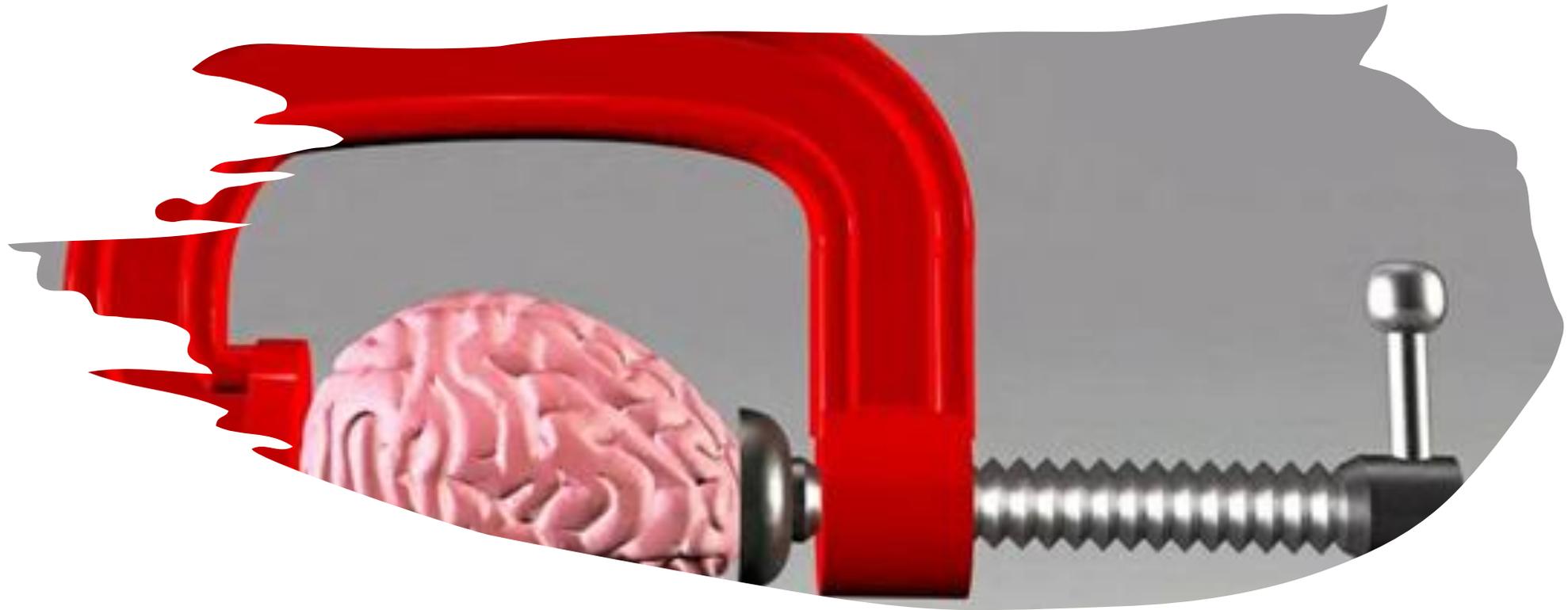


Chaire de recherche du Québec
Découvrabilité des contenus
scientifiques en français

La langue a-t-elle une influence sur les contenus scientifiques

Éric Charton Ph.D. , Centre de recherche informatique de Montréal





La langue n'est pas qu'un ensemble de mots et de verbes

La langue véhicule une culture, une vision du monde, une éducation : la langue
façonne aussi notre cerveau !



Chaire de recherche du Québec
Découvrabilité des contenus
scientifiques en français

Penser dans sa langue ...



