

LES LANGUES d'Internet¹

Dans l'édition 2018 de *La Langue française dans le monde*, il était établi² que la langue française occupait la **quatrième place sur Internet**, derrière l'anglais, le chinois et l'espagnol, avec une avance confortable sur les langues suivantes : allemand, portugais, japonais, russe, hindi et arabe.

En 2021, l'Observatoire de la langue française a demandé une actualisation des données du modèle³. Celle-ci a été réalisée tout en élargissant le nombre de langues traitées, passant de 130 langues avec plus de 5 millions de locuteurs en langue maternelle à 329 langues avec plus de 1 million de locuteurs.

Un des éléments majeurs des changements de ce modèle est l'utilisation de la dernière base de données démolinguistiques d'Ethnologue⁴, qui représente l'échantillon actualisé complet le plus précis existant et qui propose désormais la répartition des langues secondes par pays, ce qui a permis d'éliminer l'un des biais majeurs de la méthode. Toutes les élaborations statistiques qui suivent sont basées sur cette source, quoique dans plusieurs cas elle puisse proposer des valeurs différentes de la répartition de la langue française dans le monde telle qu'établie par l'OIF. Par souci de cohérence méthodologique, les données de cette source ont donc été utilisées pour l'ensemble des développements statistiques et, si les différences existent, leur impact sur les résultats produits est marginal ou faible⁵.



¹ Contribution rédigée par Daniel Pimienta, Observatoire de la diversité linguistique et culturelle dans l'Internet.

² Daniel Pimienta pour MAAYA. *Étude sur la présence de la langue française dans le cyberspace*. Voir QR code.



³ Il s'agit d'un modèle de production d'indicateurs de la place des langues dans l'Internet, mis au point en 2017 et amélioré en 2021, par l'Observatoire de la diversité linguistique et culturelle dans l'Internet et dont la méthodologie, l'ensemble des sources, l'analyse des biais et les produits sont documentés sur le site de la FUNREDES. Voir QR codes.



⁴ Global Dataset #24 de mars 2021. Voir QR code.

⁵ En tout état de cause, ce choix ne favorise pas les résultats du français car le nombre total de francophones L1+L2 comptabilisé par Ethnologue est inférieur au chiffre proposé par l'OIF.

D'autre part, les comparaisons entre les résultats de 2018 et ceux de 2022 sont à conduire avec prudence, notamment en raison des différences notables concernant les données sur le multilinguisme à l'échelle mondiale⁶.

SYNTHÈSE DES ÉVOLUTIONS

Bien que le français reste en quatrième position sur Internet :

- il est maintenant accompagné, et peut-être même déjà dépassé⁷, par l'**hindi**, qui montre une ascension spectaculaire ;
- son avance sur les langues suivantes (maintenant : portugais, russe, arabe, allemand, japonais et malais) s'est **considérablement réduite**, comme conséquence de la combinaison des deux faits suivants :

1. les taux de connexion à l'Internet des francophones des pays industrialisés sont proches de la saturation (85 % en moyenne), laissant une marge faible à l'augmentation,
2. la fracture numérique des pays francophones africains est bien plus lente à se résorber que la croissance moyenne de la connectivité dans le monde.

En ce qui concerne l'analyse thématique (qui a été affinée et élargie par rapport à l'édition précédente) les points forts du français dans l'Internet sont à trouver dans **les livres, les MOOC⁸ et la recherche** (élargie à l'ensemble des thèmes de la rubrique « Science et technologie »).

⁶ Voir encadré « Une population mondiale plurilingue » p. 314.

⁷ Les données officielles de connectivité à l'Internet de l'Inde semblent largement sous-estimées et la valeur retenue dans le modèle est restée conservatrice par rapport à d'autres sources (cette donnée est essentielle pour les calculs du modèle établi). De plus, si des langues qui sont des variantes de l'hindi (comme le rajasthani, le marwari, le magadhi, le chhattisgarhi, le bhojpuri et le awadhi) étaient comptabilisées avec l'hindi, alors il passerait devant le français. Si l'ourdou et l'hindi étaient comptabilisés comme une seule langue (l'ancien hindoustani), comme le proposent certaines sources, alors il serait largement au-dessus du français. De même, la population des locuteurs d'hindi étant supérieure à celle des locuteurs d'espagnol, et l'Inde connaissant une forte expansion de l'Internet, l'hindi est un possible candidat à moyen terme pour supplanter l'espagnol à la troisième place.

⁸ Cours en ligne ouvert et massif (également appelés CLOM).

« Une part significative de la population mondiale est composée de personnes qui maîtrisent et utilisent au moins une langue en plus de leur langue première. »

UNE POPULATION MONDIALE PLURILINGUE

Une part significative de la population mondiale est composée de personnes qui maîtrisent et utilisent au moins une langue en plus de leur langue première. En 2018, le modèle tablait sur un niveau de multilinguisme de l'ordre de 25 %, et les résultats pour langue première plus langue seconde (L1+L2) s'établissaient donc, en pourcentage, sur la base de 125 % pour prendre en compte le multilinguisme. Dans cette nouvelle édition de 2022, les données d'Ethnologue indiquent à l'échelle mondiale :

- Total (L1) = 7 231 699 136 individus (A)
- Total (L1+L2) = 10 361 716 756 individus (B)
- Donc le facteur de multilinguisme passe à 43 % (B/A = 1,4328), valeur qui est certainement plus proche de la réalité.

Compte tenu du fait que l'Internet est probablement le lieu où l'expression du multilinguisme est la plus naturelle et que pour les internautes le choix entre langue maternelle et langue(s) seconde(s) est un continuum lorsqu'ils font usage des ressources de l'Internet, la décision a été prise d'établir tous les résultats sur la base de L1+L2 et les pourcentages par rapport au total global L1+L2, qui est 43 % plus grand que la population mondiale, selon la source démolinguistique utilisée (au lieu de le faire par rapport à la population mondiale). Cela doit être pris en compte pour faire des comparaisons entre 2018 et 2022.

LA PLACE DE LA LANGUE FRANÇAISE SUR INTERNET

Le modèle utilisé dans l'édition précédente a été amélioré sur certains aspects (meilleure source de données démolinguistiques, annulation d'un des biais principaux de l'étude grâce à l'accès à des données des locuteurs seconde langue par pays, meilleure prise en compte du multilinguisme et poursuite de l'analyse systématique des biais, inévitables dans ce genre de travaux, et de la recherche de moyens pour les dépasser) et l'ensemble des données a été actualisé en mai 2021.

Les définitions des macro-indicateurs produits par le modèle restent les mêmes :

- **Puissance** (PUISS.) : moyenne de six indicateurs⁹ qui sont tous exprimés, pour chaque langue traitée, en termes de pourcentage par rapport au total mondial de locuteurs L1+L2. Cet indicateur intègre l'ensemble des paramètres qui marquent la **présence** de la langue dans l'Internet.
- **Capacité** (CAPAC.) : puissance divisée par pourcentage mondial de locuteurs (L1+L2) dans la langue. Permet de comparer la force des langues indépendamment du nombre de leurs locuteurs.
- **Gradient** (GRAD.) : puissance divisée par pourcentage mondial de locuteurs (L1+L2) connectés pour chaque langue. Permet de mesurer le dynamisme des locuteurs connectés (à créer des contenus, à faire usage de l'Internet, à générer du trafic dans leur langue, etc.).

L'ensemble des paramètres obtenus en sortie du modèle pour les quinze langues les plus puissantes de l'Internet en 2021 est présenté dans le tableau page de droite.

⁹ Voir encadré « Les indicateurs de puissance » page de droite.

