



HAL
open science

Rapport final CybergeonNet - Traductions scientifiques

Christine Kosmopoulos, Natacha Aveline-Dubach, Colette Cauvin-Reymond,
Bernard Elissalde, Maria Gravari-Barbas, Margaux Hardy, Nathalie
Lemarchand, Francisco Maturana, Denise Pumain, Benjamin Wayens, et al.

► To cite this version:

Christine Kosmopoulos, Natacha Aveline-Dubach, Colette Cauvin-Reymond, Bernard Elissalde, Maria Gravari-Barbas, et al.. Rapport final CybergeonNet - Traductions scientifiques. [Rapport de recherche] Ministère de la Recherche, de l'Enseignement Supérieur et de l'Innovation; ABES. 2021, 64 p. + 323 p. (annexes). halshs-03419355

HAL Id: halshs-03419355

<https://shs.hal.science/halshs-03419355>

Submitted on 8 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cybergeonet

2019 - 2021

31/10/2021

RAPPORT DETAILLE


**MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION**
*Liberté
Égalité
Fraternité*

 cybergeonet

Via
TOURISM REVIEW

HYPERGEO


BRUSSELS STUDIES

 CRS

InSHS


GÉOGRAPHIE-CITÉS
UMR 8504

CybergeoNet

Traductions scientifiques

2019-2021

Responsable du projet : Christine Kosmopoulos

[CNRS 8504-UMR Géographie-cités](#)

Participants [Cybergeo](#) : Natacha Aveline, Colette Cauvin-Reymond, Denise Pumain

Partenaires institutionnels et rédacteurs : Université de Paris 1 Sorbonne (Maria Gravari-Barbas), Université de Santiago (Francisco Maturana), Université de Rouen (Bernard Elissalde), Université de Mendoza (Gloria Zamorano), Université Libre de Bruxelles (Benjamin Wayens), Université Saint-Louis – Bruxelles (Margaux Hardy), Union Géographique Internationale (Nathalie Lemarchand).

Revues participantes : [Brussels Studies](#), revue interdisciplinaire spécialisée en études urbaines et régionales, [Via.Tourism Review](#), revue interdisciplinaire spécialisée en tourisme, [Hypergeo](#), Encyclopédie de géographie.

SOMMAIRE

Le projet CybergeoNet.....5

1. Lancement du projet 5
2. Focus sur le fonctionnement linguistique des partenaires..... 6
 - Cybergeo 6
 - ViaTourism Review..... 7
 - Brussels Studies..... 9
 - Hypergeo..... 10

Sur la méthode de traduction des métadonnées au fil de l'eau12

1. La traduction des métadonnées en chinois – Natacha Aveline-Dubach 12
 - Méthodologie..... 12
 - Points sur lesquels la traduction de l'anglais vers le chinois a posé des problèmes..... 13
2. La traduction des métadonnées en anglais – Natacha Aveline-Dubach..... 14
 - Des problèmes structurels 14
 - Des problèmes de formulation 15
 - Le piège des faux-amis et des traductions littérales 15
 - Les erreurs dues aux traducteurs automatiques 16
3. Tests comparatifs sur la traduction humaine et automatique des métadonnées en anglais - Colette Cauvin-Reymond..... 16
 - Méthodologie..... 16
4. Tests comparatifs sur la traduction humaine et automatique des métadonnées en espagnol – Francisco Maturana et Gloria Zamorano 17
 - Constat général sur la langue scientifique espagnole 18
 - Francisco Maturana 18
 - Gloria Zamorano 19

Elaboration d'un protocole de travail et d'organisation à l'intention des équipes éditoriales22

1. Index des mots clés et leur correspondance en plusieurs langues..... 22
2. Glossaire de cartographie 23
3. Glossaire tourisme 24
 - Méthodologie de la première étape de traduction en EN-FR-ES-IT 24

Exemples des premiers paragraphes de l'éditorial du dernier No de Via Tourism.....	25
4. Protocole de travail et d'organisation à l'intention des équipes éditoriales.....	27
Elaboration d'un guide de bonnes pratiques de rédaction et traduction des métadonnées	28
Tests des logiciels de traduction automatique gratuits	30
1. Tests en anglais et espagnol sur l'article "Mesurer l'évolution de l'accessibilité à l'aide des "espaces-temps d'action" : l'exemple de Bogotá entre 1993 et 2009"– Nathalie Lemarchand	30
Méthodologie.....	31
2. Sur le chapitre d'ouvrage : Cartographie thématique, Volume 5, Le son – Colette Cauvin-Reymond.....	31
3. Tests comparatifs de traduction humaine et automatique en espagnol sur 3 articles de Cybergeo – Francisco Maturana	32
4. Tests dans les articles de l'encyclopédie <i>Hypergeo</i> – Bernard Elissalde.....	33
Correspondance et équivalence	33
Caractéristiques du discours et traduction.....	34
Traduction Automatique et modification de sens.....	35
Traduction et tournure grammaticale	35
Capacité à prendre en compte des néologismes et des nouveaux concepts.....	36
Traductions et Figures de rhétorique	38
Auteur(e)s et écritures.....	45
Les expressions idiomatiques	47
Emplois de termes courants avec un sens scientifique particulier	47
Omissions et incomplétude	47
Problèmes linguistiques plus généraux de la TA	47
5. Conclusions.....	48
Impact des opérations de traduction pour Cybergeo - Colette Cauvin-Reymond	49
1. Au sujet de la fréquentation globale par mois et année	49
Fréquentation globale par année entre 2016 et 2020	49
Fréquentation annuelle et mensuelle entre 2016 et 2020	50
Un complément : l'analyse des durées de consultation.....	51

2. Composante géographique de la fréquentation.....	52
Nombre de pages consultées par pays et par année : une typologie nette des entités spatiales sur les 5 années	53
Une classification appuyée sur les rythmes annuels et mensuels (analyse du total des pages consultées par pays, par année et par mois).	55
3. Conclusion	60
Impact des traductions pour ViaTourism	62
Conclusion Générale.....	63

Le projet Cybergeonet

Le projet est porté par *Cybergeonet : Revue européenne de Géographie* en partenariat avec deux revues internationales *Brussels Studies* et *ViaTourism* et l'encyclopédie numérique de géographie *Hypergeo*. Le projet se concentre sur cinq grands objectifs en conformité avec l'appel à projets 2019 du MESRI sur les "traductions scientifiques" :

- 1) améliorer la qualité des traductions des métadonnées des articles déjà publiés dans *Cybergeonet* en faisant vérifier et corriger les traductions existantes par des spécialistes et un recours ponctuel à des traducteurs professionnels ;
- 2) évaluer l'apport des méthodes semi-automatiques pour la traduction scientifique ;
- 3) développer à partir de cette expérience une méthode de traduction innovante, efficace, collaborative et transférable à d'autres revues ayant un modèle économique similaire (autonome, en accès ouvert et avec petit budget) ;
- 4) garantir une montée en compétence des auteurs et des équipes éditoriales en matière de traduction et de rédaction des métadonnées ;
- 5) tester l'impact des améliorations sur la visibilité internationale de la revue au moyen des statistiques de consultation d'OpenEdition.

En partant du vocabulaire spécialisé de la géographie, le projet vise à mettre au point un protocole adaptable à d'autres disciplines des sciences humaines et sociales. Il s'agit de systématiser des actions déjà expérimentées en termes de traduction, et de tester des avancées méthodologiques qui seront partagées avec d'autres revues. En effet les revues à petits budgets en accès ouvert doivent pouvoir accéder *via* la traduction à une meilleure visibilité internationale de leurs publications, mais ce modèle éditorial reste tout de même à la fois viable et fragile.

