

7es Assises européennes du plurilinguisme

20-22 mai 2026 - Paris

Plurilinguisme et circulation des savoirs, des imaginaires
et des compétences. Quelles dynamiques et quelles vulnérabilités ?



Les langues et le multilinguisme dans l'Internet : état de l'art et perspectives

Daniel Pimienta, pimienta@funredes.org

Président de l'Observatoire de la
diversité linguistique et culturelle



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

<https://obdilci.org/>

CEUX QUI DÉTESTENT **LES CHIFFRES** VONT ME HAÏR !!!

DÉSOLÉ ! MAIS SOYEZ PATIENTS ET **DÉTENDEZ-VOUS** ...

PRENEZ-LES SIMPLEMENT À DISTANCE COMME

UNE MANIÈRE **IMPRESSIONNISTE** DE PERCEVOIR UNE RÉALITÉ INCONNUE



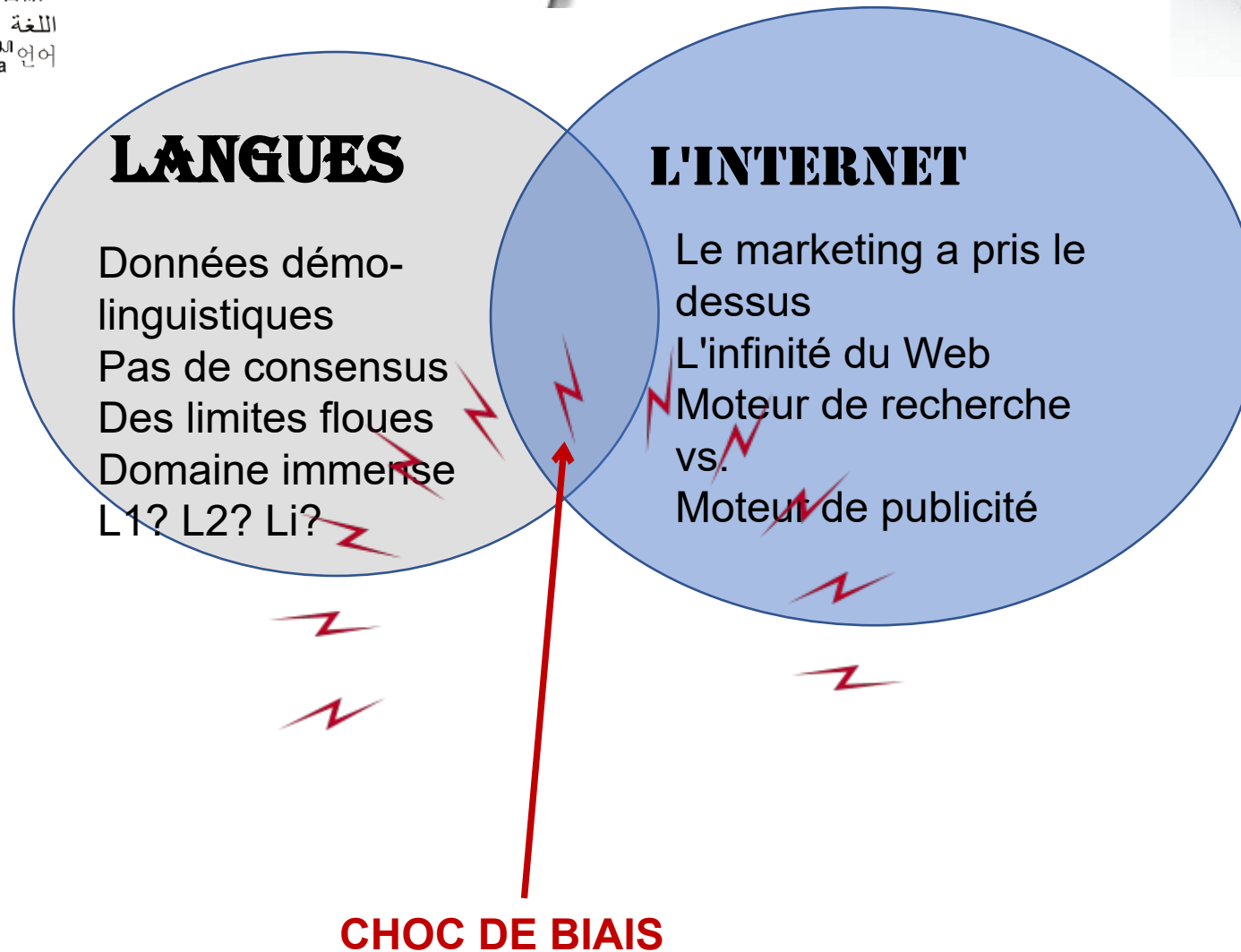
POLITIQUES PUBLIQUES

NE FAIT SENS
QUE SI DES **INDICATEURS** FIABLES ET RENOUVELABLES DE LA RÉALITÉ ACTUELLE
EXISTENT AFIN DE **DIAGNOSTIQUER** LA SITUATION
ET **MESURER L'IMPACT** DES POLITIQUES PAR LA SUITE