Ski de randonnée **Vs** **Backcountry Skiing**



Chaire de recherche du Québec
Découvrabilité des contenus
scientifiques en français

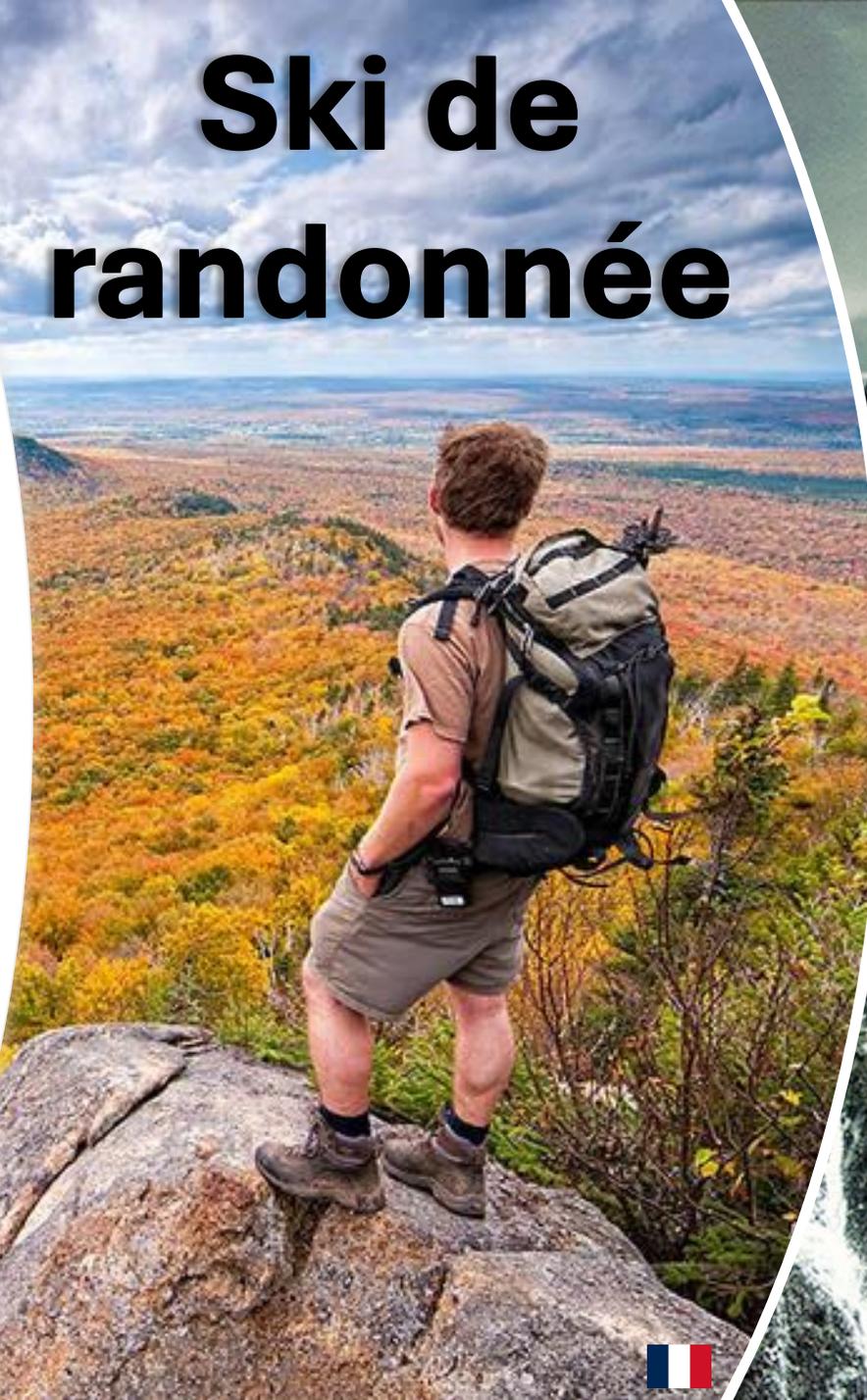
Penser dans sa langue ...