1. Lancement du projet

Un premier séminaire d'organisation du projet réunissant les revues participantes, l'encyclopédie et les traducteurs s'est tenu le 11 décembre 2019. Il a été l'occasion de faire le bilan des pratiques existantes des différents partenaires, du choix des langues, d'identification des groupes de travail et sous-groupes linguistiques (anglais, espagnol et chinois), incluant des chercheurs « native » souvent bénévoles qui assurent habituellement les traductions pour *Cybergeonet* et ceux des autres revues partenaires, ainsi que des membres des équipes éditoriales. Chaque partenaire confie au groupe de travail et sous-groupe une demande concernant un de ses problèmes de traduction (par exemple un corpus de concepts dans des articles très spécialisés, ou des métadonnées de plusieurs articles). Des réunions mensuelles des groupes de travail à partir de fin janvier 2020 sont destinées en premier lieu à définir les modalités de comparaison des traductions et d'amélioration des métadonnées, à tester la traduction automatique, et dans un second temps à mettre à plat les avancées et les difficultés rencontrées.

2. Focus sur le fonctionnement linguistique des partenaires

Cybergeo

Lors de sa création en 1996, *Cybergeo, Revue européenne de géographie* s'est d'emblée dotée d'une interface multilingue et publie depuis lors dans plusieurs langues européennes. La revue s'est donné une ambition de forte visibilité internationale, sa création ayant été souhaitée par le réseau informel qui assure, tous les deux ans depuis 1978, la tenue des colloques européens de Géographie Théorique et quantitative (150 à 200 universitaires participants). Aujourd'hui, la revue reçoit des propositions d'articles venues du monde entier et publie une cinquantaine d'articles par an (dont environ un tiers par des auteurs étrangers).

Spécialisée en géographie, *Cybergeo* couvre toutes les branches de cette discipline dont le champ lexical est assez spécifique parmi les sciences humaines et sociales : en plus des concepts élaborés par les sciences de la société, il comprend les vocabulaires de la géographie physique, de l'environnement et de l'aménagement, des méthodes statistiques particulières (autocorrélation, fractales, analyse spatiale), des sources très variées comme les images satellites, les données massives géoréférencées, des instruments propres (cartographie, systèmes d'information géographiques), des rubriques destinées à promouvoir une science reproductible (*data papers* et *model papers*). Plusieurs dictionnaires font état de cette richesse et parfois d'une certaine variabilité des définitions.

Cette spécificité au sein des sciences humaines et sociales rend la question de la traduction particulièrement complexe et coûteuse pour une revue spécialisée en géographie. Forte de son ambition internationale, la revue a néanmoins tenté depuis plusieurs années de relever ce défi. Elle a développé pragmatiquement et en interne une méthode pour assurer ses besoins de traduction, n'ayant pas les moyens suffisants pour rémunérer des traducteurs professionnels. Gérée avec les moyens du laboratoire Géographie-cités, la revue fonctionne aujourd'hui avec une personne à temps plein (ingénieure de recherche qui cumule les deux fonctions de rédactrice en chef et secrétaire de rédaction rémunérée par le CNRS) et une équipe éditoriale dont les membres bénévoles sont issus du laboratoire et d'un comité de rédaction composé de scientifiques d'autres universités et institutions de recherche ([Qui sommes-nous ?](#)).

Voici un bref rappel des différentes étapes ayant conduit au fonctionnement actuel pour la traduction :

- Plurilingue dès l'origine en 1996, avec une interface en français et en anglais, titres, résumés et mots-clés en français et en anglais et dans une autre langue européenne de publication de l'article (espagnol, italien, allemand, grec, hongrois, suédois...).
- Traduction en anglais (depuis l'origine) et en espagnol (depuis 2014) de tous les éditoriaux.

- En rejoignant le portail de Revues.org en 2007, *Cyberge* a apporté son modèle plurilingue pour Lodel développé par OpenEdition.
- En 2014, une interface en espagnol est ajoutée au français et à l'anglais.
- En 2018, la traduction en chinois des métadonnées (titres et résumés des articles récents) est lancée.
- Depuis 2011, un programme de traduction en anglais et espagnol d'articles sélectionnés (29 articles traduits en 7 ans) a pu être financé grâce à l'adhésion au modèle libre Freemium. Les traductions ont été assurées par des chercheurs bénévoles et des vacataires.
- Devant l'insuffisante qualité des résumés et des mots-clés en langue étrangère produits par les auteurs pour leurs articles, la revue a entamé en 2016 un programme rétrospectif de vérification et de correction des résumés et des mots-clés en anglais pour ses quelque 880 articles publiés.

Tous ces investissements dans l'internationalisation ont été rendus possibles en grande partie par l'implication bénévole et par les financements perçus avec Freemium depuis 2011 (soit environ 6000 euros/an). La revue percevant en outre une subvention du CNRS à hauteur de 3000 euros/an pour ses frais de fonctionnement, le budget actuel ne permet pas de rémunérer des professionnels de la traduction.

ViaTourism Review

*ViaTourism Review - Revue internationale interdisciplinaire du tourisme*¹, portée par plusieurs universités, sort 2 numéros/an. C'est une revue internationale, plurilingue et interdisciplinaire sur le tourisme, pensée comme un objet interculturel. La revue propose des dossiers thématiques, des varia, des brèves, des positions de thèses, des commentaires photographiques et cartographiques, des recensions d'ouvrages. Chaque article est publié dans au moins trois langues. Elle est en accès ouvert sous le modèle OpenEdition Freemium for Journals. Elle publie en général entre 5 et 8 articles + brèves, photos par numéro... La politique de la revue est de permettre aux auteurs de s'exprimer dans leur propre langue (la revue publie à ce jour en 7 langues²) et de leur proposer une traduction assurée par la revue dans deux autres langues. L'éditorial est traduit dans l'ensemble des 7 langues de la revue. Ce fonctionnement ne pourrait sans doute plus être assuré si la revue décidait de publier un plus grand nombre de propositions d'articles, alors que nous avons actuellement un taux d'acceptation d'environ 30%.

Ce sont les collègues bénévoles des universités partenaires qui traduisent eux-mêmes, par ex. de l'anglais vers le français ou vers l'italien. Les traductions se font toujours vers la langue maternelle des collègues. Les traductions sont ensuite envoyées aux auteurs pour vérification et validation. La revue travaille aussi avec le bureau des traducteurs de l'université de Brest

¹ <https://journals.openedition.org/viatourism/index.html>

² Exceptionnellement, le No 16/2019 a publié un article en arabe et un en chinois.

qui assure des services très personnalisés mais dont les coûts ont beaucoup augmenté³. C'est un problème puisque rien ne garantit que la revue, à petit budget, puisse continuer à travailler avec eux.

La revue teste donc d'autres traductions professionnelles, mais elle est surtout très intéressée par la traduction automatique (TA) de DeepL en complément de la traduction humaine. Le modèle éditorial de *Viatourism* fonctionne bien, mais il reste très chronophage puisqu'il rallonge les délais de publication parfois de 2 mois en raison des traductions. Il faudrait donc voir si avec la traduction automatique les délais pourraient être raccourcis. Les collègues qui traduisent pour *Viatourism* ont une expérience des questions de traduction et sont intéressés pour mettre en place un glossaire sur le tourisme qui pourrait être partagé dans le cadre de *CybergeoNet*, en s'appliquant notamment à travailler sur leur traduction de façon plus réflexive, en notant les éventuelles difficultés, problèmes etc. Un [protocole de travail](#) a été mis en place et testé depuis juillet 2020 dans le cadre de *CybergeoNet*.

Concernant le problème des prix de traduction et des délais rallongés, le groupe se pose la question de savoir si *ViaTourism* n'aurait pas intérêt à faire traduire des résumés longs plutôt que des traductions des articles intégraux. Maria Gravari-Barbas précise que cette option a été débattue, mais que le choix a été fait de la traduction du texte intégral, avec un modèle très bien défini rendu possible du fait de la thématique spécialisée de la revue et d'un nombre d'articles moins élevé qu'une revue généraliste comme *Cybergeo* par exemple. Ainsi, un article soumis en portugais est d'abord évalué dans cette langue. Ensuite, une fois accepté, il y a deux traductions publiées (dont toujours une en anglais) à partir de l'article original en portugais. Il y a une vraie valeur ajoutée pour les auteurs.