POLITIQUES PUBLIQUES POUR LES LANGUES EN LIGNE

UNE QUESTION TRÈS DIFFICILE
À L'INTERSECTION DE DEUX DOMAINES
EN MANQUE D'INDICATEURS FIABLES

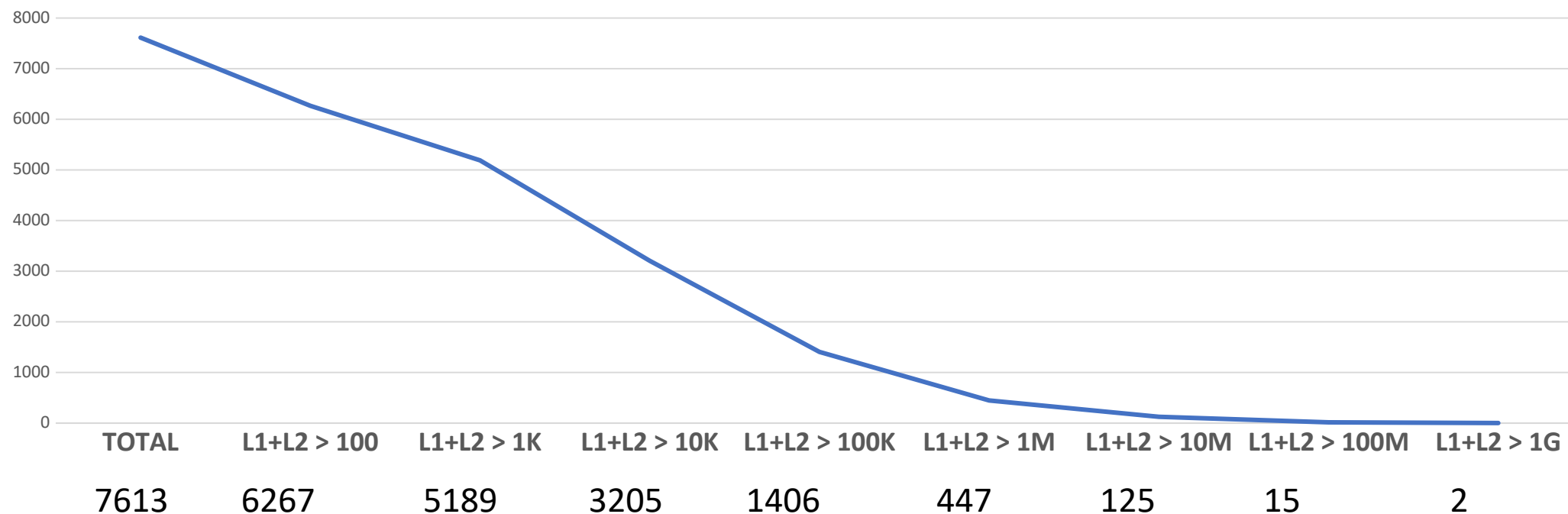




1 : Les langues dans l'Internet

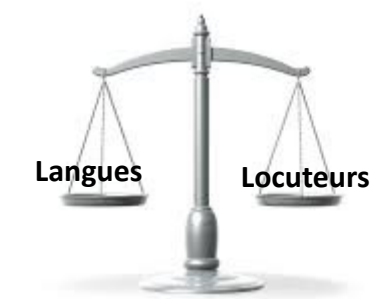


NOMBRE DE LANGUES PAR RAPPORT AU NOMBRE DE LOCUTEURS



Source : Ethnologue 2024, avant de regrouper les macro-langages

Le paradoxe ling



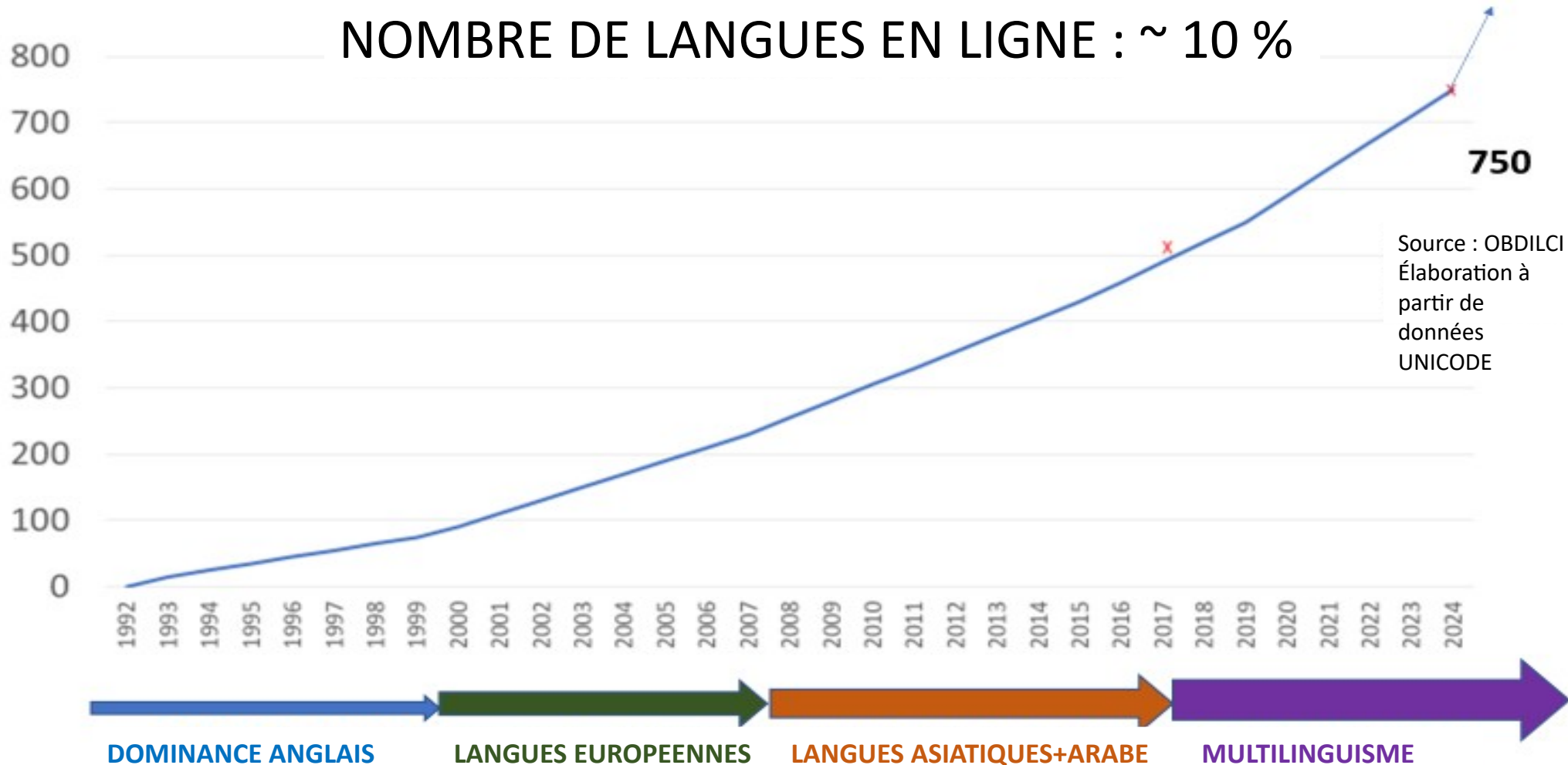
	L1>1M	TOTAL
LANGUES	336	7 226
%	4,6%	100%
LOCUTEURS L1	7 062 906 326	7 433 499 219
%	95,0%	100%
LOCUTEURS L1+L2	10 303 359 711	10 725 223 758
%	96,1%	100%

Taux de multilinguisme

1,43

Le paradoxe linguistique

NOMBRE DE LANGUES EN LIGNE : ~ 10 %





Le mythe de l'anglais comme lingua franca de l'INTERNET

ANGLAIS@MED

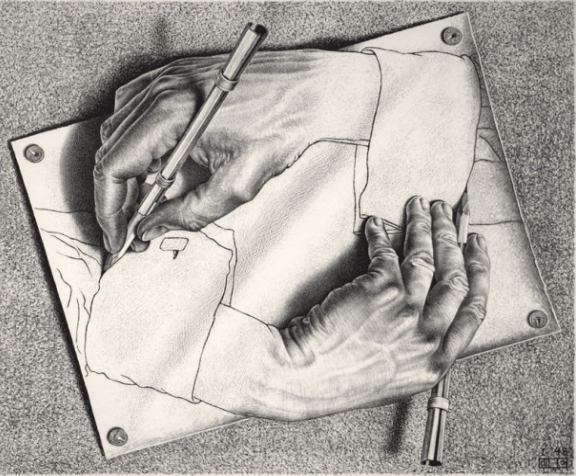
IA

Source :
W3Techs

% ANGLAIS @ WEB

SOURCE: OBDILCI





BIAIS CHRONIQUE SUR LES DONNÉES LINGUISTIQUES EN LIGNE IDÉES FAUSSES PROFONDES SUR LE **MULTILINGUISME**



METHODE : APPLICATION D'UN ALGORITHME DE DÉTECTION DE LANGUE
SUR LA PAGE D'ACCUEIL DU MILLION DE SITES WEB LES PLUS VISITÉS.