Ski de randonnée Vs Backcountry Skiing



Ski de randonnée



BACKCOUNTRY



Quelques exemples de propriétés d'une langue qui influent sur son locuteur

- **Les mots français qui n'existent pas en anglais (et vice versa)**

- dépaysement, pied à terre, frileux, affriolant, retrouvailles, flâner
- *Like* et *Love* (seul *aimer* existe en français)
- La *laïcité* n'est pas équivalente à la *sécularité*

- **Les langues sans mots ou les mots dans une seule langue**

- Le mot *art* n'existe pas en Inuktitut
- Le mot *Hipoda* en Bassa (Cameroun) signifie *cadeau minimum*
- Le mot *Saudade* en Portugais, sans équivalent en français et signifie *nostalgie profonde accompagnée d'espoir ou de désir*
- *Ragouler* en Créole Réunionnais : manger jusqu'au dégoût

- **Les mots ou locutions censés avoir le même sens dans deux langues (mais qui ne l'ont pas vraiment) ...**

- Ski de randonnée / Back country skying (ou Ski touring)
- Éventuellement (ca arrivera peut être) / *Eventually* (ca arrivera sûrement)
- N'importe quoi / Whatever ou nonsense

- **Les différences grammaticales**

- L'utilisation du masculin neutre en français
- L'absence du tutoiement en anglais
- Le nom de famille qui change de terminaison selon que l'individu est de sexe masculin ou féminin en Russe ou en Ukrainien



Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time

- Lea Borodistky, *Cognitive Psychology*, 2001
- Three experiments. English and Mandarin talk about time differently—English predominantly talks about time as if it were horizontal, while Mandarin also commonly describes time as vertical.
 - *In one study, Mandarin speakers tended to think about time vertically even when they were thinking for English.*
 - *Another study showed that the extent to which Mandarin–English bilinguals think about time vertically is related to how old they were when they first began to learn English.*
 - *In another experiment native English speakers were taught to talk about time using vertical spatial terms in a way similar to Mandarin.*

L'influence de la langue sur la compréhension des concepts est démontrée, mais quelles sont ses conséquences sur des domaines tels que la science, la culture, le développement économique ?

Quel impact des propriétés des langues sur le contenu scientifique ?

Un chercheur natif d'une langue pense et écrit selon les particularités de sa langue

- Quelques exemples :
 - Un chercheur en traitement du langage imagine des algorithmes différents pour **détecter les pronoms dans un texte** en français ou en anglais (le *tu* et le *vous* vs le *you*)
 - Le chercheur en **reconnaissance du locuteur** reconnaît la personne alors que celui qui travaille en **voice recognition** reconnaît la voix. C'est pourtant le même champ de recherche !
 - En annotation, **l'étiquetage sémantique** est l'équivalent de l'**entity linking** en anglais, mais conceptuellement le sens et l'entité sont des concepts distincts (influence Saussurienne)
 - Il n'est pas possible d'**exprimer directement dans un article en anglais certaines particularités grammaticales**, lexicales ou conceptuelles de la langue française :
 - Ordre des mots différents, prépositions plus nombreuses, genre neutre ...

Un travail de recherche écrit en français est donc subtilement différent une fois écrit en anglais (ou dans une autre langue)

Le processus de traduction fait donc discrètement perdre une partie de son contenu scientifique lié à la langue.