OpenEdition, qui héberge la revue, a configuré une trame de page Internet disponible dans les sept langues. Par contre, les métadonnées ne sont traduites que dans les deux langues de traduction de l'article (donc en tout dans trois langues). Il y a des articles dont les métadonnées ne sont pas traduites en français. A titre d'exemple, un article en anglais pourrait être traduit en espagnol et portugais mais pas forcément en français. La version originale de la plupart des articles est le français ou l'anglais (secondairement l'espagnol, l'allemand, l'italien et le portugais – aucun article original n'a été proposé en catalan). Les articles sont ensuite traduits dans deux autres langues, plus rarement trois.

Jusqu'en septembre 2020, les pays dont la revue *Via* a reçu le plus de visites sont l'Allemagne, les États-Unis, la France, le Mexique, la Colombie, l'Inde, l'Italie, l'Espagne et le Brésil, pour un nombre total de pages vues de 774 404. Ces statistiques sont toutefois à examiner avec prudence. C'est sans doute la localisation des moteurs de recherche qui explique un nombre anormalement important de consultations depuis l'Allemagne.

³14 centimes/mots + frais additionnels pour les textes considérés compliqués.

Brussels Studies

*Brussels Studies*⁴ est une revue scientifique d'études urbaines en accès libre et de large diffusion destinée à la mise en valeur de recherches originales portant sur des questions bruxelloises, et ce sans discrimination de discipline ou de langue.

Elle publie en français, en néerlandais (les deux principales langues nationales en Belgique) et en anglais (considéré comme la *lingua franca* de la recherche). Cela signifie qu'elle permet la soumission d'articles écrits dans ces trois langues, mais aussi qu'elle effectue la traduction systématique des articles qu'elle accepte, après une procédure stricte de sélection, pour que tous soient disponibles dans ces trois langues. Son objectif est ainsi d'assurer une diffusion maximale des recherches qu'elle publie. Ce trilinguisme favorise l'appréhension des études sur Bruxelles par différentes communautés linguistiques, à Bruxelles, en Belgique et à l'étranger.

La revue encourage ses auteur·e·s à écrire dans leur langue maternelle (s'il s'agit d'une des trois langues de publication) pour permettre une plus grande finesse des propos. En outre, l'expérience a appris à l'équipe éditoriale que les textes écrits en anglais par des auteur·e·s non anglophones nécessitent généralement une relecture d'un·e traducteur·trice professionnel·le, ce qui engendre des coûts supplémentaires.

Toutes les traductions de *Brussels Studies* sont effectuées par des traducteur·trice·s professionnel·le·s. Ceux-ci·celles-ci signent leurs textes. De cette façon, ils·elles sont identifié/e·s et reconnu·e·s comme des auteur·e·s. Grâce à leur collaboration récurrente avec la revue, les traducteur·trice·s ont une longue expérience de traduction de textes provenant d'un large spectre de disciplines. Par ailleurs, afin de faciliter le dialogue entre disciplines et garantir un accès à ses articles à un public non spécialisé, *Brussels Studies* demande à ses auteur·e·s d'éviter tout jargon et technicisme inutile. L'équipe éditoriale accorde une grande attention à la clarté et à la lisibilité de ces textes, ce qui facilite probablement, en partie, le travail de traduction.

Finalement, les épreuves des articles traduits sont envoyées aux auteur·e·s dans chaque langue pour relecture, en particulier des termes techniques, dans la mesure de leurs compétences linguistiques. En cas d'impossibilité pour les auteur·e·s, le comité éditorial de *Brussels Studies* procède lui-même à la vérification de traduction, en fonction de ses propres connaissances linguistiques. Au sein de l'équipe, constituée actuellement de cinq francophones et d'un néerlandophone, la connaissance active (ou au moins passive) des deux autres langues est en effet généralisée.

⁴ <https://journals.openedition.org/brussels/>

Hypergeo

*Hypergeo*⁵ est une encyclopédie électronique créée en 1999 (remaniée en 2004), proposée en accès libre (Opensource), et consacrée à l'épistémologie de la géographie francophone. Elle héberge 700 textes, tous originaux, rédigés par 120 contributeurs universitaires.

En tant qu'encyclopédie en ligne, elle a pour objectif de faire connaître à un public d'enseignants et d'étudiants les principaux concepts et théories de la géographie contemporaine. Cette visée pédagogique d'*Hypergeo* a induit une organisation du sommaire clairement différenciée. Les articles présentés sont répartis de la façon suivante : un champ général « Géographie », est entouré de trois paradigmes principaux : spatialité des sociétés, villes, territoires et régions, et relations sociétés/environnement.

Afin de se doter d'une forte visibilité internationale, l'encyclopédie est proposée en trois langues (français, anglais, espagnol) dont les textes dans les deux langues étrangères sont des traductions des textes français et non des textes originaux. Au printemps 2020, il existe 320 textes en français, 220 en espagnol et 160 en anglais. Au-delà des questions techniques soulevées par la cohésion d'un ensemble de textes et documents dans trois langues différentes, un certain nombre de problèmes de traduction ont été identifiés depuis longtemps et qui justifient la participation à un programme de recherche. Citons notamment : l'existence de champs sémantiques différents pour une notion en passant d'une langue à une autre ; la signification d'un même terme pour des écoles de pensée nationales différentes, ou encore l'antériorité d'une thématique de recherche dans une langue par rapport à une autre. Pour rendre compte de la spécificité conceptuelle d'un auteur, de sa continuité sémantique et/ou de son évolution lexicale, les articles (et les traductions) sont publiés avec la date de rédaction.

Les textes sont d'un gabarit intermédiaire (environ 9000 à 10 000 caractères) entre la définition d'un dictionnaire et un article de revue. Ils se présentent selon une organisation interne volontairement standardisée. Ils s'ouvrent par un paragraphe contenant une définition générale et les différentes acceptions de la notion ou du concept. Elle est suivie par une présentation de sa généalogie et des évolutions du sens, auxquelles succède un panorama des champs de recherche et des domaines thématiques couverts par la notion. Ils se terminent par une présentation des débats et des enjeux scientifiques contemporains développés autour de la notion ou du concept.

La structure organisationnelle du site (gestion par l'administrateur et interface des fonctionnalités proposées à l'utilisateur) a été élaborée à partir du logiciel libre *SPIP* (système de publication pour l'Internet). La base de données et l'assemblage des textes reposent sur une architecture de type hypertexte. La navigation hypertextuelle est assurée par des liens repérables dans chaque texte par des caractères en gras et une colonne en marge, permettant

⁵ <https://hypergeo.eu/>

de relier cet article à d'autres se rapportant à la même thématique. *Hypergeo* est une réalisation du GDR Libergeo du CNRS et est hébergé par les serveurs de l'UMR Géographie-cités

Sur la méthode de traduction des métadonnées au fil de l'eau

Contributeurs : Natacha Aveline-Dubach (anglais et chinois), Colette Cauvin-Reymond (anglais), Christine Kosmopoulos, Francisco Maturana (espagnol), Denise Pumain, Liubing Xie (chinois), Gloria Zamorano (espagnol)

Les tests de traduction des métadonnées ont été réalisés en anglais, en espagnol mais aussi en chinois. Les tests avec l'anglais comparent la traduction humaine et la traduction automatique des métadonnées. L'outil de traduction automatique utilisé est le logiciel partiellement libre DeepL, mais dans certains cas Google Traduction s'est avéré un instrument également utile et complémentaire.