LES INDICATEURS DE PUISSANCE

Pour chaque langue (sachant que le modèle transforme, par pondération, les données par pays en données par langue) :

- **Internautes (INTER.)** : pourcentage mondial des locuteurs connectés (obtenu en pondérant les données de connectivité par pays avec les données démolinguistiques).
- **Trafic (TRAF.)** : pourcentage mondial du trafic vers les sites Web (obtenu à partir d'applications qui rapportent la répartition du trafic par pays).
- **Usages (USAG.)** : pourcentage de la présence de la langue dans les plateformes et applications (par exemple pourcentage d'abonnés aux divers réseaux sociaux).
- **Contenus (CONT.)** : pourcentage global de pages Web.
- **Index (INDEX)** : transformation des classements des pays dans une grande variété d'indicateurs de la société de l'information (par exemple qualité du gouvernement électronique) en pourcentages par langue.
- **Interfaces (INTERF.)** : mesure de sa présence comme langue d'interface aux applications ou comme langue de traduction en ligne.

LANGUES ¹⁰	POP. MOND.	TRAF.	INTER.	USAG.	CONT.	INTERF.	INDEX	POP. CONN.	PUISS.	CAPAC.	GRAD.
Anglais	13,0 %	37,9 %	15,2 %	27,9 %	38,7 %	21,6 %	17,6 %	64,2 %	26,5 %	2,04	1,74
Chinois	14,7 %	7,9 %	17,6 %	5,5 %	8,2 %	25,0 %	19,2 %	65,6 %	13,9 %	0,94	0,79
Espagnol	5,2 %	10,8 %	6,9 %	11,7 %	5,4 %	9,8 %	7,5 %	72,5 %	8,7 %	1,66	1,26
Français	2,6 %	2,7 %	3,0 %	3,8 %	5,4 %	4,3 %	3,2 %	64,9 %	3,7 %	1,45	1,23
Hindi	5,8 %	4,9 %	4,3 %	3,2 %	0,3 %	4,1 %	3,8 %	41,2 %	3,4 %	0,59	0,79
Portugais	2,5 %	1,5 %	3,1 %	5,5 %	3,3 %	3,9 %	3,0 %	68,3 %	3,4 %	1,36	1,09
Russe	2,5 %	1,8 %	3,6 %	2,3 %	3,4 %	3,9 %	3,8 %	78,3 %	3,1 %	1,26	0,88
Arabe	3,5 %	2,3 %	3,8 %	3,0 %	2,1 %	4,2 %	2,9 %	59,0 %	3,0 %	0,86	0,80
Allemand	1,3 %	1,3 %	2,1 %	2,0 %	5,9 %	2,8 %	3,0 %	87,7 %	2,8 %	2,17	1,36
Japonais	1,2 %	2,0 %	2,1 %	1,8 %	3,5 %	2,8 %	3,0 %	92,6 %	2,5 %	2,06	1,22
Malais	2,4 %	0,9 %	2,2 %	2,8 %	0,8 %	1,9 %	2,0 %	51,0 %	1,8 %	0,74	0,80
Italien	0,7 %	0,5 %	0,9 %	1,0 %	3,4 %	1,2 %	1,2 %	75,7 %	1,4 %	2,09	1,51
Turc	0,9 %	1,0 %	1,2 %	1,6 %	0,9 %	1,4 %	1,2 %	78,0 %	1,2 %	1,46	1,02
Coréen	0,8 %	0,9 %	0,9 %	1,0 %	0,8 %	1,1 %	0,9 %	64,7 %	1,0 %	1,22	1,03
Bengali	2,6 %	1,2 %	1,2 %	1,1 %	0,3 %	0,7 %	0,8 %	24,5 %	0,9 %	0,35	0,77
RESTE	40,4 %	22,3 %	31,9 %	25,9 %	17,6 %	11,3 %	26,9 %		22,7 %		
TOTAL	100 %	100 %	100 %	100,0 %	100 %	100 %	100 %	54,9 %	100 %		

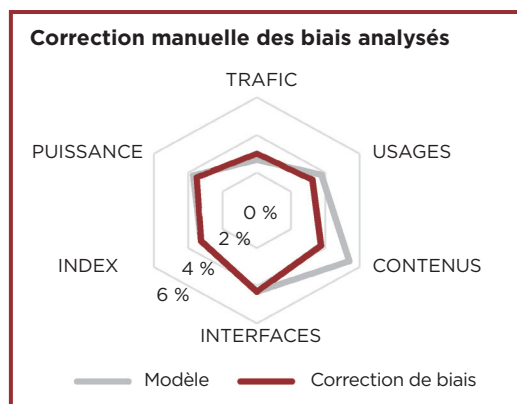
¹⁰ Dans tout le chapitre, les langues indiquées en *italique* sont des « macro-langues », c'est-à-dire le regroupement de plusieurs langues appartenant à la même famille (par exemple l'arabe égyptien et l'arabe marocain, la macro-langue *arabe* regroupant ainsi 24 langues).

Ce résultat brut doit être corrigé des biais existant dans les sources, lesquels ont été analysés ainsi :

INDICATEUR	ÉVALUATION	COMMENTAIRES
INTERNAUTES	17/20	L'indicateur est très fiable même s'il a perdu quelques points par rapport à la précédente édition, parce que l'Union internationale des télécommunications (UIT) ¹¹ n'offre plus d'estimation lorsque le pays ne fournit pas de données officielles. Cependant la Banque mondiale pallie souvent ces lacunes.
INDEX	18/20	Cet indicateur dérive d'un mélange de 25 micro-indicateurs fiables évaluant différents paramètres caractérisant la société de l'information dans un pays.
CONTENUS	8/20	80 % des sources proviennent des excellentes statistiques de Wikimedia, ce qui implique un fort biais en faveur des langues européennes et en particulier du français.
TRAFIC	11/20	Cet indicateur est issu de la mesure du trafic par pays à l'aide d'Alexa.com ¹² sur une sélection ¹³ de 338 sites du Web. Le biais au détriment des pays asiatiques présent en 2017 a été corrigé mais de nouveaux biais pénalisent maintenant les pays européens. Biais défavorable au français.
INTERFACES	19/20	Ce sont des données objectives (présence ou non d'une langue dans l'interface d'une application ou comme cible d'un service de traduction en ligne).
USAGES	12/20	Cet indicateur repose principalement sur des données d'abonnement aux réseaux sociaux par pays. Alors que les données collectées peuvent être considérées comme relativement fiables, la méthode implique un biais défavorisant les pays non occidentaux ayant des applications alternatives à Facebook, Twitter, LinkedIn, etc.

Une correction manuelle des biais analysés conduit à ce résultat pour le français :

Français	TRAFIC	USAGES	CONTENUS	INTERFACES	INDEX	PUISSANCE
MODÈLE	2,68 %	3,76 %	5,40 %	4,32 %	3,24 %	3,71 %
CORRECTION DE BIAIS	3,00 %	3,25 %	3,75 %	4,32 %	3,24 %	3,51 %



¹¹ L'institution spécialisée des Nations unies pour les technologies de l'information et de la communication.

¹² Alexa.com est un service appartenant à Amazon, qui offre des données de trafic sur les millions de sites Web les plus visités.

¹³ La sélection a été établie en essayant d'éviter les biais nationaux (trop de sites d'un même pays) et en recouvrant un éventail équilibré de thématiques. Sur les 10 millions de sites les plus visités, selon Alexa, nous en avons donc retenu 338, une première sélection dans les sites de tête à vocation la plus mondiale et une deuxième sélection s'éloignant des sites de tête pour une plus grande diversité géographique. L'ensemble des sites est détaillé sur le site de la FUNREDES. Voir QR code.