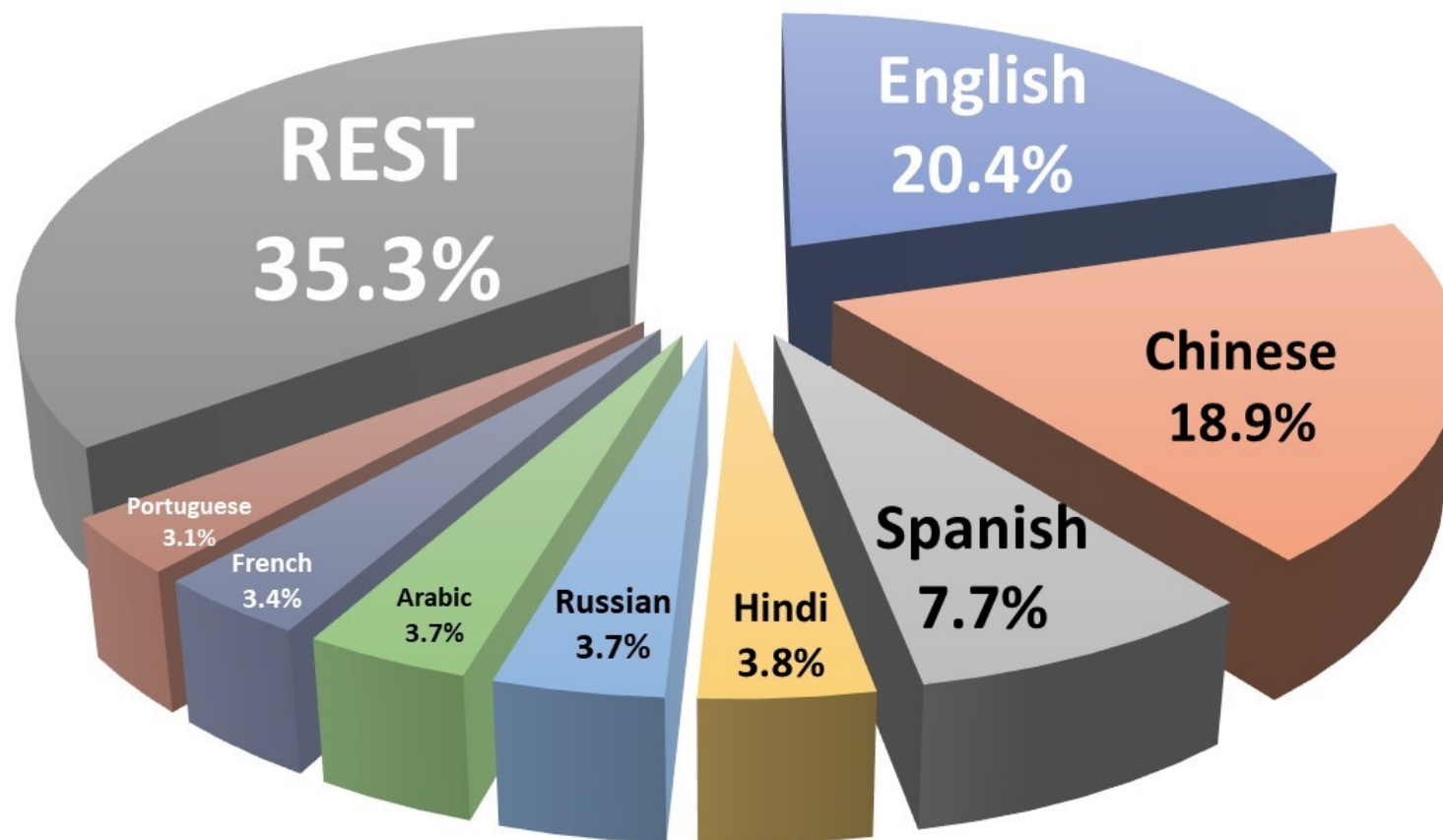
MAIS

NE PAS PRENDRE EN COMPTE LE FAIT QUE DE NOMBREUX SITES WEB SONT
MULTILINGUES

PROVOQUE UN **BIAIS** DE L'ORDRE DE GRANDEUR DE **100 %**
SUR LE POURCENTAGE DE RÉSULTATS EN ANGLAIS.

INDICATEURS OBDILCI

WEB CONTENTS %



La source des données est la dernière version du modèle OBDILCI pour créer des indicateurs de présence dans l'Internet
Pour les 361 langues comptant plus d'un million de locuteurs L1. <https://obdilci.org/main> + accès à la base de données <https://obdilci.org/base>