1. La traduction des métadonnées en chinois – Natacha Aveline-Dubach

Méthodologie

A partir de 2018, la revue *Cybergeo* a pris la décision de traduire ses métadonnées en chinois, initiative pionnière dans le champ de la géographie. Un jeune chinois non francophone spécialiste de langue et littérature anglaise, Liubing Xie, a été sollicité pour faire les traductions à partir de la version anglaise des abstracts. Il a été convenu que je relirai son travail pour garantir une restitution fidèle au champ scientifique. Je précise que je n'ai qu'un niveau intermédiaire en chinois, mais la lecture ne me pose pas de grande difficulté. En outre j'ai pu m'appuyer sur Google traduction, dont la restitution du chinois vers le français est assez satisfaisante du fait de la proximité de la structure syntaxique du chinois de celle du français ou de l'anglais —contrairement au cas du japonais, très mal traduit il y a quelques années, mais avec des progrès spectaculaires observés récemment. La traduction des abstracts par Liubing Xie m'a paru dans l'ensemble de très bon niveau — pour autant que je puisse en juger—, mais quelques difficultés se sont présentées. Une partie des problèmes provenaient du texte d'origine en anglais, souvent trop vague ou jargonneux, qui donnait lieu à confusion. Par exemple, Liubing Xie m'a demandé si "impensé socio-écologique" pouvait être traduit par "socio-ecological thought" (il n'avait pas osé proposer socio-ecological unthought !). Autre exemple : l'expression "compliance level" dans la phrase "the database allows us to capture the housing market visible by this way, at the scale of the metropolitan area and at two compliance levels: the level of the census units on the one hand, the level of the types of settlement on the other" a été traduit en chinois par "niveau de conformité", alors qu'il s'agissait plutôt d'un niveau d'observation. Ces glissements de sens m'ont donc amenée à intervenir sur les abstracts en anglais. Je me suis alors aperçue que Liubing Xie avait restructuré plusieurs phrases mal énoncées dans le texte d'origine au cours du processus de traduction, et je me suis appuyée dans certains cas sur sa traduction en chinois pour modifier l'abstract en anglais ! Liubing Xie m'a d'ailleurs confié qu'il recourait parfois au texte français

en s'aidant de Google Traduction, pour tenter de comprendre ce que l'auteur avait voulu dire. Cependant les textes en français posaient également problème, de sorte qu'il a fallu remonter à ce niveau pour contrôler la chaîne de traduction.

Nous avons alors procédé à l'inverse de la démarche initiale. Les abstracts en français et en anglais (version de l'auteur) m'étaient d'abord envoyés pour vérification, puis je les retravaillais au besoin en m'aidant de DeepL⁶ pour l'anglais, sachant que nous disposions d'une ultime relecture par un anglophone natif. La version anglaise de l'abstract était alors transmise à Liubing Xie.

Points sur lesquels la traduction de l'anglais vers le chinois a posé des problèmes

Je vais me focaliser ici sur les principaux points sur lesquels la traduction de l'anglais vers le chinois a échoué. Les problèmes ont surtout porté sur la traduction de certains termes.

Le mot « quantitative » dans le cas de « quantitative map ». Liubing avait traduit 量化, le premier caractère signifiant « quantité », le second introduisant l'idée d'un changement, d'un processus (comme dans « quantitative easing »). Nous avons finalement opté pour 定量, avec toujours le même caractère de quantité en second, et un premier caractère signifiant « fixe » mais aussi « déterminer, déterminant ». Le mot 定量 est utilisé pour des expressions comme « quantitative forecasting » ou « quantitative analysis ».

Le terme « hybridation » dans le titre « Hybridization between cartographic knowledge and vernacular knowledge » avait été traduit par le mot 融合 qui a plutôt le sens de « fusion » (de 融 « mélanger, fondre » et 合 « combiner »). Un mot alternatif aurait pu être 混合 (« mixte, composite »), avec le caractère 混 désignant également « mélanger », mais avec une connotation un peu négative (idée de confusion). Nous avons finalement retenu d'associer le caractère 整 (qui comprend entre autres le sens de « entier, complet » avec celui de 合, cette combinaison rendant l'idée d'hybridation sous la forme plus neutre d'« intégration, unification »).

Le terme « pillage » 掠夺 (掠: prendre par la force, voler ; 夺 : stratégie, plan) a été utilisé pour traduire « grabbing » dans l'expression « water grabbing » désignant l'accaparement d'une portion de littoral/bord de rivière par un agent économique . Cela me paraissait un peu excessif et j'ai suggéré à Liubing d'explorer des mots avec le caractère 抓 qui signifie « prendre, attraper, capturer, contrôler », mais aucun verbe n'a été trouvé. Il existait cependant un caractère plus compliqué tout à fait adéquat : 攫 qui recouvre bien le sens de « accaparement, emparement », plus encore s'il est associé au caractère « prendre » (取). D'où notre choix du mot 攫取 pour « water grabbing ».

⁶ Je procédais initialement à une traduction automatique avec DeepL du français vers l'anglais. Il n'y avait pas encore de passage de DeepL vers le chinois.

L'expression "short circuits" dans la phrase "establishing short circuits for financing local public action constitutes a credible alternative to the use of the Agri-environment-climate Measures" a été traduite par 快捷途径 qui signifie « moyens rapides ». La traduction finale a été : 小规模资金链 (chaîne de capitaux à petite échelle).

Pour le mot « laboratoire » (au sens d'UMR en France) nous avons le même problème qu'avec la traduction en anglais, qui réduit le terme à un espace d'expérimentation (c'est le sens du mot chinois initialement utilisé 实验室, qui signifie « salle d'expérimentation »). En anglais on traduit plutôt par « research unit » et il est possible de le traduire ainsi en chinois en associant les caractères « recherche » et « unité » (研究单元). Cependant, le terme 研究机构 (littéralement « organisation de recherche ») est plus approprié car il désigne également la structure interne d'une organisation.

Par ailleurs, comme on s'en doute, les notions se référant aux découpages territoriaux français (comme « département » ou « région ») représentent un défi pour la traduction en chinois. Des mots comme « diagonale vide » ont dû être traduits de façon littérale (空对角线: vide-diagonale) sans plus d'explication. Cependant ces problèmes ne sont pas spécifiques à la traduction en chinois.

Pour finir, il faut noter que Cybergeo reçoit davantage de soumissions de chercheurs chinois depuis que ses abstracts sont traduits dans leur langue (s'ajoutant au fait que la revue soit indexée SSCI). Nous devons donc poursuivre notre effort !

2. La traduction des métadonnées en anglais – Natacha Aveline-Dubach

La traduction de ces métadonnées du français vers l'anglais a révélé quatre grandes catégories de problèmes que je décris ci-dessous.

Des problèmes structurels

Les problèmes structurels peuvent concerner la taille des abstracts (souvent trop longs), la faible lisibilité de la problématique et des résultats, le caractère peu explicite de la méthode, et/ou un choix peu pertinent de mots-clés. Ils renvoient plus généralement à la non-conformité des métadonnées aux normes des revues scientifiques (cf. [le guide de bonnes pratiques de rédaction des métadonnées](#)). Dans de tels cas, il a été nécessaire de retravailler en profondeur la version française du résumé et parfois les mots-clés. J'ai ensuite soumis cette version à la traduction automatique en anglais par DeepL, puis procédé à d'éventuels ajustements. Pour finir, notre lecteur anglophone a effectué une dernière vérification avant la mise en ligne des métadonnées. A noter que cette ultime relecture n'a pas présenté de valeur ajoutée significative à la traduction de DeepL sous notre contrôle, ce qui atteste du haut niveau atteint aujourd'hui par ces traducteurs automatiques.

Des problèmes de formulation

Les résumés en anglais présentent les intentions et résultats de la recherche de façon très directe (« we examine », « this work highlights », « we provide evidence of » etc.) tandis que l'approche française est plus indirecte (“we aim to examine”, “though this work we would like to highlight”, “it seems that”...). Bien que les formulations françaises, plus humbles, soient tout à fait correctes, elles sont peu compatibles avec l'économie de mots exigées par le format des abstracts et peuvent donner une impression d'hésitation à un lectorat acculturé aux standards internationaux. J'ai donc dû procéder —à mon corps défendant— à la mise aux normes des abstracts aux standards anglo-saxons sur ce plan-là.