Et le classement général corrigé des biais serait le suivant :

RANG	LANGUE	PRÉSENCE
1	Anglais	25,00 %
2	Chinois	15,00 %
3	Espagnol	7,00 %
4	Français	3,50 %
4	Hindi	3,50 %
6	Portugais	3,00 %
6	Russe	3,00 %
8	Arabe	2,50 %
8	Allemand	2,50 %
8	Japonais	2,50 %
11	Malais	1,80 %
12	Italien	1,40 %
13	Turc	1,20 %
14	Coréen	1,20 %
15	Bengali	1,20 %
	RESTE	25,70 %

Le français conserve la quatrième position, où il est rejoint par l'hindi, mais son écart avec les suivants se rétrécit. Il apparaît que ce recul relatif du français, qui avait déjà été observé dans des mesures partielles réalisées en 2019 et 2020, est le résultat d'un différentiel trop important entre la croissance mondiale moyenne des taux de connectivité par pays et celle de plusieurs pays africains francophones¹⁴. Ce différentiel est de plus extrêmement fort avec les pays d'Asie et, dans une large mesure, les pays arabes, qui montrent des taux de connectivité qui tendent à s'approcher des taux occidentaux.

À long terme, la croissance démographique prévue pour l'Afrique va favoriser indirectement la présence du français dans l'Internet. Cependant, si la fracture numérique continue d'être aussi prononcée en Afrique francophone, la présence du français dans l'Internet pourrait s'en ressentir à court terme.

¹⁴ Voir « La fracture numérique en Francophonie » p. 327.

Concernant les francophones, un examen a été effectué sur des groupes de sites caractéristiques de thèmes particuliers, en comparant les pourcentages de trafic vers ces sites par rapport à la moyenne générale du trafic francophone et en notant les thèmes pour lesquels des écarts importants apparaissent. L'analyse de ces résultats doit tenir compte de trois facteurs :

1. la très grande dispersion du trafic francophone entre pays et régions ;
2. le fait que des écarts faibles ne sont pas significatifs ;
3. des biais de sélection possibles dans la liste des sites analysés.

Selon cette analyse, les thèmes qui suscitent le plus d'intérêt chez les francophones sont **les livres, les cours massifs en ligne et la recherche** (en général).

RÉPARTITION DES INTERNAUTES FRANCOPHONES

Les données démolinguistiques d'Ethnologue, pondérées avec les données de connectivité à l'Internet, permettent, avec les réserves exprimées précédemment quant aux possibles différences avec les données de l'OIF, de faire une analyse plus fine de la place du français dans le cyberspace et de montrer que son atout principal résulte du fait que c'est la langue ayant le deuxième plus haut **degré de cyber-mondialité**¹⁵ derrière l'anglais, les deux langues possédant une avance importante par rapport aux langues suivantes.

Le tableau suivant indique la répartition des francophones par pays en y associant les données de connectivité à l'Internet¹⁶.

¹⁵ Cet indicateur, qui intègre la surface mondiale d'utilisation de la langue, sa puissance en tant que langue seconde et sa population connectée, est détaillé plus loin.

¹⁶ Il est important de savoir que cette construction repose sur une hypothèse simplificatrice (la même que le modèle utilise pour ses calculs) qui n'est pas exempte de biais : le nombre de francophones connectés dans chaque pays est calculé en appliquant le pourcentage national de personnes connectées aux francophones dans le pays. Autrement dit, le taux de connectivité est présumé identique pour toutes les langues parlées dans chaque pays ; la réalité est probablement différente, avec des langues mieux connectées que d'autres, reflet d'inégalités socio-économiques ou de niveau d'éducation. Cette hypothèse ne favorise pas les résultats pour le français car la population francophone en Afrique, plus présente en ville, est probablement plus connectée dans chaque pays que ce qu'indique la moyenne nationale.

RÉPARTITION DES INTERNAUTES FRANCOPHONES¹⁷

PAYS / TERRITOIRES¹⁸ **RÉGION¹⁹** **L1+L2²⁰** **L1²¹** **L2²²** **%CONN.²³** **FRANCO-CONN.²⁴**

MEMBRES DE L'OIF						
■ Albanie	Europe	61 600		61 600	68,55 %	42 227
■ Andorre	Europe	58 200	4 200	54 000	95,00 %	55 290
■ Arménie	Asie	10 000		10 000	66,54 %	6 654
■ Belgique ²⁵	Europe	9 570 000	4 380 000	5 190 000	90,28 %	8 639 796
■ Bénin	Afrique	3 820 000		3 820 000	29,00 %	1 107 800
■ Bulgarie	Europe	164 000		164 000	67,95 %	111 433
■ Burkina Faso	Afrique	4 750 000	170 000	4 580 000	18,00 %	855 000
■ Burundi	Afrique	941 870	1 870	940 000	5,00 %	47 094
■ Cambodge	Asie	473 600	16 600	457 000	79,00 %	374 144
■ Cameroun	Afrique	10 000 000		10 000 000	34,00 %	3 400 000
■ Canada ²⁶	Amériques	10 980 000	7 170 000	3 810 000	97,00 %	10 650 600
■ Comores	Afrique	217 750	1 750	216 000	9,00 %	19 598
■ Congo	Afrique	3 181 100	11 100	3 170 000	13,00 %	413 543
■ Côte d'Ivoire	Afrique	8 235 000	45 000	8 190 000	36,29 %	2 988 395
■ Djibouti	Afrique	485 900	20 900	465 000	55,68 %	270 556
■ Dominique	Amériques	7 130		7 130	69,62 %	4 964
■ Égypte	Afrique	3 000 000		3 000 000	57,28 %	1 718 486
■ France	Europe	63 900 000	62 400 000	1 500 000	83,34 %	53 254 096
■ Gabon	Afrique	1 370 000	320 000	1 050 000	61,00 %	835 700
■ Grèce	Europe	809 200	6 200	803 000	75,67 %	612 331
■ Guinée	Afrique	3 320 000		3 320 000	23,00 %	763 600
■ Guinée-Bissau	Afrique	74 900	1 000	73 900	28,00 %	20 972
■ Guinée équatoriale	Afrique	380 000		380 000	28,00 %	106 400
■ Haïti	Amériques	4 673 200	3 200	4 670 000	33,00 %	1 542 156
■ Laos	Asie	190 000		190 000	30,00 %	57 000

¹⁷ Sources : Ethnologue, UIT, Banque mondiale et Observatoire de la diversité linguistique et culturelle dans l'Internet. En ce qui concerne Cabo Verde, Costa Rica, Kosovo, Gambie et République de Corée, Ethnologue ne comptabilise pas de présence francophone L1 ou L2 dans sa base de données. La Louisiane n'est pas traitée comme un territoire séparé par Ethnologue (comme c'est le cas de la Martinique par exemple) et en conséquence les francophones présents en Louisiane sont totalisés avec l'ensemble des francophones des États-Unis.

¹⁸ En italique dans ce tableau : territoires de la France.

¹⁹ La classification en régions est celle d'Ethnologue.

²⁰ Population francophone première et deuxième langue (selon Ethnologue).

²¹ Population francophone première langue (selon Ethnologue).

²² Population francophone deuxième langue (selon Ethnologue).

²³ Pourcentage de personnes connectées à l'Internet. La source principale est l'UIT. Cependant, l'UIT ne fournissant plus, depuis 2017, ses propres estimations, dans les situations où il n'y a pas de source nationale officielle, certains chiffres sont fournis par la Banque mondiale et le reste a fait l'objet d'une approximation par extrapolation simple dans le modèle.