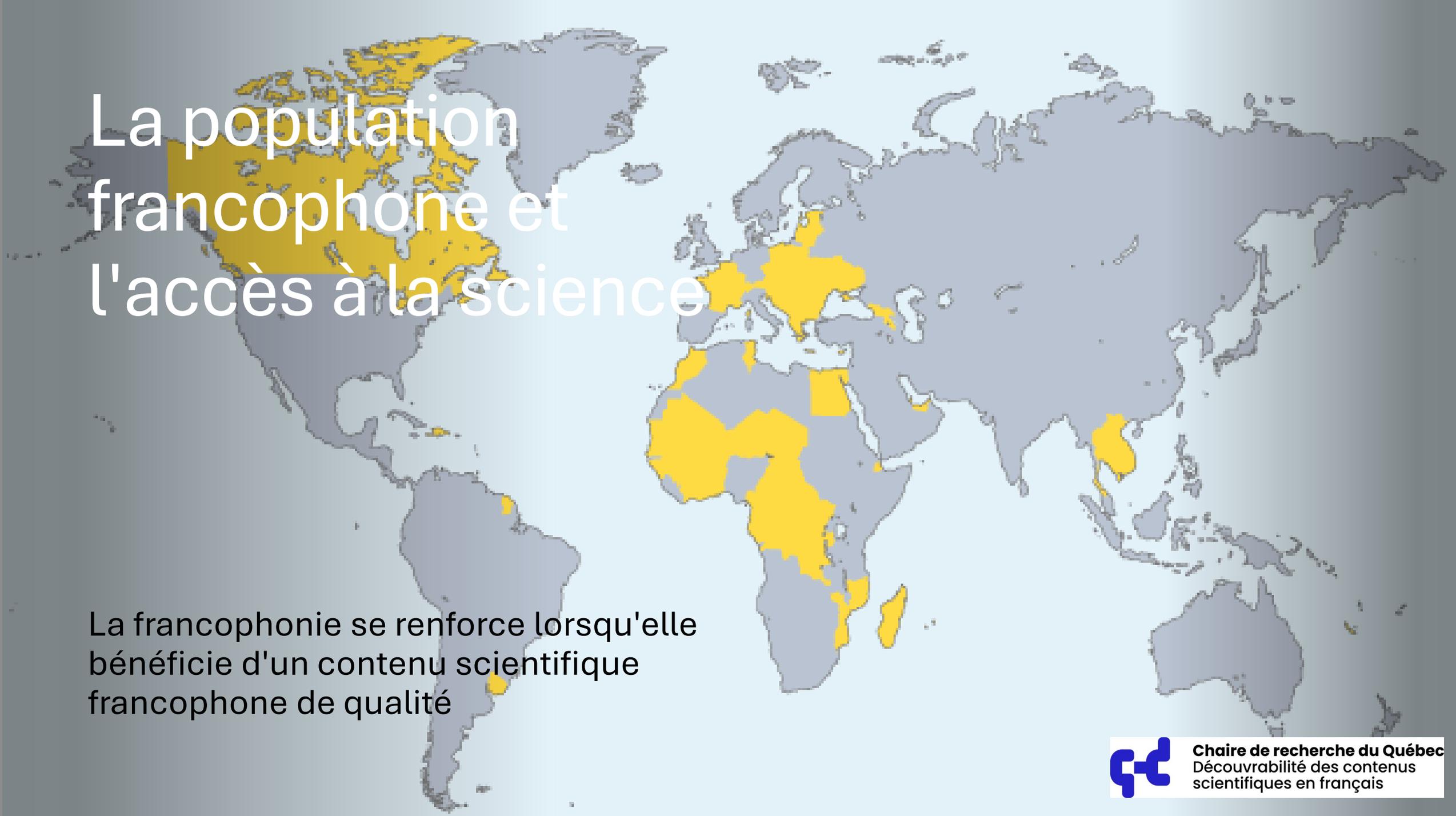
Le processus de version n'introduit pas par miracle la subtilité de la langue de destination.

Proposition d'axe de recherche : *La sémantique interlinguistique*

- Les différences de sens dans les contenus scientifiques de langues diverses créent-elles des biais ?
- Faut-il considérer ce biais dans les technologies du langage ?
- Ces différences ont-elles un impact sur l'usage des LLMs dans les sciences ?



Pourquoi c'est important

A world map where countries where French is an official language are highlighted in yellow. These include Canada, France, Belgium, Switzerland, Luxembourg, Monaco, Haiti, Benin, Burkina Faso, Mali, Niger, Chad, Cameroon, Gabon, Congo, Democratic Republic of Congo, Central African Republic, Ivory Coast, Senegal, Gambia, Guinea-Bissau, Guinea, Sierra Leone, Liberia, Ivory Coast, Mauritania, Mauritius, Madagascar, Réunion, and French Polynesia. The rest of the world is shown in light grey.