LANGUE.	LOC	LocC	% LocC	C	PV	PC
Anglais	15,79%	14,13%	70,86%	20,42%	1,45	1,29
Chinois	17,41%	14,48%	76,27%	18,88%	1.3	1,08
Espagnol	6,62%	5,22%	80,46%	7,70%	1,48	1.16
Hindi	4,34%	5,68%	48,48%	3,82%	0,67	0,88
Russe	3,28%	2,38%	87,42%	3,73%	1,57	1.14
Arabe	4,37%	4,08%	67,81%	3,65%	0,89	0,84
Français	3,05%	2,91%	66,58%	3,41%	1.18	1.12
Portugais	2,89%	2,46%	74,42%	3,09%	1,25	1.07
Japonais	1,54%	1,15%	84,98%	2,20%	1,91	1.42
Allemand	1,80%	1,25%	91,21%	2,15%	1,72	1.2
Malais	2,56%	2,29%	71,73%	2,03%	0,89	0,79
Bengali	1,90%	2,59%	47,11%	1,48%	0,57	0,78
Turc	1,12%	0,84%	85,36%	1,13%	1,34	1.01
Italien	0,85%	0,62%	87,61%	0,99%	1,59	1.16
Vietnamien	0,98%	0,80%	78,49%	0,96%	1.19	0,97
Ourdou	1,31%	2,22%	37,85%	0,91%	0,41	0,70
Persan	1,03%	1,03%	64,46%	0,88%	0,86	0,85
Coréen	0,78%	0,76%	66,02%	0,87%	1.15	1.12
Tagalog	0,90%	0,78%	74,16%	0,68%	0,87	0,76
Thaïlandais	0,79%	0,57%	89,51%	0,67%	1.18	0,84
Marathi	0,69%	0,93%	48,20%	0,58%	0,63	0,83
Télougou	0,68%	0,89%	48,99%	0,57%	0,64	0,84
Polonais	0,53%	0,39%	87,03%	0,56%	1,44	1.06
Tamoul	0,64%	0,81%	50,51%	0,54%	0,66	0,85
Javanais	0,69%	0,64%	69,26%	0,52%	0,81	0,75
Ukrainien	0,46%	0,36%	81,32%	0,43%	1.20	0,95
Néerlandais	0,35%	0,24%	96,00%	0,39%	1,66	1.11
Gujarati	0,44%	0,58%	48,43%	0,37%	0,64	0,85
Swahili	0,47%	0,91%	32,87%	0,34%	0,37	0,73
Kannada	0,41%	0,55%	48,25%	0,34%	0,62	0,83
RESTE	21,25%	27,48%		15,65%		

LOC = Pourcentage de locuteurs L1+L2 de cette langue, par rapport au total mondial L1+L2 .

LocC = Pourcentage de locuteurs L1+L2 de cette langue connectés à l’Internet, par rapport au total mondial de locuteurs L1+L2 connectés.

%LocC = Pourcentage de locuteurs L1+L2 de cette langue connectés à l’Internet, par rapport au nombre total de locuteurs L1+L2 de cette langue.

C = Pourcentage de contenus dans cette langue par rapport aux total des contenus mondiaux.

PV = Présence virtuelle, calculée comme C/LOC , le pourcentage de contenus divisé par le pourcentage de locuteurs, un indicateur sans dimension, normé à 1. Une valeur supérieure à 1 indique une bonne présence virtuelle (par rapport à la présence réelle).

PC = Productivité du contenu, calculée comme $C/LocC$, le pourcentage de contenus divisé par le pourcentage de locuteurs connectés, un indicateur sans dimension, normé à 1. Une valeur supérieure à 1 indique une bonne productivité.

EXAMINONS LES PREMIÈRES POSITIONS DES AUTRES INDICATEURS

Norvégien	98,82%
Danois	98,64%
Iban	97,42%
Néerlandais	96,00%
Limbourgeois	95,91%
Suisse allemand	95,78%
Irlandais	95,61%
Suédois	95,42%
Écossais	95,34%
Catalan	95,05%

TAUX CONNECTIVITÉ HAUTS

Norvégien	1,90
Japonais	1,89
Écossais	1,89
Irlandais	1,80
Danois	1,79
Estonien	1,75
Suédois	1,72
Allemand	1,72
Suisse allemand	1,69
Néerlandais	1,66
Finlandais	1,66
Catalan	1,61

PRÉSENCE VIRTUELLE SUPÉRIEURE

Japonais	1,43
Anglais	1,29
Écossais	1,27
Norvégien	1.23
Irlandais	1.21
Estonien	1.21
Allemand	1.20
Okinawa	1.18
Occitan	1.18
Macédonien	1.18
Danois	1.17
Grec	1.16
Italien	1.16
Espagnol	1.16
Suédois	1.16

PRODUCTIVITÉ DE CONTENU PLUS ÉLEVÉE

CYBER-GÉOGRAPHIE DES FAMILLES DE LANGUES

Famille de langues	Afrique	Amériques	arabe	Asie	Europe	Pacifique	Non inclus	TOTAL
% Locuteurs connectés	36,3%	69,0%	68,0%	60,8%	89,2%	62,2%	49,5%	64,2%
Contenus	4,2%	0,3%	3,6%	44,3%	45,9%	0,03%	1,7%	100%
Présence virtuelle	0,4	0,8	0,9	0,9	1.4	0,5	0,5	1
Productivité des contenus	0,6	0,8	0,8	0,9	1.2	0,7	0,7	1
Locuteurs	11,7%	0,4%	4,1%	48,4%	32,1%	0,05%	3,4%	100%
% Locuteurs connectés	6,7%	0,4%	4,3%	47,5%	38,5%	0,04%	2,6%	100%
Nombre de langues	152	8	1	147	51	2		361

2 : Le multilinguisme dans l'Internet



LA RÉALITÉ DU **MULTILINGUISME SUR INTERNET** EST À LA FOIS :

- LA CAUSE DES **ÉNORMES** BIAIS SUR LES DONNÉES LINGUISTIQUES DANS L'INTERNET
- LE **SECRET** LE MIEUX CACHÉ DE L'INTERNET



OBDILCI A RÉUSSI EN 2025 À LANCER **DES RECHERCHES** SUR CE SUJET :

- LES RÉSULTATS SONT **APPROXIMATIFS**
- CEPENDANT ILS SONT **SIGNIFICATIFS** ET PERMETTENT DE MONTRER UNE PREMIÈRE IMAGE CARACTÉRISÉE PAR
 - TRÈS **GRANDE VARIANCE** SELON LES PAYS ET LES LANGUES
 - DES CHIFFRES DE TOUTE FAÇON **ASSEZ ÉLEVÉS EN MOYENNE** ET UNE MODÈLE DE **CROISSANCE RAPIDE**
 - UNE CONCLUSION QUE L'INTERNET POURRAIT ÊTRE (OU DEVENIR BIENTÔT) L'ENDROIT LE **PLUS MULTILINGUE** AYANT JAMAIS EXISTÉ SUR TERRE.