²⁴ Nombre de locuteurs francophones L1 ou L2 connectés à l'Internet.

²⁵ La Fédération Wallonie-Bruxelles est également membre.

²⁶ Le Canada est membre, ainsi que le Nouveau-Brunswick et le Québec. L'Ontario est membre observateur.

PAYS / TERRITOIRES	RÉGION	L1+L2	L1	L2	%CONN.	FRANCO-CONN.
■ Liban	Asie	2 311 800	31 800	2 280 000	78,18 %	1 807 383
■ Luxembourg	Europe	543 600	57 600	486 000	97,06 %	527 625
■ Macédoine du Nord	Europe	42 700		42 700	81,00 %	34 587
■ Madagascar	Afrique	5 250 000	30 000	5 220 000	15,00 %	787 500
■ Mali	Afrique	3 238 000	18 000	3 220 000	26,00 %	841 880
■ Maroc	Afrique	12 705 400	5 400	12 700 000	74,38 %	9 449 808
■ Maurice	Afrique	921 000	51 000	870 000	64,00 %	589 439
■ Mauritanie	Afrique	610 210	6 210	604 000	25,00 %	152 553
■ Moldavie	Europe	55 000		55 000	80,00 %	44 000
■ Monaco	Europe	37 700	11 100	26 600	98,00 %	36 946
■ Niger	Afrique	2 835 000	5 000	2 830 000	5,25 %	148 851
■ Rép. centrafricaine	Afrique	1 349 000	9 000	1 340 000	5,00 %	67 450
■ Rép. dém. du Congo	Afrique	8 680 000		8 680 000	13,00 %	1 128 400
■ Roumanie	Europe	2 340 000		2 340 000	73,66 %	1 723 585
■ Rwanda	Afrique	727 030	3 030	724 000	26,00 %	189 028
■ Sainte-Lucie	Amériques	2 910		2 910	55,00 %	1 601
■ Sao Tomé-et-Principe	Afrique	42 200		42 200	32,00 %	13 504
■ Sénégal	Afrique	4 210 000	100 000	4 110 000	40,00 %	1 684 000
■ Seychelles	Afrique	51 190	690	50 500	79,00 %	40 440
■ Suisse	Europe	5 732 000	1 912 000	3 820 000	93,15 %	5 339 134
■ Tchad	Afrique	1 963 000	3 000	1 960 000	10,00 %	196 300
■ Togo	Afrique	5 686 200	6 200	5 680 000	19,00 %	1 080 378
■ Tunisie	Afrique	6 081 200	1 200	6 080 000	66,70 %	4 056 160
■ Vanuatu	Pacifique	88 800	1 800	87 000	28,00 %	24 864
■ Vietnam	Asie	659 700	5 700	654 000	68,70 %	453 214
MEMBRES ASSOCIÉS DE L'OIF ²⁷						
■ Chypre	Asie	79 140	1 140	78 000	86,06 %	68 111
■ Émirats arabes unis	Asie	250 000		250 000	99,15 %	247 875
■ Ghana	Afrique	244 000		244 000	53,00 %	129 320
■ Nouvelle Calédonie	Pacifique	276 300	68 300	208 000	85,00 %	234 855

²⁷ En l'absence de données démolinguistiques de la part d'Ethnologue, le Kosovo, membre associé de l'OIF, qui affichait 89,44 % de personnes connectées, selon Eurostat, donnée reprise par l'UIT, n'est pas mentionné dans ce tableau.

RÉPARTITION DES INTERNAUTES FRANCOPHONES

PAYS / TERRITOIRES	RÉGION	L1+L2	L1	L2	%CONN.	FRANCO-CONN.
■ Qatar	Asie	114 500	8 500	106 000	99,65 %	114 102
■ Serbie	Europe	276 000		276 000	77,42 %	213 670
OBSERVATEURS DE L'OIF						
■ Argentine	Amériques	16 000	16 000		85,50 %	13 680
■ Autriche	Europe	1 146 900	16 900	1 130 000	87,75 %	1 006 430
■ Bosnie-Herzégovine	Europe	28 000		28 000	69,95 %	19 585
■ Croatie	Europe	97 500		97 500	79,08 %	77 103
■ Estonie	Europe	18 900		18 900	89,06 %	16 832
■ Géorgie	Asie	16 400		16 400	68,85 %	11 291
■ Hongrie	Europe	114 000		114 000	80,37 %	91 624
■ Irlande	Europe	512 540	2 540	510 000	84,52 %	433 211
■ Lettonie	Europe	21 600		21 600	88,90 %	19 202
■ Lituanie	Europe	68 200		68 200	81,58 %	55 639
■ Malte	Europe	56 900	30 600	26 300	85,78 %	48 808
■ Mexique	Amériques	30 000		30 000	70,07 %	21 021
■ Monténégro	Europe	12 900		12 900	73,48 %	9 478
■ Mozambique	Afrique	93 900	2 300	91 600	15,00 %	14 085
■ Pologne	Europe	948 890	8 890	940 000	80,44 %	763 248
■ Rép. dominicaine	Amériques	159 100	2 100	157 000	74,82 %	119 045
■ République tchèque	Europe	235 060	2 060	233 000	80,87 %	190 086
■ Slovénie	Europe	50 000		50 000	83,11 %	41 554
■ Slovaquie	Europe	114 000		114 000	82,85 %	94 453
■ Thaïlande	Asie	580 900	13 900	567 000	66,65 %	387 184
■ Ukraine	Europe	52 800		52 800	70,00 %	36 960
■ Uruguay	Amériques	20 500	20 500		77,37 %	15 861
ÉTATS OU GOUVERNEMENTS HORS OIF						
■ Afrique du Sud	Afrique	11 000	11 000		68,00 %	7 480
■ Algérie	Afrique	13 814 300	14 300	13 800 000	58,00 %	8 012 294
■ Allemagne	Europe	12 540 000	140 000	12 400 000	88,13 %	11 052 068
■ Arabie saoudite	Asie	33 000	33 000		95,72 %	31 589
■ Australie	Pacifique	70 900	70 900		86,54 %	61 360

PAYS / TERRITOIRES	RÉGION	L1+L2	L1	L2	%CONN.	FRANCO- CONN.
■ Danemark	Europe	426 800	6 800	420 000	98,00 %	418 264
■ Espagne	Europe	2 108 000	108 000	2 000 000	90,72 %	1 912 349
■ États-Unis ²⁸	Amériques	2 123 000	1 250 000	873 000	88,50 %	1 878 832
■ Fédération de Russie	Europe	619 000		619 000	82,64 %	511 555
■ Finlande	Europe	440 400	4 400	436 000	89,61 %	394 631
■ Guadeloupe	Amériques	377 400	6 400	371 000	83,34 %	314 524
■ Guyane française	Amériques	291 000	111 000	180 000	83,34 %	242 519
■ Inde	Asie	10 000	10 000		41,00 %	4 100
■ Israël	Asie	500 000	37 000	463 000	86,79 %	433 939
■ Italie	Europe	9 756 000	116 000	9 640 000	74,39 %	7 257 214
■ Japon	Asie	6 300	6 300		92,73 %	5 842
■ Koweït	Asie	8 200	8 200		99,54 %	8 162
■ Libye	Afrique	8 220	8 220		25,00 %	2 055
■ Martinique	Amériques	312 000	15 000	297 000	83,34 %	260 020
■ Mayotte	Afrique	163 850	5 850	158 000	83,34 %	136 552
■ Norvège	Europe	162 500	16 500	146 000	98,00 %	159 250
■ Nouvelle-Zélande	Pacifique	49 100	49 100		93,00 %	45 663
■ Panama	Amériques	43 000	43 000		63,63 %	27 360
■ Pays-Bas	Europe	3 227 900	27 900	3 200 000	93,29 %	3 011 262
■ Polynésie française	Pacifique	278 600	55 600	223 000	75,00 %	208 950
■ Porto Rico	Amériques	2 100	2 100		77,74 %	1 632
■ Portugal	Europe	2 590 600	10 600	2 580 000	75,35 %	1 951 923
■ Réunion	Afrique	777 000	189 000	588 000	83,34 %	647 550
■ Royaume-Uni	Europe	12 767 000	167 000	12 600 000	94,82 %	12 105 669
■ Saint-Martin	Amériques	3 000	3 000		83,34 %	2 500
■ Saint-Barthélemy	Amériques	6 750	6 750		83,34 %	5 625
■ Saint-Pierre-et-Miquelon	Amériques	6 340	6 340		83,34 %	5 284
■ Suède	Europe	922 700	13 700	909 000	94,54 %	872 315
■ Syrie	Asie	10 100	10 100		34,25 %	3 460
■ Turquie	Asie	4 300	4 300		77,67 %	3 340
■ Wallis-et-Futuna	Pacifique	9 660	1 130	8 530	20,00 %	1 932
■ Zambie	Afrique	7 200	7 200		19,00 %	1 368
TOTAL	Monde	266 959 240	79 579 970	187 379 270	65,48 %	174 811 214

