

## Le multilinguisme dans la recherche

Analyse I/IST n°34 - Décembre 2020

« La recherche est internationale » : signé en 2019 par 120 organisations, l'initiative d'Helsinki marque un tournant important dans la reconnaissance du plurilinguisme en recherche<sup>1</sup>. Depuis plusieurs décennies, la publication de documents scientifiques en d'autres langues que l'anglais est plutôt appréhendée comme un *problème* qui gêne le bon fonctionnement des grandes bases de données commerciales comme le *Web of Science* ou *Scopus* (Sivertsen, Gunnar 2018). L'initiative d'Helsinki prend ce problème à l'envers : la diversité linguistique est une « bonne chose ».

Ce nouveau mouvement de préservation du multilinguisme est une conséquence indirecte du développement du libre accès. La disparition des *paywalls* rend la recherche techniquement accessible à tout le monde et non seulement au monde académique. Cette ouverture reste cependant limitée par d'autres barrières plus intangibles. L'usage généralisé de l'anglais, même sur des publications portant sur des problématiques fondamentalement locales, restreint la circulation de l'information scientifique.

Le multilinguisme permet de continuer à mener des recherches pertinentes au niveau local. Préservons-le! La diffusion des résultats de la recherche dans notre propre langue crée de l'impact. Soutenons-la! Il est essentiel d'interagir avec la société et de partager les connaissances au-delà des milieux universitaires<sup>2</sup>.

Cette note de synthèse fait le point sur un mouvement de fond émergent qui pourrait transformer profondément les conditions de publication scientifique.

### L'écriture scientifique avant l'anglais : une tour de Babel ?

Depuis près d'un demi-siècle, la production scientifique de nombreuses disciplines s'écrit presque exclusivement en anglais. De nombreuses revues publiées dans d'autres langues ont disparu ou se sont reconverties. En 1998, sept revues historiques de sciences physiques publiées dans sept langues différentes (*Acta Physica Hungarica*, *Anales de Física*, *Il Nuovo Cimento*, *Journal de Physique*, *Portugaliae Physica* et *Zeitschrift für Physik*) ont ainsi fusionné pour former *European Physical Journal* une revue européenne publiant exclusivement en anglais.

---

<sup>1</sup> « Helsinki Initiative on Multilingualism in Scholarly Communication », <https://www.helsinki-initiative.org/>

<sup>2</sup> Version française de l'initiative d'Helsinki.



# EPRIST

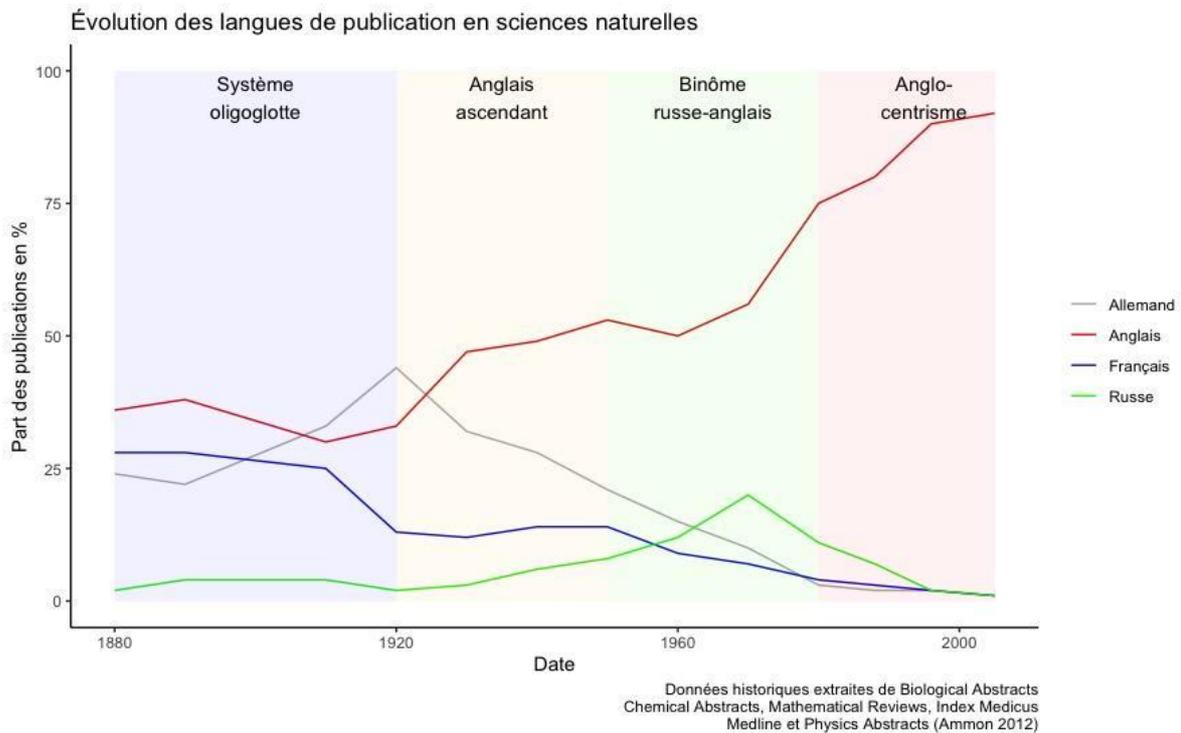
Ce monolinguisme exclusif n'a pas de précédent historique. Dans une synthèse très documentée publiée en 2015, l'historien Michael D. Gordin montre que le monde scientifique a toujours été plurilingue à différents degrés. Les précédentes *lingua franca* s'associaient toujours avec d'autres grandes langues véhiculaires et, plus ponctuellement, avec des langues locales : la recherche scientifique prend la forme d'une sorte de tour de Babel (*scientific babel*) s'étendant par cercles concentriques à partir de quelques idiomes de référence.