SOURCE : BASE DE DONNÉES DE DATAPROVIDER.COM

RECHERCHES OBDILCI 2025 SUR LE MULTILINGUISME DANS L'INTERNET

<https://www.obdilci.org/projects/other/mlreports/>

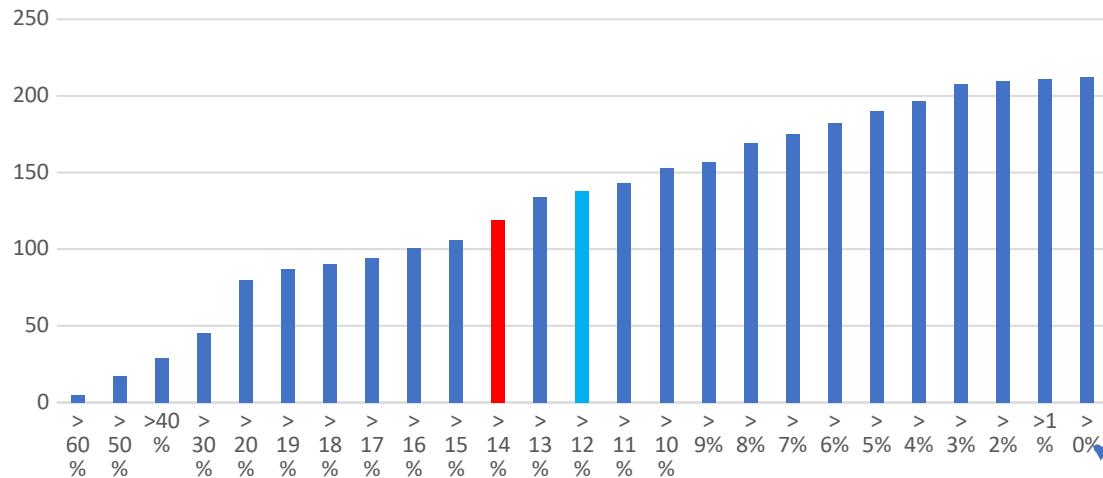


L'état du multilinguisme du WWW

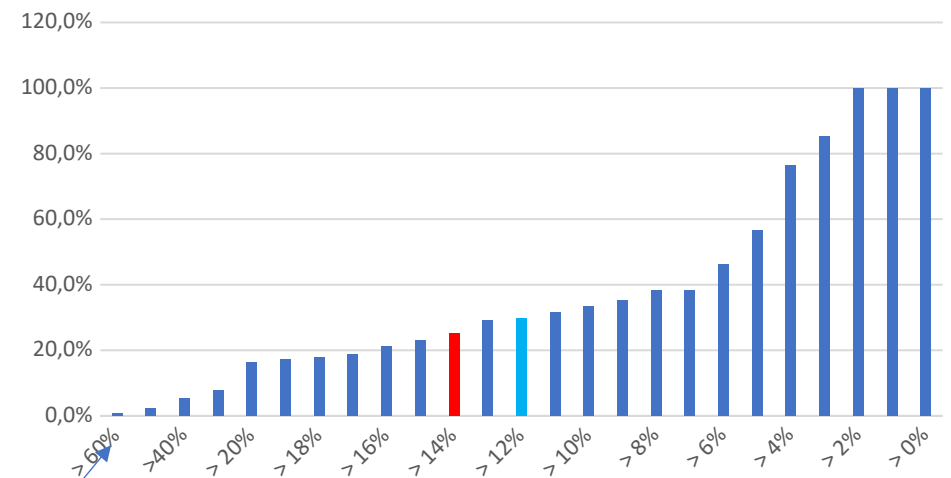
- L'état de multilinguisme de la Toile, rapport technique #7: [Propension à l'utilisation de la langue française dans l'écosystème Web des ccTLDs d'Afrique francophone](#). 10/2025
- Technical Report #6: [Wikimedia multilingualism](#), 9/2025
- Technical Report #5: [Web Multilingualism analyzed by ccTLD languages gTLDs and much more: winners and losers](#), 8/2025
- Informe No 4: [Una evaluación de la reciprocidad en el uso mutuo del español y del portugués en la web lusófona e hispanófona](#). 6/2025
- Rapport Rapport technique #3 : [Une caractérisation du Web francophone à partir d'une série de paramètres en comparaison avec d'autres langues dominantes sur la Toile](#). 5/2025
- Technical Report #2: [Exploring web presence and multilingualism of European minority languages with associated gTLDs](#). 4/2025
- Technical Report #1: [Approximation of the Rate of multilingualism of the WWW](#) 2/2025 [En français](#).

CORRÉLATION DU MULTILINGUISME DANS LE WEB PAR RAPPORT AUX PAYS OU À LA POPULATION

Nombre de pays par tranche de multilinguisme



% Population par tranche de multilinguisme



Indicateur : pourcentage de sites web multilingues

- **Faible multilinguisme dans 48 % des pays ...**
 - **MAIS**
 - **cela représente 76% des locuteurs**

MULTILINGUISME@WEB PAR CONTINENT ET RÉGION



RÉGION (*)	#WEB	MULTIL%
Europe	18 987 454	17,8%
Amérique du Nord	28 119 411	5 2%
Amérique latine et Caraïbes (**)	2 703 407	8 4%
Afrique	1 055 493	7 9%
Océanie (***)	1 350 548	4 3%
Pays arabes	417 094	43 3%
Asie	4 424 336	8 9%
Langue officielle anglaise	33 291 720	5 1%

PAYS	MULTIL
USA	4,8%
CANADA	13,8%

(*) Basé sur ccTLD et filtré par les domaines
piratés
(**) Inclut le Mexique
(***) Comprend l'Australie et la Nouvelle-
Zélande

LANGUES VERSUS MULTILINGUISME DANS LE WEB

LANGUE	#WEB	MULTIL
TOUS	95 612 368	10 95%
Basque	8 072	63 52%
Ukrainien	221 043	58 67%
Letton	43 428	53 81%
Catalan	82 966	52 05%
Estonien	57 245	50 89%
Grec	231 074	50 60%
Kazakh	9 285	45 58%
Lao	1 109	45 31%
Macédonien	10 607	39 15%
Lituanien	91 380	38 69%
Sindhi	9 350	35 16%
slovène	86 936	34 71%
Bulgare	130 100	34 24%
Azerbaïdjanais	21 617	32 78%
Tadjik	1 613	32 08%
Macro arabe	332 071	31 56%
Gallois	12 642	31 23%
Croate	86 665	28 95%
Yiddish	400	28 75%
Irlandais	2 510	26 59%
Bosniaque	119 615	26 49%
Amharique	1 210	25 83%
Finlandais	258 164	24 56%
Néerlandais	1 976 285	23 62%
Roumain	278 962	23 36%
Italien	2 092 445	22 84%
Mongol	9 289	22 28%
Birman	3 016	21 97%
Serbo-croate	11 821	21 93%



LANGUE	#WEB	MULTIL
Anglais	52 456 229	8 00%
Indonésien	1 220 050	7 94%
Xhosa	979	6 89%
Haoussa	3 237	6 72%
Afrikaans	20 039	6 70%
Espéranto	1 757	6 26%
Somali	2 622	6 20%
Haïtien	1 856	6 20%
Népalais	19 233	6 02%
ÉCOSSAIS	1 408	5 86%
Hmm	1 365	5 13%
Luxembourgeois	31 987	5 12%
Soundanais	6 099	4 34%
Marathi	33 902	4 31%
Japonais	3 341 482	4 16%
Shona	2 648	3 97%
Coréen	596 650	3 60%
Javanais	4 853	3 50%
Corse	3 353	3 43%
Zoulou	2 935	2 64%
Samoan	1 061	2 59%
Chinois	4 969 986	2,08%
Malgache	4 225	1 78%