²⁸ La Louisiane est membre observateur.

Finalement, le tri, cette fois-ci par population connectée à l'Internet, limité aux pays ou territoires où le nombre de personnes connectées est supérieur à 200 000, révèle où sont les internautes francophones les plus nombreux, dans les deux tableaux page de droite et ci-dessous, pour les pays membres et pour les pays non membres de l'OIF²⁹.

Il est intéressant de noter qu'environ 70 % des francophones connectés sont présents dans des pays membres, avec un taux moyen de connectivité légèrement

supérieur à 60 % tandis que 30 % le sont dans des pays non membres avec un taux moyen légèrement supérieur à 80 %, le taux moyen mondial de connectivité des francophones étant légèrement supérieur à 65 %.

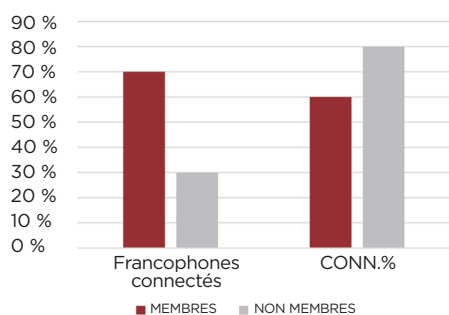
LES ATOUTS DE LA LANGUE FRANÇAISE DANS LE MONDE NUMÉRIQUE

Quels sont les atouts du français dans la compétition linguistique au sein du cyberspace ?

Les deux indicateurs de la langue française les plus significatifs dans ce contexte sont :

1. le **rapport (L1+L2)/L1**, un marqueur de la fréquence relative d'utilisation d'une langue en tant que deuxième langue, pour lequel le français est la cinquième langue du monde ;
2. le **nombre de pays où il existe des locuteurs** d'une langue donnée, pour lesquels le français est en seconde position derrière l'anglais³⁰.

Pourcentage de francophones connectés et taux moyen mondial de connectivité



Pays/territoires non membres	% CONN.	FrancoCONN.	% TOTAL
Royaume-Uni	94,82 %	12 105 669	6,92 %
Allemagne	88,13 %	11 052 068	6,32 %
Algérie	58,00 %	8 012 294	4,58 %
Italie	74,39 %	7 257 214	4,15 %
Pays-Bas	93,29 %	3 011 262	1,72 %
Portugal	75,35 %	1 951 923	1,12 %
Espagne	90,72 %	1 912 349	1,09 %
États-Unis	88,50 %	1 878 832	1,07 %
Suède	94,54 %	872 315	0,50 %
Fédération de Russie	82,64 %	511 555	0,29 %
Israël	86,79 %	433 939	0,25 %
Irlande	84,52 %	433 211	0,25 %
Danemark	98,00 %	418 264	0,24 %
Finlande	89,61 %	394 631	0,23 %
Total	80,60 %	50 337 612	29,80 %

²⁹ La ligne Total prend en compte l'ensemble des pays et territoires, et pas seulement ceux mentionnés dans les tableaux respectifs, lesquels montrent seulement les cas où le nombre de personnes connectées est supérieur à 200 000.

³⁰ Calculs réalisés dans le cadre des 329 langues de plus de 1 million de locuteurs L1, avec les données de connectivité de septembre 2021 et sur la base des 215 pays (ou territoires) sélectionnés dans le modèle.

Pays/territoires membres	% CONN.	FrancoCONN.	% TOTAL
France	83,34 %	53 254 096	30,46 %
Canada	97,00 %	10 650 600	6,09 %
Maroc	74,38 %	9 449 808	5,41 %
Belgique	90,28 %	8 639 796	4,94 %
Suisse	93,15 %	5 339 134	3,05 %
Tunisie	66,70 %	4 056 160	2,32 %
Cameroun	34,00 %	3 400 000	1,94 %
Côte d'Ivoire	36,29 %	2 988 395	1,71 %
Liban	78,18 %	1 807 383	1,03 %
Roumanie	73,66 %	1 723 585	0,99 %
Égypte	57,28 %	1 718 486	0,98 %
Sénégal	40,00 %	1 684 000	0,96 %
Haïti	33,00 %	1 542 156	0,88 %
Rép. dém. du Congo	13,00 %	1 128 400	0,65 %
Bénin	29,00 %	1 107 800	0,63 %
Togo	19,00 %	1 080 378	0,62 %
Autriche	87,75 %	1 006 430	0,58 %
Burkina Faso	18,00 %	855 000	0,49 %
Mali	26,00 %	841 880	0,48 %
Gabon	61,00 %	835 700	0,48 %
Madagascar	15,00 %	787 500	0,45 %
Guinée	23,00 %	763 600	0,44 %
Pologne	80,44 %	763 248	0,44 %
Réunion	83,34 %	647 550	0,37 %
Grèce	75,67 %	612 331	0,35 %
Maurice	64,00 %	589 439	0,34 %
Luxembourg	97,06 %	527 625	0,30 %
Vietnam	68,70 %	453 214	0,26 %
Congo	13,00 %	413 543	0,24 %
Thaïlande	66,65 %	387 184	0,22 %
Cambodge	79,00 %	374 144	0,21 %
Guadeloupe	83,34 %	314 524	0,18 %
Djibouti	55,68 %	270 556	0,15 %
Martinique	83,34 %	260 020	0,15 %
Émirats arabes unis	99,15 %	247 875	0,14 %
Guyane française	83,34 %	242 519	0,14 %
Nouvelle-Calédonie	85,00 %	234 855	0,13 %
Serbie	77,42 %	213 670	0,12 %
Total	60,87 %	124 473 602	71,20 %

Cela permet de renseigner un nouvel indicateur : le **degré de mondialisation d'une langue** (DML). Il est défini de cette manière : $DML = (L1+L2)/L1 \times$ pourcentage de pays où la langue est pratiquée.

Ce nouvel indicateur permet de classer les pays en combinant les deux critères qui marquent la surface géographique et démographique de la langue : cette combinaison produit une valeur dont la traduction en

pourcentage exprime une hiérarchie et souligne des écarts. Le DML, par construction, est indépendant de la présence de la langue dans le cyberespace.