La population francophone et l'accès à la science

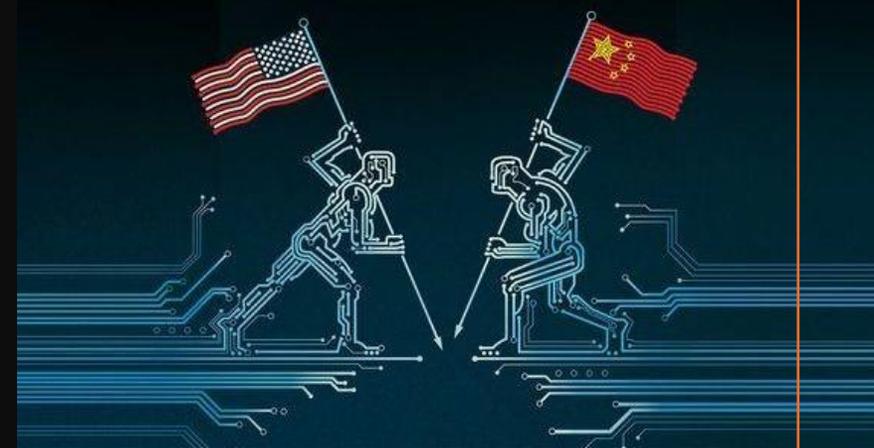
La francophonie se renforce lorsqu'elle bénéficie d'un contenu scientifique francophone de qualité



Chaire de recherche du Québec
Découvrabilité des contenus
scientifiques en français

Les atouts d'une existence de contenus scientifiques en français

- Un agenda scientifique non dicté par les besoins des grandes puissances (EU, Chine), ou de leurs industries (majors de l'IA)
- Une vivacité et une autonomie de la vie scientifique francophone (conférences, revues, sociétés savantes ...)
- Des champs de recherche plus exhaustif car ils ne sont pas contraints par les particularités culturelles lexicales et sémantiques de l'anglais ou du mandarin
- La découvrabilité ne peut exister ... que si le contenu à découvrir existe !



ST National Institute of
Standards and Technology
U.S. Department of Commerce



Merci !



Chaire de recherche du Québec
Découvrabilité des contenus
scientifiques en français

Peut-on étudier les 7117 langues parlées avec une théorie anglo-centrée ?

Qui a dit qu'un extraterrestre qui débarquerait sur terre et lirait les théories de Chomsky penserait que tous les Terriens parlent anglais ?

◆ Non, l'idée qu'un extraterrestre lisant les théories de Chomsky en conclurait que tous les Terriens parlent anglais est un malentendu. Voici pourquoi :

Ce que Chomsky dit réellement :

- **Grammaire universelle:** Chomsky propose que tous les humains possèdent une capacité innée à acquérir le langage, une sorte de "grammaire universelle" sous-jacente à toutes les langues.
- **Structure profonde vs. structure de surface:** Il distingue la "structure profonde" (commune à toutes les langues) de la "structure de surface" (les différences visibles entre les langues).
- **Langue comme organe:** Chomsky considère le langage comme un "organe" biologique, au même titre que le cœur ou le foie, et non comme un simple outil culturel.

Pourquoi cela ne veut pas dire que tout le monde parle anglais :

- **La grammaire universelle n'est pas l'anglais:** La grammaire universelle est un ensemble de principes abstraits, pas une langue spécifique. Elle permet d'acquérir n'importe quelle langue, pas seulement l'anglais.
- **Les langues divergent en surface:** Si toutes les langues partagent une structure profonde, elles divergent considérablement en surface (vocabulaire, grammaire, prononciation). L'anglais n'est qu'une de ces manifestations superficielles.
- **Chomsky ne s'intéresse pas à l'histoire des langues:** Son travail porte sur les mécanismes cognitifs du langage, pas sur l'évolution historique des langues ou leur répartition géographique.

L'idée d'un extraterrestre induit en erreur par Chomsky est un sophisme :

- **Analogie trompeuse:** On fait une analogie entre la grammaire universelle et l'anglais, alors que ce sont des choses distinctes.
- **Interprétation erronée:** On prête à Chomsky l'idée que la grammaire universelle est une sorte d'anglais caché, ce qui est faux.
- **Décontextualisation:** On sort les théories de Chomsky de leur contexte pour leur faire dire ce qu'elles ne disent pas.