PAYS (ccTLD) VERSUS MULTILINGUISME DU WEB

Nom du pays	ccTLD	#WEB	MULTIL
TOUS		87 199	
		527	14 39%
Monaco	MC	1 069	72 03%
Moldavie	MD	15 186	70 43%
Koweït	KW	1 091	63 93%
Ukraine	UA	211 516	61 74%
Mauritanie	M	561	60 61%
Luxembourg	LU	22 859	56 86%
Lettonie	LV	46 142	54 95%
Géorgie	GE	17 345	54 67%
Arabie Saoudite	SA	27 756	53 99%
Andorre	AN	1 052	53 47%
Arménie	SUIS	11 182	53 43%
Grèce	GR	243 677	53 43%
Jordanie	JO	1 552	53 00%
Estonie	EE	65 173	52 12%
Irak	QI	3 185	51 18%
Groenland	GL	1 062	51 08%
Cuba	CU	358	50 28%
Chypre	CY	7 704	48 94%
Egypte	EG	2 591	48 92%
Yémen	OUI	209	47 85%
Bahreïn	BH	1 005	46 52%
Porto Rico	PR	360	45 14%
Tadjikistan	TJ	1 633	43 94%
Birmanie	MM	1 052	43 73%
Bulgarie	BG	40 863	43 63%
Albanie	AL	11 114	43 32%
Oman	OM	1 796	42 73%
Macédoine du Nord	MK	13 737	42 13%
Azerbaïdjan	AZ	14 081	40 36%
Lituanie	LT	84 366	39 94%
Saint-Marin	SM	580	39 22%
Cambodge	KH	1 305	38 89%
Hong Kong	HK	47 269	38 59%
Portugal	PT	152 372	36 51%
Panama	PA	2 654	35 80%
Slovénie	SI	68 650	34 08%



Nom du pays	ccTLD	#WEB	MULTIL
Brésil	BR	1 767 293	6 96%
Philippines	PH	25 248	6 87%
Sri Lanka	LK	21 103	6 75%
Indonésie	ID	265 085	6 63%
Barbade	BB	272	6 43%
États-Unis	NOUS	270 045	5 78%
États-Unis (pays)	US	26 955 911	4 81%
Nouvelle-Zélande	NZ	223 600	5 55%
Pakistan	PK	55 940	5 43%
Îles Caïmans	KY	1 888	5 43%
Guinée équatoriale	GQ	1 286	5 25%
Tanzanie	TZ	13 773	5 21%
Japon	JP	900 409	5 11%
Royaume-Uni	UK	2 178 399	4 97%
Soudan du Sud	SS	151	4 97%
Inde	IN	903 802	4 49%
Îles Malouines	FK	56	4 46%
Kenya	KE	42 705	4 19%
Îles Vierges américaines	VI	122	4 10%
Sierra Leone	SL	316	3 96%
Australie	AU	1 121 532	3 93%
Bermudes	BM	1 081	3 70%
Papouasie-Nouvelle-Guinée	PG	952	3 68%
Nigeria	NG	59 913	3 61%
Bangladesh	BD	15 848	3 52%
Népal	NP	32 255	3 44%
Corée du Sud	KR	249 401	3 43%
Afrique du Sud	ZA	406 452	3 24%
Botswana	BW	3 038	3 21%
Lesotho	LS	803	3 11%
Chine	CN	861 667	2 73%



TLD, E -COMMERCE ET IMPACT ÉCONOMIQUE VS. MULTILINGUISME

GROUPE	#WEB	MULTIL
ccTLD	31 498 261	14 73%
gTLD	49 571 411	10 88%
Nouveau gTLD	5 521 510	8 92%
sTLD	200 022	15 24%
IDN	87 199 527	12 12%
Entreprise	60 839 086	15 63%
Non Entreprise	21 532 154	3 65%
B2C	26 589 834	16 93%
B2B	8 971 766	18 22%
Boutique en ligne	9 761 547	25 25%
Pas de boutique en ligne	77 437 980	10 46%



ÉCONOMIQUE IMPACT	#WEB	MULTIPLE
EI > 90	110	45 45%
80 < EI < 90	1 111	76 51%
70 < EI < 80	7 270	64 75%
60 < EI < 70	40 895	56 66%
50 < EI < 60	165 685	39 78%
40 < EI < 50	903 613	37 44%
30 < EI < 40	4 026 546	26 39%
20 < EI < 30	17 424 223	17 22%
10 < EI < 20	51 084 470	9 18%
EI < 10	4 927 618	2 21%
TOTAL/MOY	78 581 541	11 83%



PRINCIPALES CONCLUSIONS



- **Un multilinguisme Web élevé est corrélé à** **communauté hispanique**, forte empreinte économique, pays européens, pays arabes.
- **Un faible multilinguisme sur le Web est corrélé aux** pays anglophones et à certains pays asiatiques (Chine, Corée du Sud, Inde, Japon, Indonésie).
- **Faible multilinguisme sur le Web dans** 48 % des pays représentant 76 % des locuteurs
- **Gagnants linguistiques :** basque, ukrainien, letton, catalan, estonien et grec
- **Les perdants linguistiques :** le chinois, le coréen et le japonais
- **Gagnants des domaines nationaux :** Monaco, Moldavie, Koweït, Ukraine et Mauritanie
- **Perdants des domaines nationaux :** Zimbabwe, Chine, Afrique du Sud, Corée du Sud
- **Gagnants des domaines urbains :** Montréal Kiev Barcelone Bruxelles Hambourg Lisbonne
- **Perdants du domaine municipal :** Los Angeles, New York, Houston, Moscou, Tokyo
- **Gagnants des TLD :** la plupart des gTLD linguistiques et géographiques, ccTLDs, sites orientés vers les entreprises
- **Les perdants des TLD :** le reste des sites gTLD, sites non orientés vers les affaires
- **Singularité :** Luxembourg très élevé mais Luxembourgeois très faible
- **Singularité :** le portugais est bas malgré le Portugal élevé car le Brésil est très bas
- **Singularité :** l'Amérique du Nord est faible malgré un Canada moyen car les États-Unis sont très faibles