Les deux tableaux suivants présentent les classements respectifs selon ces deux critères : le premier tableau établit la liste des langues dont le rapport $(L1+L2)/L1$ est supérieur à 2,5 ; le second tableau montre les langues qui sont parlées dans plus de 35 pays.

LANGUE	L1+L2	% CONN ³¹	# CONN ³²	L1	L1+L2/L1
Dioula	12 504 000	30,68 %	2 226 000	3 836 321	5,62
Swahili	80 338 410	20,35 %	18 378 410	16 349 242	4,37
Anglais	1 348 145 040	64,21 %	369 935 120	865 663 602	3,64
Bamanankan	14 181 800	26,05 %	4 181 800	3 694 231	3,39
Français	266 959 240	64,88 %	79 579 970	173 196 639	3,35
Ourdou	230 052 270	24,37 %	69 006 470	56 060 710	3,33
Malais	244 761 296	51,00 %	81 592 296	124 833 138	3,00
Sotho du Nord	13 731 000	67,99 %	4 631 000	9 336 310	2,97
Thaïlandais	60 715 470	66,85 %	20 715 470	40 586 942	2,93

LANGUE	Nombre de pays avec locuteurs L1 ou L2
Anglais	148
Français	108
Espagnol	70
Arabe	61
Allemand	59
Russe	51
Chinois	47
Italien	46
Portugais	44
Grec	41
Turc	39
Romani	38
Ukrainien	37
Hindi	35

En multipliant la valeur du degré de mondialisation d'une langue (DML) par le pourcentage de locuteurs connectés, on obtient le **degré de cyber-mondialisation** (DCL) de la langue ($DCL = DML \times \%Conn.$), un indicateur qui synthétise les atouts des langues pour la mondialisation dans le monde numérique. En rajoutant la dimension de connectivité moyenne à l'Internet, le DCL corrige la valeur du DML en prenant en compte la présence dans l'Internet de chaque langue.

La vision en répartition par pourcentage permet de bien mesurer les écarts et de constater que l'anglais et le français possèdent une nette avance sur le reste des langues, selon ce critère qui indique dans quelle mesure la langue est bien insérée dans le mouvement de mondialisation en cours (DML) et dans la marche vers le multilinguisme qui est portée par l'Internet (DCL). Les valeurs indiquées en pourcentage montrent la répartition des deux indicateurs entre toutes les langues considérées : ainsi,

³¹ Pourcentage de locuteurs de cette langue connectés à l'Internet.

³² Nombre de locuteurs de cette langue connectés à l'Internet.

anglais et français, ensemble, couvrent près de 25 % de cette répartition, avec une marge de distance importante devant l'allemand, le russe et l'espagnol, suivis par l'arabe, le malais, l'italien et le chinois.

« L'anglais et le français possèdent une nette avance sur le reste des langues. »

LANGUE	DML	DML%	DCL	DCL%
Anglais	2,51	12,68 %	1,61	14,24 %
Français	1,69	8,52 %	1,09	9,66 %
Allemand	0,48	2,45 %	0,42	3,75 %
Russe	0,40	2,01 %	0,31	2,76 %
Espagnol	0,37	1,90 %	0,27	2,40 %
Arabe	0,30	1,51 %	0,18	1,56 %
Malais	0,33	1,69 %	0,17	1,51 %
Italien	0,22	1,13 %	0,17	1,50 %
Chinois	0,25	1,27 %	0,16	1,46 %
Portugais	0,23	1,15 %	0,15	1,37 %
Thailandais	0,23	1,17 %	0,15	1,37 %
Romani	0,20	1,01 %	0,15	1,35 %
Turc	0,19	0,98 %	0,15	1,34 %
Grec	0,19	0,97 %	0,15	1,31 %
Ukrainien	0,21	1,05 %	0,15	1,31 %
Polonais	0,16	0,81 %	0,13	1,15 %
Persan	0,20	0,99 %	0,12	1,10 %
Roumain	0,16	0,80 %	0,12	1,06 %
Hindi	0,29	1,44 %	0,12	1,04 %

Les évolutions dans le temps du degré de cyber-mondialisation (DCL) seront plus rapides que celle du degré de mondialisation d'une langue (DML) puisque le premier est affecté par les progrès dans la connectivité Internet tandis que le second est lié uniquement à des facteurs démolinguistiques dont les évolutions sont plus lentes. Une vision prospective à l'horizon 2050, tenant compte des deux facteurs, permet d'estimer que la très forte démographie africaine³³ (et une probable résolution de la fracture numérique à l'échelle planétaire) provoquera un accroissement de l'écart entre les deux premières

langues et les suivantes comme résultat indirect de la démographie africaine sur les langues européennes de forte présence sur le continent africain (français, anglais et portugais). À ce propos, il est intéressant de noter quelle est la présence relative de ces trois langues en Afrique.

NOMBRE DE LOCUTEURS EN AFRIQUE

Anglais ³⁴	33 236 808 000
Français	109 245 420
Portugais	32 531 250



³³ Selon les Nations unies, la population de l'Europe devrait diminuer en 2050 alors que celle de l'Afrique aura doublé. Voir QR code.

³⁴ Dont 111 millions au Nigeria (tous L2).

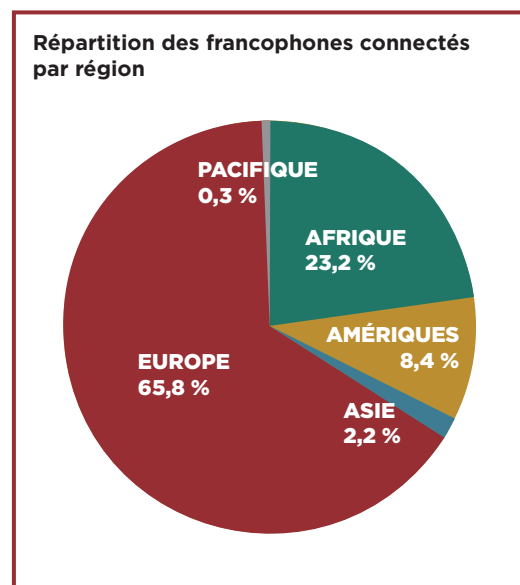
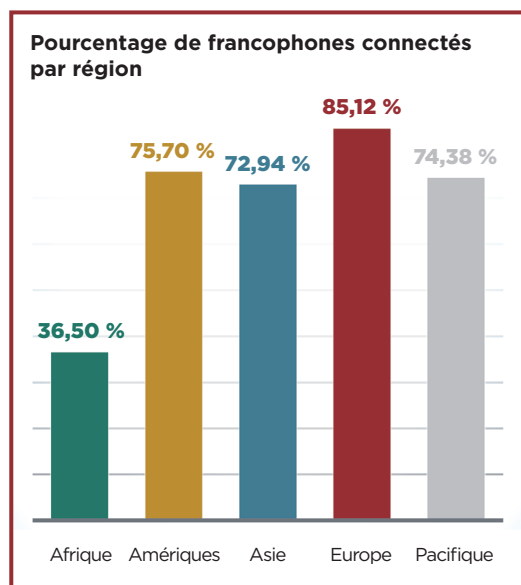
UNE VISION GÉO-LINGUISTIQUE

L'examen de la présence francophone dans l'Internet, déclinée par région, permet de mieux en cerner les forces et les faiblesses. Le tableau ci-dessous est organisé en regroupant les langues par région de présence et en obtenant ainsi les indicateurs correspondant à chaque groupe. Toutes les données sont calculées pour les populations L1+L2.

L'Afrique francophone représente 41 % des locuteurs du français mais seulement 23 % des francophones connectés, comme conséquence d'un taux de connexion à l'Internet très bas.