Par ailleurs, le français fait un usage abondant du mode passif, lequel passe assez mal en anglais en raison des impératifs évoqués ci-dessus. Par exemple : “In this paper, XX is presented”=> “this paper presents XX” ou bien “an analysis of newspaper articles is carried out on the basis of a corpus of 180 texts”=> « we carried an analysis of newspaper articles out on the basis of a corpus of 180 texts”.

Enfin, le français utilise le futur dans des cas où l'anglais reste au présent. Par exemple, la formulation “the geographical database compiled in this work is an extremely useful tool for understanding and monitoring wildfires and *will provide* managers and decision-makers with a valuable basis for decision-making” paraît plus naturelle en anglais sous la forme : “the geographical database compiled in this work is an extremely useful tool for understanding and monitoring wildfires and *provides* managers and decision-makers with a valuable basis for decision-making”. Il convient de préciser que les traducteurs automatiques n'interviennent pas sur ces différences de formulation.

Le piège des faux-amis et des traductions littérales

En raison des racines communes anciennes de l'anglais et du français, des erreurs ont également porté sur la traduction de certains mots. On en distingue deux principaux types :

- l'usage de faux-amis : par exemple, le terme « qualify » utilisé pour “qualifier”, au lieu de « characterize ».

- la traduction littérale de mots ou d'expressions qui sont compréhensibles en anglais mais guère pratiqués dans cette langue. Ces erreurs sont beaucoup plus fréquentes. C'est le cas par exemple du terme « articulation », souvent utilisé par les auteurs français pour mettre en lien les questionnements ou les champs de connaissances, comme dans la phrase « articulation entre pluridisciplinarité et savoirs pratiques ». Dans les métadonnées en anglais, on utilise plutôt les mots « link », « relationship ». De même, si l'expression « mutations urbaines » peut être traduite littéralement en anglais, on lui préfère généralement les formulations de « urban transformations », « urban change » ou « urban restructuring » selon la nature des mutations. Le mot « appropriation » (d'une idée, d'un concept, d'un objet, d'un bien...) se retrouve également souvent inchangé dans les traductions, alors qu'il est nécessaire de rechercher le

terme anglais le plus adéquat en fonction du contexte (to adapt, to embrace, to take advantage, to take up, to take control, to take over, to take ownership). Les erreurs de ce type sont rarement corrigées par les traducteurs automatiques. Il convient donc d'être vigilant lors de la phase de relecture finale.

Les erreurs dues aux traducteurs automatiques

Les traducteurs automatiques génèrent également des erreurs systématiques. J'en ai croisé en particulier deux types dans la traduction des métadonnées :

-modification automatique des acronymes par renversement des lettres, ce qui est particulièrement problématique quand les acronymes sont issus d'un mot anglais.

-assimilation de mots d'origine étrangère à des termes proches en français. Par exemple, le terme algérien « jessour » (ouvrages de conservation des eaux et des sols qui barrent les ravines creusées dans les loess) a été automatiquement traduit par « jour » par Google traduction.

Cependant, il semble que ces risques d'erreurs aient été très fortement réduits récemment. En testant les performances actuelles de DeepL et Google traduction, je n'ai plus rencontré d'erreur de ce type. Ces logiciels ne transforment que les acronymes les plus courants (exemple : OMC est traduit par WTO).

En conclusion, la grande qualité atteinte par les logiciels de traduction automatique permet de restituer les métadonnées en français dans un anglais tout à fait satisfaisant, à condition de procéder à une relecture humaine avant la mise en ligne. L'effort doit donc porter sur la qualité de l'abstract en français afin d'éviter les problèmes structurels, les erreurs de formulation et l'usage de faux-amis. Le guide que nous avons préparé à l'attention des auteurs devrait permettre de minimiser notre effort.

3. Tests comparatifs sur la traduction humaine et automatique des métadonnées en anglais - Colette Cauvin-Reymond

Methodologie

C. Cauvin est partie du résumé en comparant l'outil de traduction automatique DeepL à la traduction faite lors de la publication de l'article par une traductrice. En partant de la traduction anglaise, le résumé a été retraduit en français avec DeepL, puis dans les deux sens à partir la version française initiale avec DeepL.

Français initial > Anglais traductrice humaine > Français DeepL

Français initial > Anglais DeepL > Français DeepL

Nous mettons en [annexe](#) les résultats du test comparatif suivant :

"Proposition pour un archétype urbain logico-géographique : l'association « logique trichotomique - logique ternaire antagoniste » et son intérêt géographique"⁷

La différence flagrante porte essentiellement sur le fait que la traduction automatique (TA) est assez littérale. En effet, la traduction de DeepL rend assez bien le contenu du texte, mais ne modifie pas la structure des phrases. La traductrice quant à elle est passée au niveau supérieur avec l'accord des auteurs et a utilisé des tournures plus anglaises. Le retour au français à partir de la traduction anglaise montre d'ailleurs un résultat plus clair voire plus facile à lire que le texte original en français.

Une difficulté est cependant apparue au sujet du terme "power".

Les deux traductions provoquaient, lors du retour en français, la même insatisfaction à propos du terme "power", utilisé par les deux types de traduction en anglais pour exprimer le mot "pouvoir". Power lors du retour en français devient puissance. Puissance et pouvoir ne sont pas équivalents en français, en particulier dans le texte concerné où « pouvoirs » était au pluriel. *Conclusion* : un retour au français après la traduction serait intéressant pour éviter des erreurs involontaires. Le détail est présenté en annexe 1.

"Mesurer l'évolution de l'accessibilité à l'aide des "espaces-temps d'action" : l'exemple de Bogotá entre 1993 et 2009"

Un essai similaire a été mené sur "Mesurer l'évolution de l'accessibilité avec les "espaces-temps d'action"⁸. Les conclusions sont identiques à celles sur l'article "Logique et géographie" présentées ci-dessus, mais il faut souligner la difficulté de traduction, en particulier dans le titre, pour les termes nouveaux, comme ici "espaces-temps d'action". Cette expression a été traduite de la même manière dans le titre et le texte lors de la traduction "humaine" alors qu'avec DeepL, la traduction du titre était inexacte, mais dans le texte, la TA diffère du titre et traduit de manière correcte et identique à celle de la traduction humaine (cf annexe 2).

4. Tests comparatifs sur la traduction humaine et automatique des métadonnées en espagnol – Francisco Maturana et Gloria Zamorano

Concernant l'harmonisation des mots clés en espagnol, les articles ont été mis à jour dans *Cybergeo*. Les mots clés nouvellement traduits en espagnol sont entrés dans le fichier Excel qui recense les mots-clés de la revue avec leur correspondance dans les différentes langues (cf [Index des mots clés et leur correspondance en plusieurs langues](#)). Nous partageons le

⁷ Henri Reymond et Colette Cauvin-Reymond, « Proposition pour un archétype urbain logico-géographique : l'association « logique trichotomique -logique ternaire antagoniste » et son intérêt géographique », *Cybergeo : European Journal of Geography* [En ligne], Systèmes, Modélisation, Géostatistiques, No.860, mis en ligne le 11 juillet 2018, consulté le 24 février 2020. URL : <http://journals.openedition.org/Cybergeo/29177> ; DOI : <https://doi.org/10.4000/Cybergeo.29177>

⁸ Florent Demoraes, Vincent Gouëset et Mégane Bouquet, « Mesurer l'évolution de l'accessibilité à l'aide des "espaces-temps d'action" : l'exemple de Bogotá entre 1993 et 2009 », *Cybergeo : European Journal of Geography* [En ligne], Cartographie, Imagerie, SIG, document 926, mis en ligne le 17 décembre 2019, consulté le 11 mai 2020. URL : <http://journals.openedition.org/cybergeo/33638> ; DOI : <https://doi.org/10.4000/cybergeo.33638>

fichier avec les partenaires du projet (*ViaTourism*, *Brussels Studies* et *Hypergeo*) pour qu'ils puissent enrichir cet index multilingue.