L'usage du latin en Europe jusqu'au XVIIIe siècle est un cas d'école. À la différence de l'anglais contemporain, le latin médiéval et renaissant est une langue *déterritorialisée* qui n'a pas de locuteur natif et son usage est limité à certains aspects de la vie quotidienne. Notamment le latin n'a jamais été une langue d'affaires, laissant le champ libre aux correspondances commerciales italiennes, hollandaises ou anglaises. Il n'était pas davantage une langue globale et dans d'autres espaces géographiques que l'Europe, d'autres langues de prestige pouvaient jouer un rôle similaire.

En raison de son usage circonscrit, le latin n'a jamais été la seule option. Le financement de la recherche repose alors fortement sur le mécénat et les soutiens potentiels ne lisaient pas forcément le latin. Au gré des stratégies personnelles, les intellectuels, scientifiques et inventeurs sont amenés à s'exprimer successivement en latin et dans une ou plusieurs langues vernaculaires (à l'image de Galilée passant de l'italien au latin). À l'échelle collective, l'emploi du latin fluctue fortement selon les lieux, les époques et les cultures scientifiques locales, avec des mouvements d'adoption et de reflux parfois difficilement explicables. Par exemple, les médecins français reviennent massivement au latin au XVIIe siècle après une phase de bilinguisme avec le français au siècle précédent.

Après son déclin définitif amorcé entre 1700 et 1850 selon les différents pays d'Europe, le latin n'a pas de successeur. La *lingua franca* se recompose autour d'un système oligoglotte centré sur trois langues majeures, le français, l'allemand et l'anglais. C'est aussi la période où se multiplient les projets de langue scientifique universelle, depuis la langue philosophique de John Wilkins (1668) jusqu'à l'Esperanto. Ces essais infructueux annonçaient néanmoins la situation historique actuelle où toute la recherche s'exprime dans une seule *langue universelle*.

Tout comme le latin n'a pas connu de déclin progressif continu depuis l'introduction de l'imprimerie, ce système oligoglotte ne disparaît pas d'un coup. Il se décompose et se reforme plusieurs fois sous l'effet de tendances de fond (la consolidation de l'anglais comme langue globale) et de grandes ruptures historiques (comme les deux conflits mondiaux). La transformation la plus inattendue a lieu après la Seconde Guerre Mondiale avec l'émergence du russe comme grande langue de recherche (Aronova 2017). Au moins un cinquième des articles consignés dans plusieurs grandes bases en sciences naturelles et médicales sont publiés en russe au cours des années 1950. Encore aujourd'hui, plusieurs disciplines perpétuent le mythe du mystérieux article russe du milieu du XXe siècle qui aurait déjà effectué l'expérience très spécifique sur laquelle un ou plusieurs chercheurs étaient en train de travailler.



