites à très faible empreinte économique
- Nouveaux gTLD

Et un niveau de multilinguisme relativement faible en moyenne pour .net, .org, gTLDs et .com.

Le tableau suivant présente la vérification de la corrélation du multilinguisme avec la langue principale des sites Web.

Tableau 6: Variabilité de Multi% selon la langue principale du site

ÉCHANTILLON	%Hreflang	Multi%
Ukrainien	23,6%	59,0%
Catalan	22,2%	55,5%
Estonien	21,6%	54,0%
Grec	20,3%	50,8%
Italien	10,9%	27,3%
Hindi	10,0%	25,0%
Néerlandais	9,6%	24,0%
Finlandais	9,6%	24,0%
Français	8,6%	21,5%
Hébreu	8,1%	20,3%
Espagnol	7,6%	19,0%
Turc	7,6%	19,0%
Vietnamien	7,5%	18,8%

⁴ https://en.wikipedia.org/wiki/Sponsored_top-level_domain

Allemand	6,8%	17,0%
Persan	6,8%	17,0%
Polonais	6,5%	16,3%
Ourdou	6,2%	15,5%
Swahili	6,0%	15,0%
Thaïlandais	5,4%	13,5%
Suédois	5,1%	12,8%
Télougou	5,0%	12,5%
MOYENNE	4,8%	12,0%
Afrikaans	4,5%	11,3%
Portugais	4,2%	10,5%
Anglais	3,8%	9,5%
Malais/indonésien	3,4%	8,5%
Russe	3,0%	7,5%
Marathi	1,8%	4,5%
Japonais	1,7%	4,3%
Coréen	1,5%	3,8%
Chinois	1,0%	2,5%

Concernant la langue principale des sites, plus de 50 % des sites en ukrainien, catalan, estonien et grec sont multilingues. À l'inverse, les sites en chinois, coréen ou japonais sont multilingues dans moins de 5 % des cas, ⁵et le portugais, contrairement à l'italien, au français, à l'espagnol et aux autres langues européennes, est inférieur à la moyenne.

Tableau 7: Analyse des continents

CONTINENT	Nombre total sites	%	Nombre HREF	Occurrences de la balise HrefLang	Href %
Amérique du Nord	81 700 171	42,85	30 585 750	775 311	2,53%
Europe	48 926 074	31,19	29 284 084	2 294 511	7,84%
Asie	24 338 209	19,35	16 402 470	613 182	3,74%
Amérique du Sud	3 956 843	2,74	3 103 861	119 150	3,84%
Océanie	3 154 042	1,95	1 884 996	35 682	1,89%
Afrique	1 477 744	1,91	933 273	31 704	3,40%
Antarctique	2 418	0,00	588	15	2,55%
TOTAL/MOY.	163 555 501	100	82 195 022	3 869 555	4,71%

⁵Notez cependant qu'il existe un biais potentiel : il est possible que les concepteurs de sites Web asiatiques soient réticents à utiliser la balise hreflang qui les oblige à spécifier la langue en utilisant l'alphabet latin.

Les sites Web européens sont les leaders en matière de multilinguisme et l'Amérique du Nord a encore un long chemin à parcourir ⁶.

Si l'hypothèse de 5 langues en moyenne par site multilingue est correcte, le TM global serait de 1,6, supérieur au taux équivalent pour l'humanité, mais pas supérieur à 2, comme nous l'attendions des mesures manuelles (Pimienta, 2024). Cependant, le nombre moyen de langues par site multilingue reste difficile à estimer, la seule donnée fiable étant celle de l'échantillon de l'étude de l'Université Ionienne sur les ccTLD de l'UE, avec une valeur de 4,91 et une variance relativement faible (voir la colonne Nm du premier tableau).

Il est intéressant de constater, grâce à ces tableaux, que même si ces chiffres sont encore provisoires, car fondés sur une hypothèse à vérifier, la variance est importante selon le pays, le type de site ou la langue principale. Ce résultat confirme toutefois deux hypothèses antérieures de l'OBDILCI :

- Le taux de multilinguisme des sites e-commerce est bien supérieur à la moyenne.
- En ce qui concerne l'échantillon Tranco, il n'est pas encore possible d'avoir des chiffres, mais si le TM global est globalement de 1,6, il ne sera pas surprenant que le chiffre pour l'échantillon Tranco se rapproche de 2, car les sites les plus visités sont plus susceptibles d'être multilingues que le reste du Web.

ANALYSE DE LA VARIABILITÉ DU MULTILINGUISME DU WEB

Le recoupement réalisé par catégories et consigné dans les 3 tableaux précédents fournit des apports éclairants sur le multilinguisme du Web.

- ✓ La variance est extrêmement élevée : les pourcentages varient de 1 % à 100 % pour le multilinguisme selon la catégorie et de 2,5 % à 60 % selon la langue principale du site. Pourtant, il est probable que la moyenne actuelle se situe autour de 12 % des sites web multilingues, avec une tendance à la hausse attendue grâce à la généralisation et au développement des outils linguistiques basés sur l'IA.
- ✓ L'Europe est beaucoup plus multilingue que les autres continents, le taux le plus bas étant celui de l'Amérique du Nord, les États-Unis étant l'un des endroits où le multilinguisme est le plus faible, bien que l'Australie et le Japon affichent des données encore plus faibles (respectivement 1,8 % et 1,5 %).
- ✓ Il existe une corrélation claire et forte entre les activités de commerce électronique et un niveau élevé de multilinguisme.
- ✓ Certaines langues sont plus enclines au multilinguisme (l'ukrainien, le catalan, l'estonien, le grec) tandis que d'autres ont encore un long chemin à parcourir (le chinois, le coréen, le japonais, le russe).
- ✓ Les leaders de l'économie numérique sont fortement multilingues, la corrélation entre leadership économique et multilinguisme est spectaculaire avec le paramètre « empreinte économique ».

⁶Et le Canada, avec 5,5 %, au-dessus de la moyenne de 4,8 %, n'est pas à blâmer ! De toute évidence, le pays où le multilinguisme est le moins présent aujourd'hui est les États-Unis, avec 1,9 % de balises hreflang correspondant à environ moins de 5 % de sites multilingues.

IV RÉFÉRENCES

Giannakouloupoulos, A., Pergantis, M., Konstantinou, N., Lamprogeorgos, A., Limniati, L., and Varlamis, I. (2020). Exploring the dominance of the English language on the websites of EU countries. *Fut. Int.* 12, 76.

<https://doi.org/10.3390/fi12040076>

Pimienta, D. (2024). Is it true that more than half the Web contents are in English? If Web multilingualism is paid due attention then no! *Forum for Linguistic studies*, Vol. 6 , Iss. 5

<https://doi.org/10.30564/fls.v6i5.7144> - [Version en français](#)