Constat général sur la langue scientifique espagnole

Il n'y a pas de grosses différences entre le vocabulaire utilisé en Argentine, au Chili, au Pérou, en Colombie, en Bolivie, au Mexique et en Espagne. L'usage de la langue espagnole est homogène en ce qui concerne le vocabulaire scientifique pour les concepts de géographie dans l'ensemble de l'Amérique latine et avec l'Espagne, même si dans la langue courante on trouve des expressions plus spécifiques à certaines régions. Cela peut s'expliquer par le fait que beaucoup de livres scientifiques qui circulent en Amérique Latine viennent d'Espagne et de nombreux chercheurs en Amérique Latine y ont été formés.

Francisco Maturana

Méthode

Personnellement pour faire les traductions, j'entre le texte en français et je le traduis automatiquement, sur Google traduction, après je le corrige. Il y a toujours des problèmes, peu, mais il y en a. De plus, il est souvent nécessaire de modifier des mots inadéquats ou certaines expressions.

Exemples

Exemple : texte original en français :

« Il s'agit d'identifier les évolutions exceptionnelles qu'ont pu connaître certaines villes depuis 1961, notamment avec la chute du Mur de Berlin en 1989, qui signale un changement de paradigme de développement urbain en Europe centrale et orientale ».

Traduction automatique sur Google traduction :

“El objetivo es identificar los desarrollos excepcionales que ciertas ciudades han conocido desde 1961, en particular con la caída del Muro de Berlín en 1989, lo que indica un cambio en el paradigma del desarrollo urbano en Europa central y oriental”.

Correction

“El objetivo corresponde a identificar la extraordinaria evolución que evidenciaron determinadas ciudades desde 1961, particularmente con la caída del Muro de Berlín en 1989, implicando un cambio de paradigma en el desarrollo urbano de Europa central y oriental”.

On peut constater ici qu'il y a des problèmes de singulier et de pluriel, d'expressions et de sens, entre autres. Le vocabulaire n'est pas non plus très formel ou élégant. Par exemple Google Traduction utilise le mot "excepcional" (traduction textuelle), mais à mon avis ce serait plutôt « extraordinaria ». Un autre mot, par exemple, pour Google « ciertas » est retenu mais à mon avis « determinadas » serait préférable. Un autre exemple l'expression « lo que indica » mais serait mieux rendu par « implicando », qui permet de simplifier la rédaction.

Il est absolument nécessaire de bien relire la traduction automatique (TA) parce qu'il y a des mots erronés et que le style est pratiquement toujours à reprendre. Or, l'exigence scientifique impose une rédaction claire et structurée. On rencontre aussi des problèmes pour les noms de lieux avec lesquels le traducteur en espagnol n'est pas familier. Le choix est fait si possible de les traduire selon leur usage en espagnol, ce qui est utile à la fois pour le lecteur hispanophone, et pour les autres lecteurs qui découvrent comment, par exemple, on écrit une certaine ville en espagnol.

Gloria Zamorano

Pour les mots scientifiques, généralement il n'y a pas beaucoup de différences, mais il arrive qu'on se heurte à des difficultés de choix entre deux traductions pour un même concept.

Exemples

Exemple : « décroissance urbaine »

On peut traduire par « *deterioro* », mais aussi par « *decrecimiento* ». Cette expression en géographie pose problème dans toutes les langues. Dans les années 80 en France, on disait "déclin urbain", puis il y a eu des discussions quant à la nuance péjorative associée à ce terme, au fur et à mesure que le processus se généralisait et se diversifiait et on a préféré utiliser un terme neutre comme "décroissance urbaine". Mais en espagnol on utilise le terme de « *deterioro* », qui a une connotation un peu moins négative que « déclin » en français, et parfois aussi « *decrecimiento* » qui serait l'équivalent neutre de « décroissance ». Derrière l'utilisation de ces concepts se reflète une représentation plus ou moins catastrophiste de l'évolution urbaine.

Il est aussi intéressant de voir comment les processus ont amené l'émergence de mots différents selon les langues, par exemple les concepts en français de "faubourg" et de "banlieue" pour lesquels en espagnol il n'y a qu'un seul mot : *Suburbio*. Pour les Français le "suburbain", ce sont des morceaux de ville qui sont réunis par des flux de travail, en périphérie des villes ; on emploie aussi le terme de "périurbain" ou parfois "rurbain", pour désigner des lieux habités surtout par des employés urbains mais qui ne sont pas construits en continuité avec la ville et qui en dépendent complètement pour l'emploi ou certains services. On n'a pas deux termes distincts comme faubourg et banlieue en espagnol, peut-être parce qu'en Espagne les villes sont restées très longtemps compactes et donc il n'y a pas eu lieu de forger deux concepts distincts. Il reste que traduire faubourg ou banlieue en *suburbio* peut poser un problème, cela ne recouvre pas toujours des formes urbaines similaires.

Autre exemple, la traduction d'une phrase dans le résumé de l'article 651⁹.

⁹ Serge Lhomme, Richard Laganier, Youssef Diab et Damien Serre, "La résilience de la ville de Dublin aux inondations : de la théorie à la pratique", *Cybergeo : European Journal of Geography* [En ligne], Environnement,

J'ai fait la comparaison avec Google traduction et DeepL.

- En français : « La ville de Dublin doit faire face à des inondations d'origines diverses et à un développement urbain continu qui requiert d'aménager certaines zones identifiées comme étant inondables. »

- En espagnol : « La ciudad de Dublin debe hacer frente a inundaciones de orígenes diversos y a un desarrollo urbano continuo que requiere ordenar ciertas zonas identificadas como inundables. »

Dans ce cas, comme indiqué dans l'annexe 3, dans la première partie de la phrase, j'ai fait la même traduction proposée par Google, en exprimant en espagnol « hacer frente a inundaciones de orígenes diversos », car il s'agit des mots spécifiques de la géomorphologie. En DeepL on préférerait parler de « diversas fuentes », par exemple. Quant à la deuxième partie de la phrase, « qui requiert d'aménager certaines zones identifiées comme étant inondables », j'ai préféré aussi la traduction de Google, plus courte et directe, correspondant aux expressions linguistiques postmodernes, « ciertas zonas identificadas como inundables », tandis que la traduction de DeepL était : « ciertas áreas identificadas como propensas a inundaciones ».

Commentaires

Toutes ces remarques montrent que la relecture de la traduction par un spécialiste de la langue et du domaine est indispensable. Il arrive que ces traducteurs géographes, en l'occurrence, fassent eux-mêmes appel à des collègues lorsqu'il s'agit de termes très spécifiques (sur les images satellites, la géographie physique, etc.)

On souligne une fois de plus l'importance d'un texte original clair, précis et bien structuré pour permettre une bonne traduction, avant d'entreprendre les traductions de résumés, il est important de noter les choses qui ne sont pas claires dans la langue originale. C'est au moment de la traduction qu'on se rend compte des confusions voire des contradictions du texte original, résumé ou article.

Le sujet de la spécificité du vocabulaire scientifique est évidemment très important, notamment pour les SHS. Par exemple, une série d'articles récents traitent des villes à partir de points de vue de la science physique, et identifient ce qu'ils appellent des « Objets d'intérêt », qui correspondent en fait à des équipements ou des services urbains signalés sur les réseaux sociaux par des internautes pour être rendus visibles aux autres. La captation de ces informations par Internet permet de constituer à bas coûts d'énormes bases de données,

et de tester divers modèles de la présence des équipements et des services en ville. On peut alors retirer l'impression d'une approche totalement nouvelle, d'une forme inédite de « science des villes », alors qu'en fait ces travaux retrouvent, sans le dire et sans utiliser le vocabulaire déjà construit, des formes et des niveaux de « fonctions urbaines » et de « centralité urbaine » qui sont des concepts depuis longtemps théorisés et formalisés par les travaux de géographie. Une « traduction » entre disciplines soulève ainsi parfois des questions du même ordre que celle de la traduction entre langues différentes.