## Évolution des préférences linguistiques en sciences naturelles dans plusieurs grandes bases de données. Visualisation d'après les données de Ammon (2012)

La domination sans partage de l'anglais est une réalité historique tardive qui ne s'affirme pas définitivement avant les années 1970 voire 1980 et reste une anomalie dans l'histoire longue de l'écrit scientifique. L'informatisation précoce de l'édition scientifique a été un facteur déterminant. Dès sa création, au début des années 1960, l'index de citation d'Eugene Garfield est anglo-centré: de nombreuses revues russes de référence ne sont pas intégrées, alors que c'était à l'époque la seule langue de recherche capable de concurrencer l'anglais (Wouters 1999, 89). L'intégration de ressources multilingues représente un investissement important pour ces premiers essais de moteur de recherche, où une bonne partie de l'indexation est effectuée à la main. Le prestige considérable de ces outils incite plutôt les revues à changer en profondeur les usages. Aujourd'hui, les revues scientifiques ne peuvent pas être incluses dans le *Web of Science* si elles ne publient pas au moins un résumé en anglais par article. Cette prédilection pour l'anglais s'est insinuée dans la plupart des grands programmes internationaux de financement de la recherche, en particulier au niveau européen (Ochsner et al. 2020, 32).

Significativement, les principaux moteurs de recherche académique n'ont aujourd'hui pas de champ de recherche par langue. Sur Google Scholar, il est tout au plus possible de

"personnaliser" les résultats à l'échelle d'un pays, ce qui peut faire remonter des publications en d'autres langues que l'anglais mais ne donne aucune idée de leur ampleur<sup>3</sup>. Sur le nouveau concurrent émergent de Scholar, Dimensions, la recherche par langue est tout simplement exclue.

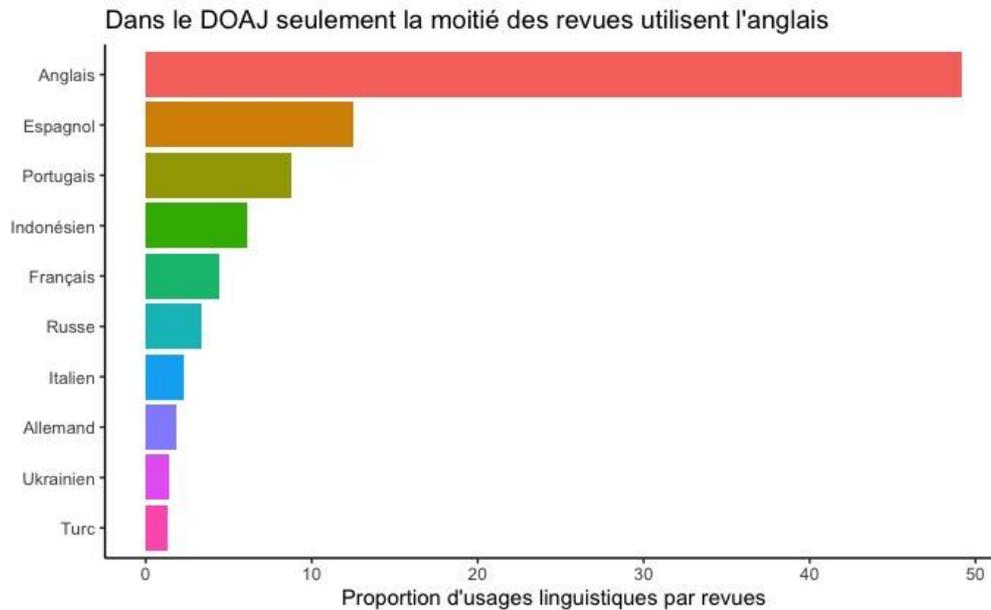
## Le plurilinguisme a-t-il disparu ou est-il devenu invisible?