Les graphiques ci-dessous montrent en filigrane l'importance de la fracture numérique en Afrique et son impact sur la présence mondiale du français dans l'Internet.

	Répartition mondiale des francophones	Taux connexion Internet	Répartition mondiale des francophones connectés
Afrique	40,92 %	36,50 %	23,24 %
Amériques	7,14 %	75,70 %	8,41 %
Asie	1,97 %	72,94 %	2,24 %
Europe	49,68 %	85,12 %	65,79 %
Pacifique	0,29 %	74,38 %	0,34 %
TOTAL	100 %	54,89 %³⁵	100 %



³⁵ Il s'agit du pourcentage mondial de personnes connectées à l'Internet.

LA FRACTURE NUMÉRIQUE EN FRANCOPHONIE

Quels sont les pays membres de la Francophonie avec les taux de connexion à l'Internet les plus faibles? Le tableau suivant regroupe les pays où le taux de connexion est inférieur à 30 %. La colonne UIT2017 indique les valeurs inscrites par l'Union internationale des télécommunications (UIT) en 2017, et la colonne 2021 indique soit les valeurs inscrites par l'UIT en 2021 (valeurs fournies par un organisme officiel du pays), soit des valeurs proposées par la Banque mondiale en 2021³⁶ (indiquées en italique).

Certains pays montrent des augmentations très nettes depuis 2017 (marqués en bleu), tandis que d'autres ont des valeurs en 2021 inférieures aux estimations de 2017 (marqués en rouge). Le tableau est complété avec un autre indicateur important de la fracture numérique, celui des personnes connectées en bande large.

Il faut noter que tous les pays concernés sont africains. Une étude plus fine de la fracture numérique ne se préoccuperait pas uniquement de ce pourcentage de personnes connectées mais s'orienterait également sur des données d'accès à la bande large : la profondeur de la fracture numérique est encore plus flagrante avec ce critère pour lequel seul le Mali dépasse les 1 % alors que la moyenne mondiale est de 17 % et que la France atteint les 47 %.

PAYS/ TERRITOIRES ³⁷	UIT 2017	2021	LARGE BANDE 2021 ³⁸
Bénin	14,12 %	29,00 %	0,25 %
Burkina Faso	15,88 %	18,00 %	0,07 %
Burundi	5,59 %	5,00 %	0,04 %
Comores	8,48 %		0,10 %
Congo	8,65 %	13,00 %	0,01 %
Guinée	11,40 %	23,00 %	0,01 %
Guinée-Bissau	3,93 %	28,00 %	0,12 %
Madagascar	9,80 %	15,00 %	0,10 %
Mali	12,72 %	26,00 %	1,20 %
Mauritanie	20,80 %		0,41 %
Mozambique	20,77 %	15,00 %	0,23 %
Niger	10,22 %	5,25 %	0,04 %
Rép. centrafricaine	4,34 %		0,01 %
Rép. dém. du Congo	8,62 %	13,00 %	0,01 %
Rwanda	21,77 %	26,00 %	0,14 %
Tchad	6,50 %	10,00 %	0,00 %
Togo	12,36 %	19,00 %	0,64 %

³⁶ Données de la Banque mondiale. Voir QR code. Des mises à jour de cette source (par rapport à mai 2021) sont survenues pendant la rédaction de ce rapport ; les valeurs retenues dans ce rapport sont les dernières de septembre 2021. Pour les cas où ni l'UIT ni la Banque mondiale n'ont proposé de valeurs, le modèle a consigné une estimation à partir d'une extrapolation approximative plus ou moins forte, selon les résultats d'une recherche de valeurs dans d'autres sources.



³⁷ Le Cameroun est sorti du tableau avec la dernière estimation de la Banque mondiale, qui lui donne un chiffre de 34 % alors qu'en 2017 l'UIT estimait son taux de connectivité à 23,20 %.



³⁸ Pourcentage de personnes connectées en bande large. Données de la Banque mondiale. Voir QR code.

LA PLACE DES AUTRES LANGUES

En ce qui concerne les classements selon les critères de capacité et gradient, les langues qui dominent sont clairement les langues nationales des pays (ou territoires) qui sont considérés comme bien placés dans l'économie numérique. Parmi les dix premières places : hébreu, norvégien, estonien, finlandais, suédois, hollandais et danois, ainsi que l'italien et le catalan.

Les langues les plus connectées à l'Internet sont les suivantes (données pratiquement exemptes de biais), à comparer avec le pourcentage pour le français, qui est de 65,5 %.

Langues	Pourcentage de locuteurs connectés
Norvégien	97,87 %
Danois	97,83 %
Suédois	93,49 %
Japonais	92,62 %
Hollandais	92,02 %
Luxembourgeois	91,90 %
Suisse allemand	91,56 %
Catalan	90,50 %
Flamand occidental	90,43 %

■ Le cas de l'anglais

En ce qui concerne l'anglais, sa place relative continue de se restreindre et passe de 30 % en 2017 à 25 % en 2021, même si les médias, s'appuyant sur des résultats fortement biaisés, faute de considération sérieuse du multilinguisme, continuent à annoncer des chiffres supérieurs à 50 %.

L'Observatoire de la diversité linguistique et culturelle dans l'Internet a conduit des mesures sur la place de l'anglais et d'autres langues depuis 1998, avec une longue interruption entre 2007 et 2017, due à l'évolution des moteurs de recherche³⁹. À partir de 2011, une entre-

³⁹ La première méthode utilisée reposait sur le comptage des occurrences dans la Toile d'une collection de mots choisis pour garantir la meilleure équivalence sémantique et syntaxique entre différentes langues. Après 2007, les chiffres annoncés par les moteurs de recherche ont perdu toute crédibilité et la méthode a dû être abandonnée.

prise spécialisée dans des sondages d'utilisation des technologies dans le Web, W3Techs⁴⁰, inclut dans sa liste de sondages la langue des sites Web. Elle est devenue, en l'absence d'autre choix, la référence⁴¹ pour les données de présence des langues sur la Toile. Ses résultats au 1^{er} octobre 2021, pour les premières langues avec plus de 2 % des contenus, sont présentés dans le tableau suivant. Les valeurs produites par l'Observatoire de la diversité linguistique et culturelle dans l'Internet sont inscrites dans la dernière colonne, à des fins de comparaisons. Les différences entre les deux sources sont extrêmement importantes et posent question. La première et majeure différence est la mesure de la présence de l'anglais, et d'autres écarts considérables concernent le chinois et l'hindi, respectivement 10 et 30 fois plus bas chez W3Techs.

Langue	W3TECHS		OBSERV.	
	Rang	% du web	Rang	% du web
Anglais	1	62,8 %	1	26,50 %
Russe	2	7,3 %	7	3,13 %
Turc	3	3,8 %	13	1,24 %
Espagnol	4	3,7 %	3	8,69 %
Persan	5	3,5 %	19	0,69 %
Français	6	2,5 %	4	3,74 %
Allemand	7	2,1 %	9	2,83 %
Chinois	10	1,3 %	2	13,88 %
Arabe	11	1,2 %	8	3,04 %
Portugais	13	0,7 %	6	3,37 %
Hindi	32	0,1 %	5	3,42 %

De plus, W3Techs conserve l'historique de ses résultats depuis ses débuts en 2011 et indique une stabilité de l'anglais, voire une croissance, de 57,6 % en 2011 à 63,2 % en 2021⁴².



⁴⁰ W3Techs couvre plus de 20 différents types de « technologies » telles que système opérationnel des serveurs, hébergeurs ou outil d'analyse de trafic. Voir QR code.



⁴¹ Contenus dans l'Internet par langue selon W3Techs. Voir QR code.