Elaboration d'un protocole de travail et d'organisation à l'intention des équipes éditoriales

Contributeurs : Natacha Aveline-Dubach, Colette Cauvin-Reymond, Maria Gravari-Barbas, Margaux Hardy, Christine Kosmopoulos, Denise Pumain

1. Index des mots clés et leur correspondance en plusieurs langues

Un premier fichier des mots clés a été constitué, entre 1996 et 2020, par la revue *Cybergeo* à partir de la révision complète des 850 articles publiés. L'élaboration de ce fichier a exigé la collaboration de 8 personnes et l'engagement de huit mois de travail à temps plein.

Etape 1

Dès 2017, les listes de mots clés enregistrés dans *Cybergeo*, en anglais et en français, ont été corrigées en éliminant les coquilles, les doublons et les orthographe fautive. Le niveau conceptuel de certains termes ou expressions a été harmonisé par les spécialistes du domaine dans les deux langues. Ces listes validées ont été employées pour réviser les mots clés présents dans chacun des 850 articles.

Par ailleurs, l'indexation géographique était notée de manière très hétérogène dans les mots clés généralistes d'origine. Un index géographique conforme aux normes internationales ISO des Nations Unies a été établi. Il a été utilisé pour compléter systématiquement les mots clés des articles.

Etape 2

Dès le début du projet *CybergeoNet* (décembre 2019), un fichier/index a été créé par Christine Kosmopoulos pour établir les correspondances des mots clés entre les langues de publication des articles dans *Cybergeo* (français, anglais, espagnol, allemand, portugais, italien, suédois, russe). Ce fichier a été utilisé pour compléter les mots clés manquants, notamment en anglais, dans tous les articles de *Cybergeo* dans le courant de l'année 2020. Les correspondances multiples ont été vérifiées par des traducteurs humains et des traductions automatiques.

Etape 3

Ce fichier a été transmis aux partenaires pour être complété et validé. Etant formé par les apports successifs des articles publiés, il ne constitue pas un glossaire de la géographie, (contrairement aux exemples spécialisés, cartographie et tourisme, qui sont joints à ce rapport). Mais les mots choisis par les auteurs pour distinguer leur publication posent souvent des problèmes d'ajustement entre spécificité (ou originalité) et généralité (ou généralité), qui se retrouvent à la traduction. La tendance est à ce que les catégories du vocabulaire soient

d'un niveau plus fin, plus précis ou plus « détaillé » dans la langue initiale et qu'elles soient de niveau plus élevé dans la hiérarchie (approximative) des concepts dans la langue traduite. Un échantillon de ce glossaire plurilingue de Cybergegeo est fourni en annexe 11. Il sera mis intégralement en ligne sur la collection Dataverse Cybergegeo¹⁰.

2. Glossaire de cartographie

Un glossaire avait été construit par les auteurs d'un traité de cartographie thématique publié en français (5 volumes) en 2007-2008 et traduit en anglais (3 volumes) en 2010 (Cauvin, Escobar, et Serradj)¹¹.

Ce glossaire est produit en annexe 4. Il a été constitué selon plusieurs approches :

- Pour les termes de cartographie classique, les deux glossaires du Comité Français de Cartographie ont été utilisés¹², complétés et transformés en fonction de l'évolution de certains termes dans leur signification ou leur utilisation.
- La plupart des termes, surtout récents ou plus techniques, ont été collectés au fur et à mesure des lectures dans des articles ou des ouvrages. Il semble que certains termes soient liés à un auteur ou un autre (d'où l'indication de certains auteurs dans le glossaire).
- Le web a été utilisé prudemment mais il ne faut pas oublier que la traduction a été effectuée entre 2008 et 2009 et qu'à ces dates, le web avait moins de possibilités qu'en 2020. Nous avons toujours vérifié les propositions et nous continuons encore maintenant.
- Des mots ont posé de nombreux problèmes comme *sign* ou *symbol*. En principe les sens des termes sont inversés en anglais. Mais selon qu'ils sont utilisés en référence à F. De Saussures ou à C.S. Peirce, le terme est différent. Nous avons précisé en note¹³ quel avait été notre choix pour éviter toute ambiguïté ultérieure.
- Certains mots sont dangereux : de vrais "faux amis" comme "*contour*" qui n'est pas le contour d'une forme mais qui est utilisé en particulier dans l'expression "*contour map*", carte en courbes de niveau.
- Les choix initiaux ont souvent été effectués après de nombreux échanges entre les 3 auteurs avec la nécessité de l'harmonisation soit vers l'anglais soit vers l'américain (cf. s et z) mais tout n'a pas pu être harmonisé du fait des spécificités au sein même de

¹⁰ <https://dataverse.harvard.edu/dataverse/cybergegeo>

¹¹ Cauvin C., Escobar F., Serradj A., 2007-2008, Cartographie thématique, Hermès Science, Lavoisier, Traité IGAT, Paris, (Vol.1. Une nouvelle démarche. Vol.2. Des transformations incontournables. Vol.3. Méthodes quantitatives et transformations attributaires. Vol.4. Des transformations renouvelées. Vol.5. Des voies nouvelles à explorer. Cauvin C., Escobar F., Serradj A., 2010, Thematic Cartography and Transformations, Volume 1, Cartography and the Impact of the Quantitative Revolution, Volume 2, New Approaches in Thematic Cartography, Volume 3, Iste Ltd, Wiley & sons, Inc., Great Britain.

¹² Comité Français de Cartographie, 1970, "Glossaire français de cartographie", Comité Français de Cartographie, Fascicule No.46, Bulletin No.4, p. 127-260. Comité Français de Cartographie, 1990, "Glossaire de cartographie", Comité Français de Cartographie, Bulletin No.123/124, 171 p.

¹³ Dans l'ouvrage français, note 4, pages 25 à 27 (Vol.2, Ch.1). Dans l'ouvrage anglais, notes 4 et 14, pages 215 à 217 (Vol.1, Ch.7).

notre groupe pour la langue anglaise : anglais pour tous les 3, américain pour CC et australien pour FE et AS.

- Dans le cadre de *Cybergeo*Net le glossaire a été révisé et complété à la marge par l'ensemble des partenaires du projet pour ce qui est du français vers l'anglais.

3. Glossaire tourisme

La création du glossaire sur le tourisme s'est fait sur la base des entrées choisies dans les 7 langues de *Via Tourism Review*¹⁴.

Méthodologie de la première étape de traduction en EN-FR-ES-IT

La traduction des concepts dans les différents articles des numéros de la revue *Via Tourism* a été réalisée en plusieurs étapes. Dans un premier temps, le travail a consisté à repérer un ensemble de concepts dans la langue d'origine des articles pertinents par rapport au projet de *Cybergeo* (mentionnés par « VO » ou « V.O. »). Dans un second temps, chaque concept a été recherché dans les traductions de l'article considéré et selon la disponibilité des traductions (français, anglais, italien, espagnol). Pour de nombreux articles, les 4 traductions ne sont pas disponibles. On retrouve le plus souvent la triangulaire français-anglais-espagnol. Les articles italiens sont moins nombreux et sont souvent associés à l'anglais et/ou au français. Une fois l'ensemble des concepts et leurs traductions relevés, un tri a été effectué afin de ne retenir que les concepts les plus pertinents par rapport au projet *Cybergeo*. Par ailleurs, une colonne « Remarques éventuelles » a été ajoutée afin de noter les occurrences et/ou les multiples traductions pour certains concepts. Du point de vue des résultats, on observe globalement que les articles sont disponibles la plupart du temps en 2 ou 3 langues ce qui explique l'absence de traduction systématique. Dans quelques articles, le concept repéré dans la langue d'origine était traduit sous la forme d'une périphrase dans les autres langues (par exemple « expériences touristiques de l'inauthentique », « quête de l'esthétique touristique » et « quotidienneté urbaine » relevés en français, traduits en espagnol mais pas directement en anglais même si l'article était disponible dans cette langue). Pour plusieurs articles, on note l'abondance des traductions dans une langue. A titre d'exemple, les concepts de « communautés locales », « destinations touristiques », « mise en tourisme » « mise en scène », « lieux touristiques » et « processus de patrimonialisation » sont traduits différemment d'un article à l'autre ou au sein d'un même article. Dans cette perspective, une réflexion semble avoir été menée en amont sur le choix du terme à utiliser au regard du contexte d'écriture. Enfin, certains concepts n'ont pas été traduits dans les autres langues. Il s'agit de concepts spécifiques initialement définis en anglais et utilisés dans les autres langues de la même manière (par exemple « dark tourism », « thanatourism », « lifestyle migration », « privileged migration », « agency » et « city branding »).