Le graphe précédent, censé résumer l'histoire du multilinguisme en recherche, est en partie faux. Il existe aujourd'hui des bases de données massives enregistrant des millions d'articles scientifiques depuis deux siècles, mais aucune n'est exacte. La généralisation de l'anglais n'a pas seulement affecté le présent mais durablement brouillé notre image du passé. Les principaux index scientifiques restent les supports majeurs des enquêtes bibliométriques, bien qu'ils privilégient même rétroactivement les productions en anglais et, encore plus crucialement, qu'ils sous-évaluent massivement des champs disciplinaires préservant une forme de plurilinguisme à cause justement de leur plurilinguisme. Paradoxalement, l'expansion récente des collections indexées a eu pour effet de faire « décroître » la part de l'anglais : le *Web of Science* accorde ainsi plus de place aux revues en espagnol ou en portugais depuis la fin des années 2000 (Liu 2017).

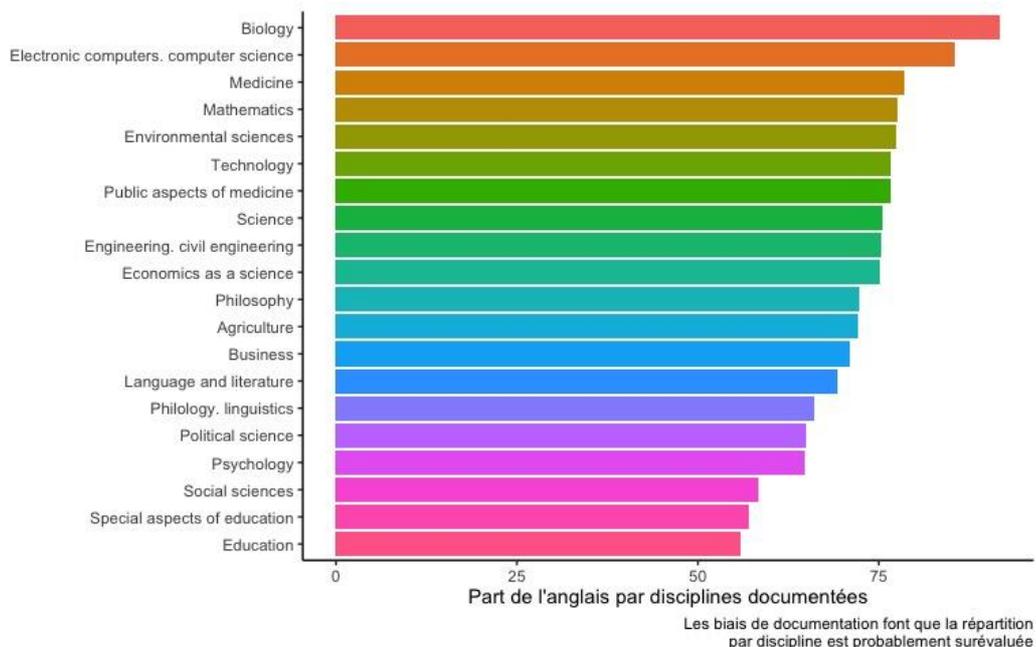
La visualisation suivante met en évidence une situation très différente. Elle repose sur les métadonnées du *Directory of Open Access Journal* qui recense les revues en libre accès selon une méthodologie considérablement plus ouverte et internationalisée que le *Web of Science* ou *Scopus*. L'anglais n'est pas l'unique langue par défaut ; ce n'est même pas la langue majoritaire avec un peu moins de 50% des revues concernées. Six langues sont représentées par au moins 500 revues actives : l'Espagnol (2922 revues, soit 13%), le Portugais (2053 revues), l'Indonésien (1423 revues), le Français (1033 revues), le Russe (789 revues) et l'Italien (541 revues)..