Qui a dit qu'un extraterrestre qui débarquerait sur terre et lirait les théories de Chomsky penserait que tous les Terriens parlent anglais ?

 Copilot

Je n'ai pas trouvé de source spécifique attribuant cette citation à une personne en particulier. Cependant, cette idée est souvent discutée dans le contexte des critiques des théories linguistiques de Noam Chomsky, notamment en ce qui concerne l'universalité de la grammaire générative et la prédominance de l'anglais dans les études linguistiques.

Si vous avez d'autres questions ou si vous cherchez des informations supplémentaires, n'hésitez pas à me le faire savoir !



Chaire de recherche du Québec
Découvrabilité des contenus
scientifiques en français

L'accès à la science selon les compétences linguistiques

- **Au Québec**, entre les deux derniers recensements, la proportion de personnes ayant le français **comme seule langue maternelle** a diminué, passant de 77,1 % en 2016 à 74,8 % en 2021.
- Il n'existe pas de données précises sur le nombre exact de **francophones unilingues en France**. (selon les statistiques de l'Organisation internationale de la Francophonie 97 % de la population française parle le français – ce qui est assez peu surprenant ...)
- La langue française est parlée par plus de **115 millions de personnes en Afrique**. Une grande majorité de la population africaine (1,373 milliards d'habitants) parle une langue autre que le français.

Quelle langue pour la science Africaine ?

- En Afrique, il existe une grande diversité linguistique avec environ 2 000 langues parlées sur le continent
- Parmi ces langues, seules 197 sont utilisées et soutenues par des institutions autres que la maison et la communauté, par exemple en tant que langues officielles ou langues minoritaires reconnues



Aptitude à lire et écrire au Mali selon le recensement de 2009
(population résidente de 12 ans et plus)¹⁶

Aptitude à lire et écrire	Pourcentage ↕	Nombre ↕
Sait lire et écrire le français seul	22,56 %	1 917 501
Sait lire et écrire uniquement une langue nationale	1,66 %	141 325
Sait lire et écrire uniquement une autre langue	3,27 %	277 533
Sait lire et écrire le français et une langue nationale	2,32 %	197 141
Ne sait ni lire, ni écrire	59,25 %	5 035 502
Non déterminé	10,94 %	930 008
Total	100,00 %	8 499 010



Qui parle français en Iraq ?



Les langues officielles de l'[Irak](#) sont l'[arabe](#) et le [kurde](#), cette dernière ayant été ajoutée par la [constitution de 2004](#). L'[anglais](#) est surtout parlé en seconde langue par l'élite, et est la première langue étrangère enseignée devant le [farsi](#) (ou persan). Avec le retour de la paix en Irak, l'anglais sera très utile, car le pays a un fort potentiel touristique (monuments mésopotamiens, [architecture islamique](#), etc.). L'anglais est parfois utilisé dans l'administration, surtout dans les grandes villes, même s'il n'a pas le statut d'une [langue officielle](#).

Français et autres langues [\[modifier \]](#) [\[modifier le code \]](#)

- Selon l'[OIF](#), « Dans un Irak en reconstruction, la diffusion du français est restructurée. L'enseignement du français reste aussi difficile du fait de la présence forte de l'anglais. Une centaine d'écoles (sur un total de 4 000 à 5 000), dont la moitié située dans le gouvernorat de Bagdad, propose aujourd'hui le français comme seconde langue étrangère facultative. Le nombre total d'élèves oscillerait entre 2 500 et 3 000. »⁴.
- L'allemand, l'espagnol, et l'italien sont aussi trois langues proposées, avec un total d'élèves lui aussi de 2 500 et 3 000.
- Le russe pouvait lui aussi être étudié avant 2003, et surtout entre 1959 et 1992. La [première guerre du Golfe](#) fut fatale à l'enseignement de cette langue, mais toutefois, elle devrait bientôt être enseignée de nouveau dans le pays, vu que l'Irak a des liens commerciaux et diplomatiques avec la fédération de Russie.