MULTILINGUISME DE WIKIMEDIA

ÉLÉMENT	THÈME
WIKIMEDIA	
Wikipédia	Encyclopédie en libre accès
WikiBooks	Livres éducatifs en libre accès
WikiSource	Bibliothèque en libre accès
WikiQuote	Citations en libre accès
WikiVersité	Ressources éducatives en libre accès
Wiktionnaire	Dictionnaire en libre accès
WikiNews	Dépôt d'actualités en libre accès
WikiVoyage	Guides de voyage en libre accès
WikiCommons	Photos et autres médias en libre accès
WikiData	Bases de données en libre accès
WikiSpecies	Base de données d'espèces en libre accès
Fonctions Wiki	Base de fonctions en libre accès
MédiaWiki	Base de logiciels en libre accès



WIKIMEDIA	NOMBRE DE LANGUES
Wikipédia	343
Wiktionnaire	174
WikiSource	81
WikiBooks	78
WikiQuote	74
WikiNews	31
WikiVersité	18
WikiVoyage	18

Source des statistiques linguistiques : https://meta.wikimedia.org/wiki/List_of_Wikipedias
[https://meta.wikimedia.org/wiki/Wikiquote /](https://meta.wikimedia.org/wiki/Wikiquote/)



PRÉSENCE EN TOUS	PRÉSENCE DANS TOUS SAUF 1	PRÉSENCE DANS TOUS SAUF 2
Chinois	Arabe	Bengali
Tchèque	Néerlandais	Catalan
Anglais	Espéranto	Indonésien
Finlandais	Hébreu	Limbourgeois
Français	Hindi	Norvégien
Allemand	Coréen	Serbe
Grec	Persan	Slovène
Italien	Polonais	Tamoul
Japonais	Roumain	Turc
Portugais	Suédois	Ukrainien
Russe		Vietnamien
Espagnol		



WIKIMEDIA

CLASSEMENT GENERAL

Anglais	22 71%
Allemand	9 95%
Français	7 98%
Italien	4 85%
Hébreu	3 78%
Polonais	3 40%
Russe	2 92%
Chinois	2 86%
Vietnamien	2 74%
Espagnol	2 26%
Japonais	2 08%
Portugais	2 03%
Néerlandais	1 69%
Persan	1 60%
Ukrainien	1 20%
Suédois	1 19%
Tchèque	1 17%
Arabe	0 96%
Serbe	0 96%
Finlandais	0 83%
Indonésien	0 80%
Turc	0 78%
Bengali	0 75%
Coréen	0 68%
Grec	0 62%
Norvégien	0 62%

Moyenne des pourcentages
dans une série d'indicateurs
Incluant toutes les applications
Wikimedia (sauf Wikinews).



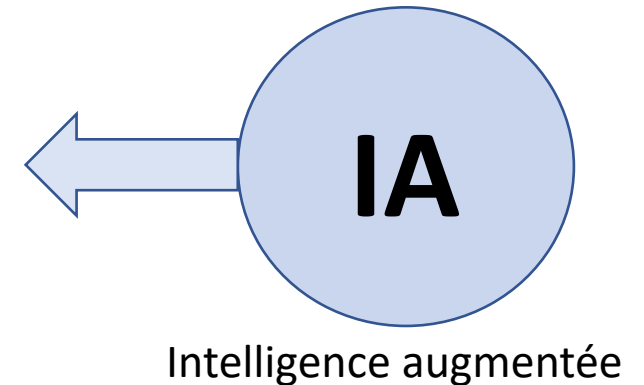
- **EST L'ENDROIT LE PLUS MULTILINGUE D'INTERNET**
- **IL N'EST PAS RÉSERVÉ AUX LANGUES PRINCIPALES**
- **CEPENDANT, LA PLUPART SE PORTENT BIEN ET L'ANGLAIS EST LE LEADER, À NOUVEAU AVEC UN CHIFFRE PROCHE DE 20 %**
- **CERTAINES LANGUES AUTOCHTONES FONT UNE PERCÉE**
- **CERTAINES LANGUES MENACÉES OU DISPARUES L'UTILISENT COMME MÉMOIRE**
- **C'EST UNE DÉMONSTRATION EN TEMPS RÉEL QUE LE MULTILINGUISME EST POSSIBLE**

3 : IA et langues numériques



IA ET DIVERSITÉ LINGUISTIQUE DANS L'INTERNET

- INTERNET EST DÉSORMAIS L'ESPACE LE PLUS **MULTILINGUE** JAMAIS CONSTRUIT
- LA LINGUA FRANCA DE L'INTERNET N'EST PAS L'ANGLAIS, MAIS **LA TRADUCTION**
- NOUS VIVONS **UNE RÉVOLUTION** DANS LE DOMAINE DE LA TRADUCTION
- ET LE DÉBUT D'UN **CHANGEMENT DE PARADIGME** AVEC LES LANGUES : DE L'AIDE À L'INTERCOMPRÉHENSION À **LA TRADUCTION ASSISTÉE PAR L'IA**



CEPENDANT

**LE DÉFI DE L'INTÉGRATION DES LANGUES MINORITAIRES DEMEURE IMMENSE
PLUS DE 95 % DES LANGUES = MOINS DE 5 % DE LOCUTEUR...**

AUJOURD'HUI : AIDE À L'INTERCOMPRÉHENSION

- CONSERVER LE FORMAT DANS L'AIDE À LA TRADUCTION DU DOCUMENT
- AIDE À LA TRADUCTION POUR LA CRÉATION/ÉDITION DE SITES WEB
- Intercompréhension dynamique automatique sur des sites Web en 250 langues
- Sous-titres automatiques sur YouTube en 250 langues
- INTÉGRATION DES SOUS-TITRES DANS LES RÉUNIONS ZOOM

DEMAIN : TRADUCTION AUTOMATIQUE

- DOCUMENTS TRADUIT INCLUANT DES IMAGES
- GÉRER LA TRADUCTION DANS LA CRÉATION/ÉDITION DE SITES WEB
- Traduction automatique de sites Web en 250 langues
- SOUS-TITRES AUTOMATIQUES SUR YOUTUBE EN 250 LANGUES
- INTÉGRATION DES SOUS-TITRES DANS LES RÉUNIONS ZOOM

- RÉUNIONS EN PRÉSENTIEL OU À DISTANCE, CHACUN UTILISE SA LANGUE MATERNELLE

CEPENDANT

QUE FAISONS-NOUS DES MILLIERS DE LANGUES LAISSÉES DE CÔTÉ, LES LANGUES AUTOCHTONES ?

- COMPRENDRE QUE PARTICIPER À LA VIE NUMÉRIQUE DEVRAIT ÊTRE **UN DROIT LINGUISTIQUE**
- COMPRENDRE L' ÉQUATION **ÉCONOMIQUE PARADOXALE**
- MOTIVER LES FAMILLES LINGUISTIQUES À **S'UNIR POUR** RELEVER LE DÉFI :

ALGIC : 56 langues, moins de 25 000 habitants. Ensemble, plus de 150 000 (Canada et États-Unis).