---

<sup>3</sup> <https://pitt.libanswers.com/faq/162088>



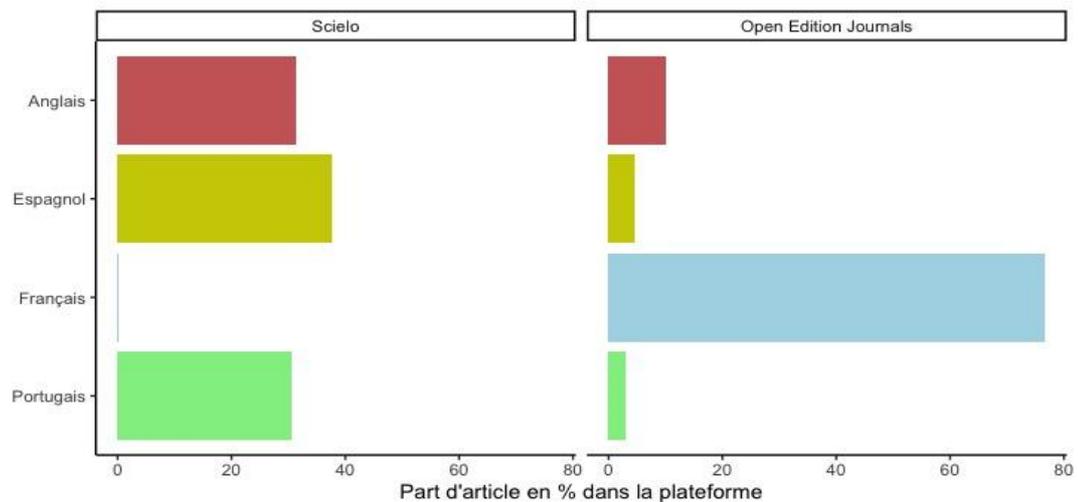
À la différence des grandes bases commerciales, le DOAJ a une meilleure couverture des revues de sciences humaines et sociales. Ce n'est toutefois pas le seul facteur. Nous avons croisé l'usage de la langue avec la discipline revendiquée par la revue. Malgré les limites des données disponibles (les disciplines ou sujets ne sont pas uniformisés), la diversité linguistique reste tangible dans des domaines très variés :



**Part de l'anglais dans les principales disciplines déclarées par les revues du DOAJ**

# EPRIST

La diversité linguistique du DOAJ résulte, pour partie, du développement récent de grandes plateformes de publication où l'anglais n'est pas la langue prédominante. Sur la bibliothèque numérique *Scielo*, il n'est qu'une langue de référence parmi d'autres (*Scielo*, où l'on constate une répartition quasi égale avec l'Espagnol et le Portugais). Dans les revues d'*Open Edition Journals*, son usage est même largement minoritaire, très loin derrière le français :



L'analyse de publications rendues invisibles dans les grandes bases commerciales met en évidence la diversité des modèles éditoriaux ancrés dans des territoires et des cultures scientifiques différenciés où l'usage par défaut de l'anglais ne va pas de soi. Cette préservation du multilinguisme peut s'expliquer par la perpétuation d'une autre langue véhiculaire. Le choix de langue peut être aussi déterminé par un écosystème plus large de circulation de la recherche scientifique impliquant des locuteurs natifs qui ne maîtrisent pas forcément l'anglais ou du moins pas l'anglais spécialisé des publications scientifiques.

Dans de nombreux domaines, l'usage de la langue reste une question de situations, de contextes et de supports. Les études sociologiques sur les pratiques de publication des chercheurs dévoilent une réalité très différente des données d'ensemble extraites des grands index bibliométriques : dans sept pays européens avec des langues locales à diffusion limitée le multilinguisme est non négligeable, un tiers des auteurs en sciences humaines et sociales publient communément dans au moins deux langues différentes (Kulczycki et al. n.d.). Seulement 25-30% de la production scientifique étudiée est accessible sur *Scopus* ou *Web of Science* : le contenu occulté provenant très largement d'autres langues que l'anglais.

## Uniformisation ou diversification ? L'écrit scientifique à la croisée des chemins

Contrairement aux apparences, l'écriture scientifique n'est pas monolingue aujourd'hui. L'anglais joue évidemment un rôle fondamental de langue véhiculaire servant à la communication internationale. Le manque de diversité linguistique se situe à un autre plan : dans l'inadéquation entre des infrastructures d'indexation qui font toujours autorité aujourd'hui et la diversité des usages en pratique. En ce sens, la déclaration d'Helsinki n'est pas une initiative désespérée mais une invitation à reconsidérer le paysage effectif de l'édition scientifique aujourd'hui, au-delà des effets de prestige induits par quelques indicateurs et moteurs de recherche de référence.