⁴² Données historiques de contenus par langue selon W3Techs. Voir QR code.

Pourtant, l'Internet a changé considérablement dans les dix dernières années avec une entrée massive d'internautes locuteurs de langues asiatiques et de l'arabe ! Le pourcentage mondial de locuteurs anglophones connectés (L1+L2) était estimé par l'Observatoire de la diversité linguistique et culturelle dans l'Internet, en 2007, à 32 %, et il est passé maintenant à 15 %. Pourquoi la courbe des contenus serait-elle différente ? Comment est-il possible que les valeurs produites par W3Techs et celles de l'Observatoire de la diversité linguistique et culturelle dans l'Internet soient si différentes ? Est-il crédible, comme l'indique W3Techs, que les locuteurs de chinois et de l'hindi connectés à l'Internet, qui représentent ensemble environ 22 % du total des personnes connectées⁴³, ne réunissent que 1,4 % des contenus de la Toile ?

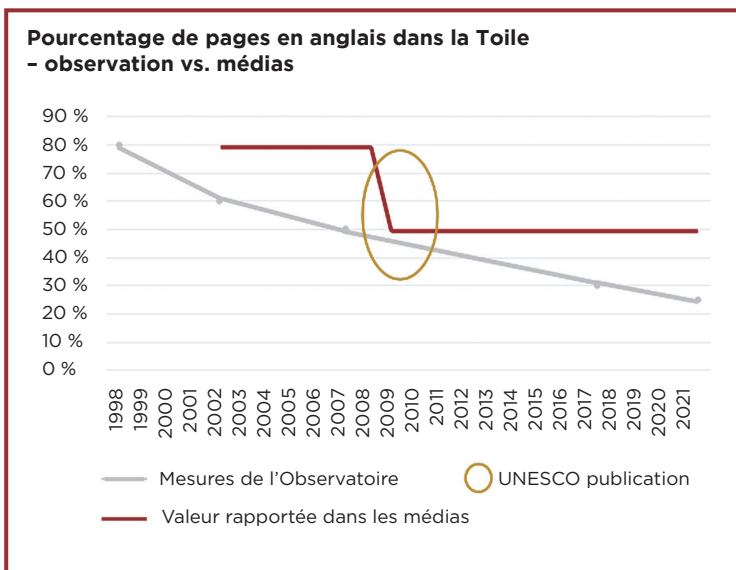
La réponse à ces questions se trouve dans l'**absence de gestion du multilinguisme** dans la méthode de W3Techs. W3Techs, sans aucun doute une entreprise experte en gestion de sondages d'utilisation des technologies, applique une méthode comparable pour les langues, qui sont cependant une « technologie »⁴⁴ assez différente de celles dont l'entreprise a la maîtrise.

La méthode de l'entreprise est d'appliquer un algorithme de reconnaissance de langue sur la page d'entrée des dix millions de sites Web les plus visités, sans explorer d'autres pages du site et sans se préoccuper du multilinguisme. Ainsi, des sites d'une seule page sont comptabilisés de manière identique à des sites de centaines ou de milliers de pages, et les nombreux sites qui offrent plusieurs options de langue pour leur interface ne sont comptabilisés que dans la langue implicite du robot programmé, qui est sans doute l'anglais. La présence de quelques mots en anglais dans la page d'entrée d'un site (par exemple de navigation ou de

copyright ou bien un résumé, comme dans le cas d'articles scientifiques) suffit probablement pour qu'il soit comptabilisé comme site anglophone, même si toutes les autres pages ne comprennent aucun mot en anglais. Dans ces conditions, la taille de l'erreur peut être très importante. De plus, les pourcentages sont calculés sur la population mondiale alors qu'ils devraient l'être sur la population de locuteurs L1+L2, ce qui gonfle artificiellement les pourcentages et cache l'erreur dans les valeurs pour le reste des langues qui ne sont pas exhibées.

La Toile se transforme chaque jour de plus en plus en l'un des endroits où l'expression du **multilinguisme** est la plus forte ; s'il y avait vraiment une *lingua franca* dans la Toile, ce serait sans aucun doute la traduction⁴⁵, la béquille indispensable du multilinguisme.

Le même phénomène de différence marquée entre l'observation et les mesures avait été observé entre 1998 et 2007 et avait cessé, pour un temps, quand l'UNESCO avait publié à ce sujet.



⁴³ Les locuteurs du chinois connectés étant aujourd'hui plus nombreux que les anglophones (17,6 % versus 15,2 %).

⁴⁴ "Is language a technology or a culture?" Imminent Question of the Year – 2021. Voir QR code.



⁴⁵ Umberto Eco déclarait déjà en 1993 que « la langue de l'Europe, c'est la traduction », lors des dixièmes Assises de la traduction littéraire, à Arles, comme le rappelle Barbara Cassin dans un texte inspiré par cette déclaration : « La langue de l'Europe? », Po&sic 2017/2-3 - N° 160-161. Voir QR code.

« Le centre de gravité de l'Internet est en train de se déplacer rapidement des langues européennes vers les langues asiatiques et l'arabe. »

Aujourd'hui, la réalité est que le centre de gravité de l'Internet est en train de se déplacer rapidement des langues européennes vers les langues asiatiques et l'arabe, tandis que les langues européennes approchent, pour certaines, des niveaux maximum de taux de connectivité, où une forte progression n'est donc plus possible, et que les langues africaines prennent un retard préoccupant dans cette évolution.

Un travail de regroupement des résultats du dernier modèle portant sur les 329 langues de plus de 1 million de locuteurs L1 montre les signes de cette évolution.

Le tableau se lit de cette manière : il y a 47 langues européennes de plus de 1 million de locuteurs (anglais, espagnol, français, etc.). 81 % des locuteurs de ces langues sont, en moyenne, connectés à l'Internet dans le monde. Ces locuteurs représentent ensemble près de 55 % de la présence linguistique dans l'Internet, avec des indicateurs de capacité et gradient moyens supérieurs à 1. Ensemble, les locuteurs (L1+L2) des langues d'origine européenne représentent 31 % de la population mondiale de locuteurs L1+L2 et 40 % des personnes connectées.

	Langues africaines	Langues américaines	Arabe	Langues asiatiques	Langues européennes	Reste des langues ⁴⁶
% moyen de locuteurs connectés	28,6 %	59,7 %	60,2 %	46,6 %	81,1 %	54,2 %
Puissance	2,55 %	0,19 %	3,11 %	35,71 %	54,93 %	3,47 %
Capacité	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 %
% population mondiale connectée	5,18 %	0,32 %	3,89 %	44,60 %	39,51 %	6,45 %
Nombre de langues avec L1>1M	138	8	1	135	47	0

Ainsi, en termes de personnes connectées, les langues asiatiques ont dépassé les langues européennes de 5 %, avec un potentiel de croissance très fort (47 % de locuteurs de langues asiatiques connectés versus la quasi-saturation pour les langues européennes). En même temps, les langues africaines, malgré le fait que ce sont les plus nombreuses dans la catégorie de plus de 1 million de locuteurs, montrent une très faible capacité, conséquence de taux de connexion moyens qui n'arrivent pas à dépasser les 30 %.

Les perspectives démographiques pour l'Afrique devraient assurer à ces langues, ainsi qu'aux langues européennes des pays africains (anglais, français et portugais principalement) une présence grandissante dans l'Internet, à condition que la fracture numérique soit surmontée.

⁴⁶ Il n'y a pas de langues du Pacifique avec plus de 1 million de locuteurs. La colonne « Reste des langues » rassemble les données de toutes les langues non incluses dans le modèle.