ESKIMO : 21 langues (Canada, USA, Groenland, Russie)

EYAK-ATHABASKAN : 47 langues, dont le Navajo et l'Apache, totalisant plus de 200 000 (Canada, États-Unis)

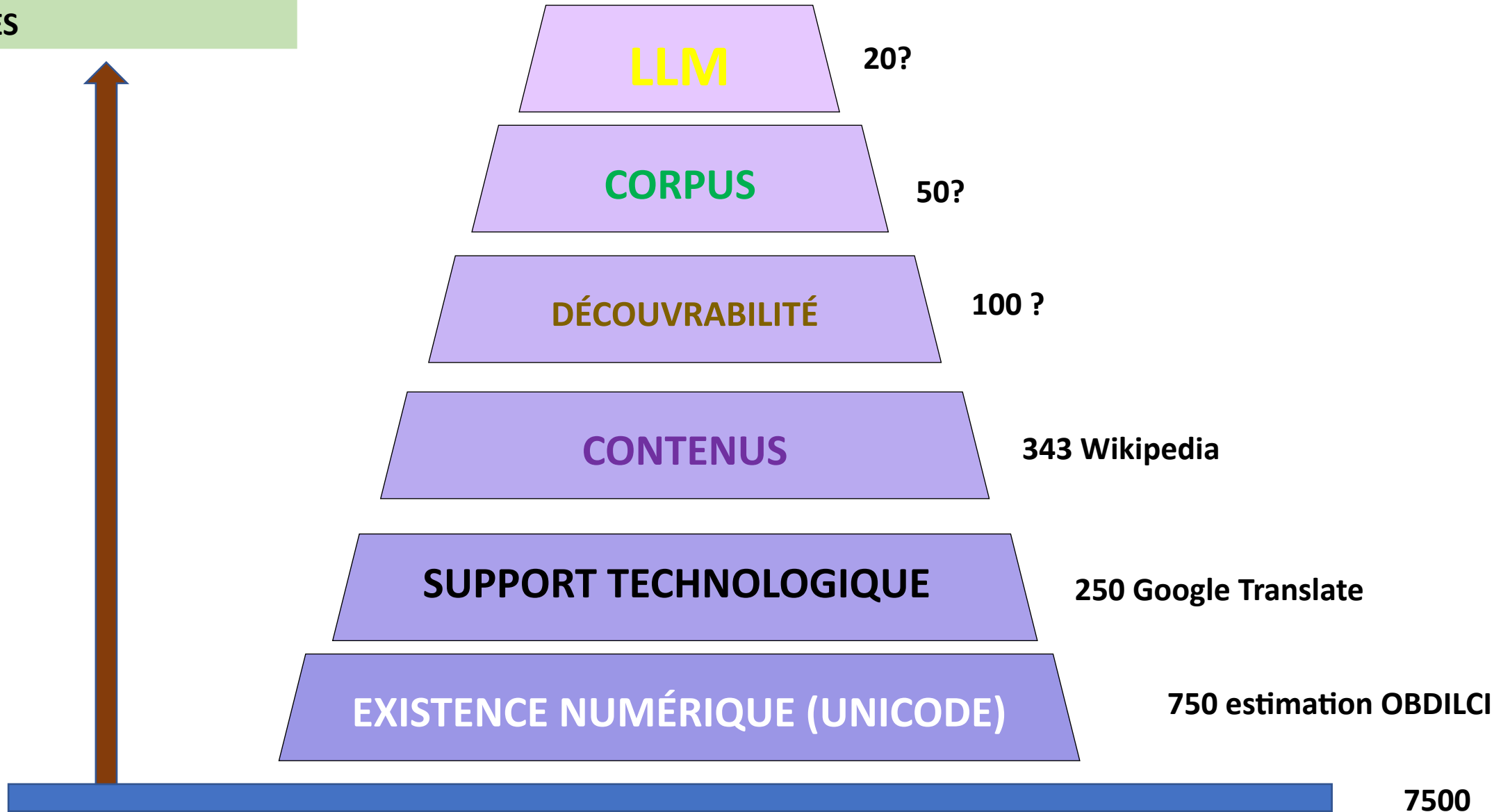
MAYA : 20 langues > 2,5 millions

OTO-MANGUÉEN : 148 langues, dont le mixtèque et le zapotèque, soit au total > 1,7 million

UTO-AZTECAN : 48 langues comme le nahuatl, au total > 1,9 million

TUPIAN : 63 langues au Brésil, au total > 75 000, proches du guarani (langue officielle du

**LES NIVEAUX DE NUMÉRISATION
DES LANGUES**



RÉFÉRENCES

Pimienta, D., 2021. *Internet and linguistic diversity: the cyber-geography of languages with the largest number of speakers*, dans *LinguaPax Review* 2021, Technologies linguistiques et diversité linguistique, page 9. En anglais, espagnol et catalan. - <https://linguapax.org/wp-content/uploads/2022/02/LinguapaxReview9-2021-low.pdf>

Pimienta, D., 2022. *Resource: Indicators on the presence of languages in Internet*. Actes de la 1re réunion annuelle du groupe d'intérêt ELRA/ISCA sur les langues sous-financées ; 24-25 juin 2022 ; Marseille, France. pp. 83–91. <https://aclanthology.org/2022.sigul-1.11/>
Version française : https://obdilci.org/wp-content/uploads/2024/01/Res.Ind_.lang_.Internet.fr_.pdf

Pimienta, D., Blanco, A., Müller de Oliveira, G., 2023. *The method behind the unprecedented production of indicators of the presence of languages in the Internet*, *Frontiers in Research Metrics and Analytics*. Vol. 8 - <https://doi.org/10.3389/frma.2023.1149347> -
Version française : https://obdilci.org/wp-content/uploads/2024/01/METHODV3.fr_.pdf

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 française : https://obdilci.org/wp-content/uploads/2024/11/FLS-7144-6-5.en_.fr_.docx

Pimienta, D. 2025. *Inventory and comparisons of all methods for the measure of languages online and implications on Internet's lingua franca* , LT4ALL2025, Paris-UNESCO, <https://obdilci.org/wp-content/uploads/2025/01/LangOnlineReport.docx>
Version française : https://obdilci.org/wp-content/uploads/2025/07/LangOnlineReport.en_.fr_.pdf

Pimienta, D. 2026. *La diversité linguistique dans l'Internet* in « De Babel à l'intelligence artificielle – Le plurilinguisme de Dante à nos jours », Christian Tremblay, José Carlos Herreras (coord.). Collection plurilinguisme, édité par l'Observatoire européen du plurilinguisme. ISBN : 9791042488130 – 1/2026 - <https://obdilci.org/wp-content/uploads/2026/01/BabelIA-DLI.pdf>