L'accessibilité des documents scientifiques multilingue pourrait être également transformée par les développements récents d'une technique ancienne: la traduction automatique (Gordin 2015, 323). Ironiquement, les premiers projets dans ce domaine, au début des années 1950, visaient déjà à améliorer l'accès aux publications scientifiques en russe aux États-Unis — avec un succès alors limité. Depuis quelques années, la traduction automatique de textes spécialisés a beaucoup progressé grâce à l'emploi de nouveaux algorithmes de *deep learning* capable d'identifier automatiquement le lexique, le style et les usages caractéristiques d'une discipline scientifique. Les meilleurs outils actuels comme *DeepL* commencent à donner des résultats suffisamment corrects pour donner un aperçu général d'une publication ou un premier jet de traduction à corriger.

L'impact de la traduction automatique reste cependant difficile à estimer. Des études récentes pointent aussi le risque d'un « multilinguisme uniforme » (Ramati and Pinchevski 2018). La traduction depuis ou vers l'anglais est, de longue date, la direction prioritaire. Depuis 2016, la généralisation des modèles de *deep learning* permet, en principe, de mieux intégrer les usages propres à chaque langue même si l'anglais reste de loin la langue la mieux documentée et celle sur laquelle se focalisent la plupart des tests d'évaluation. Malgré ses limites, la traduction automatique pourrait créer un contexte moins défavorable à la publication de recherches en d'autres langues que l'anglais.

Plutôt que par une solution purement technique, le renouveau du multilinguisme en science pourrait être encouragé par un mouvement social et collectif : la *bibliodiversité*. Ce concept a été initialement élaboré par des éditeurs indépendants latino-américains en réaction à la concentration industrielle du secteur (Hawthorne 2016). Il a été récemment élargi à l'édition scientifique, notamment dans le cadre de l'appel de Jussieu pour la science ouverte et la bibliodiversité. De nouveaux acteurs majeurs de l'écosystème comme Scielo ou Open Edition ne sont pas seulement divers sur le plan linguistique mais aussi par leur structure, leur modèle économique et leurs productions éditoriales, qui ne se limitent pas aux articles de revues mais intègre potentiellement de nombreuses autres formes (livres, textes de vulgarisation,

rapports...). Le développement de la bibliodiversité scientifique contribue ainsi à valoriser des pratiques demeurées jusqu'à présent marginales et peu visibles.

## Références

- Aronova, Elena. 2017. "Russian and the Making of World Languages during the Cold War." *Isis* 108 (3): 643–50. <https://doi.org/10.1086/694163>.
- Gordin, Michael D. 2015. *Scientific Babel: How Science Was Done Before and After Global English*. University of Chicago Press.
- Hawthorne, Susan. 2016. *Bibliodiversité, Manifeste Pour une Édition Indépendante*. 1st Edition. ECLM.
- Kulczycki, Emanuel, Raf Guns, Janne Pölonen, Tim C. E. Engels, Ewa A. Rozkosz, Alesia A. Zuccala, Kasper Bruun, et al. 2020 "Multilingual Publishing in the Social Sciences and Humanities: A Seven-Country European Study." *Journal of the Association for Information Science and Technology*, <https://doi.org/10.1002/asi.24336>.
- Liu, Weishu. 2017. "The Changing Role of Non-English Papers in Scholarly Communication: Evidence from Web of Science's Three Journal Citation Indexes." *Learned Publishing* 30 (2): 115–23. <https://doi.org/10.1002/leap.1089>.
- Ochsner, Michael, Nina Kancewicz-Hoffman, Marek Hołowiecki, and Jon Holm. 2020. *Overview of Peer Review Practices in the SSH*. figshare. <https://doi.org/10.6084/M9.FIGSHARE.12032589>.
- Ramati, Ido, and Amit Pinchevski. 2018. "Uniform Multilingualism: A Media Genealogy of Google Translate." *New Media & Society* 20 (7): 2550–65. <https://doi.org/10.1177/1461444817726951>.
- Sivertsen, Gunnar. 2018. "Balanced Multilingualism in Science." *BiD: Textos Universitaris de Biblioteconomia i Documentació*, no. 40.
- Wouters, P. F. 1999. "The Citation Culture." <https://dare.uva.nl/search?identifier=b101b769-100f-43e5-b8d2-cac6c11e5bbf>.