En ce qui concerne les concepts repérés, on peut observer qu'ils relèvent souvent des lexiques spécifiques à différentes disciplines, notamment la sociologie (e.g. arène, pratique), la

¹⁴ <https://journals.openedition.org/viatourism/>

géographie (en particulier les termes qui relèvent des notions d'espace et de territoire), l'urbanisme, l'économie, le marketing, les études sur le patrimoine. Cela reflète sans doute la nature transdisciplinaire des études sur le tourisme. La thématique de chaque numéro joue également un rôle. Ainsi, dans le numéro 16 on trouve beaucoup de concepts qui relèvent du champ des études postcoloniales, qui empruntent à leur tour à la théorie du discours et à la sémiotique. On a choisi d'inclure également des concepts qui relèvent de la méthodologie utilisée par les auteurs (e. g. « enquête ethnographique » ou « entretien compréhensif »).

Quelques exemples des premiers paragraphes de l'éditorial du dernier No de Via Tourism

Nous présentons ci-après quelques paragraphes de l'éditorial du No 16 de Via (2019) où nous avons surligné en jaune les mots ou ensembles de mots qui ont été retenus pour le glossaire.

Alors que la France représente dans l'histoire mondiale le deuxième empire colonial et que celui-ci a été un puissant vecteur de la mondialisation touristique (Zytnicki et Kazdaghli, 2009 ; Baranowski et Furlough, 2001), peu de travaux francophones portent encore sur la dimension postcoloniale des pratiques, **lieux et imaginaires du tourisme** contemporain comme sur la production normative qui régit son organisation économique, sociale et politique (Boukhris, Chapuis, 2016).

(...)

Notre discours est également situé depuis la France, qui semble faire figure d'exception dans la production et la réception du **paradigme postcolonial** du fait des difficultés et du silence assourdissant qui caractérisent leurs mobilisations dans les travaux de recherche ou les enseignements. On retrouve quelques tentatives précoces d'introduction de ces approches en géographie (Staszak et al., 2003 ; Collignon, 2007 ; Chivallon, 2007), suivies néanmoins d'oppositions virulentes (Lacoste, 2010). Notre perspective située nous oblige à revenir sur cette réception du paradigme postcolonial en France, invitant les lecteurs internationaux de la revue Via à réaliser ce même exercice d'analyse critique de leurs traditions académiques nationales respectives ou à le réexplorer. Cet examen s'avère être un laboratoire des enjeux de nos sociétés contemporaines et en particulier, de la façon dont celles-ci engagent le dialogue avec leur passé esclavagiste, colonial et/ou impérial.

Cette situation n'a rien de spécifique à la **géographie ou aux études touristiques** : elle tient plus globalement à la place à la fois marginale et polémique qu'occupe le **paradigme postcolonial** en France (Bancel, 2012). Cette « constellation intellectuelle dense et protéiforme » (Mbembe, 2010) s'installe tardivement dans le paysage académique français : alors que la contribution majeure d'Edward Saïd est publiée en 1978, il faut attendre 1980 pour qu'elle soit traduite en français et 2006 pour qu'un premier colloque soit organisé sur la "situation postcoloniale" à l'EHESS (Smouts, 2007), donnant lieu à des critiques sévères (Amselle, 2008 ; Bayart, 2010).

(...)

L'objectif est plutôt de comprendre non seulement l'ensemble des implications de la colonisation dans le fonctionnement touristique des **sociétés colonisées** (passées, présentes et à venir), mais aussi les structures de la pensée coloniale, ce qu'Edward Saïd nomme l'impérialisme, soit la pratique, la théorie et les attitudes d'un centre métropolitain dirigeant un territoire lointain, perdurant bien après le colonialisme. Or, le tourisme a été un puissant vecteur de circulation de ces structures de pensée. En cela, le paradigme postcolonial participe d'une critique non seulement du fonctionnement de nos **sociétés touristiques** contemporaines mais aussi d'une espérance en la réparation, la réciprocité et le partage, à laquelle **l'économie touristique** peut participer (Berger, 2006 ; Mbembe et al., 2006).

Nous avons ainsi retenu comme indiqué ci-dessus des termes qui relèvent du vocabulaire touristique ('lieu touristique', 'imaginaires du tourisme', 'société touristique', 'économie touristique', 'géographie du tourisme', 'études touristiques'...) mais également des mots et associations de mots qui semblent incontournables pour des articles en sciences du tourisme postcoloniales ('sociétés colonisées', 'paradigme postcolonial').

Le glossaire plurilingue sur le tourisme est fourni en annexe 10.

4. Protocole de travail et d'organisation à l'intention des équipes éditoriales

CybergeoNet s'est engagé à fournir une méthode de traduction innovante, efficace, collaborative et transférable à d'autres revues ayant un modèle économique similaire, en accès ouvert avec de petits budgets. Le protocole a été développé incluant une méthode de travail de la traduction des métadonnées au fil de l'eau et d'organisation à destination des équipes éditoriales de revues pour assurer à moindre coût une bonne qualité de traduction.

Le protocole s'appuie sur les pratiques éditoriales mises en place à *Cybergeo*, décrites ci-dessous, sur [l'index plurilingue des mots clés de *Cybergeo*](#) et sur les glossaires spécialisés en [cartographie](#) et sur le [tourisme](#) et fournis en annexes du rapport.

Dans la note aux auteurs publiée sur le site de la revue *Cybergeo*, un [guide de bonnes pratiques sur la rédaction et la traduction des métadonnées](#) a été inséré. Ce guide est adressé par les équipes éditoriales à tous les auteurs dont l'article a été accepté pour publication afin de les aider à mettre en conformité les métadonnées de leur article, la rédaction et la traduction dans les différentes langues demandées. Lorsque l'article corrigé selon les recommandations éditoriales et le [guide de bonnes pratiques sur la rédaction et la traduction des métadonnées](#) est retourné, il est adressé au rédacteur en chef de la revue. La vérification de la qualité de rédaction (dans la langue originale) et de la traduction des métadonnées en anglais, espagnol et selon le cas en chinois est faite par plusieurs experts géographes et un traducteur anglophone.

A ce stade, le pré-print peut-être renvoyé aux auteurs par le service éditorial avec une demande d'amélioration des métadonnées, du titre et/ou résumé et/ou des mots clés. Les auteurs auront fourni, selon le cas, à la demande de la revue des termes spécifiques sur la thématique de leur article avec des propositions de traductions en anglais et en espagnol. Dans le cas de *Cybergeo*, la revue n'ajoute que de manière très exceptionnelle un mot clé. En effet, même si le contenu de la science se renouvelle tout le temps le logiciel de la plateforme d'OpenEdition actuellement en vigueur ne permet pas de modifier la liste de mots-clés en appliquant une mise à jour rétroactive automatique. Seuls des concepts tout à fait nouveaux seront examinés pour intégration. Actuellement l'index plurilingue de *Cybergeo* comporte près de 2200 mots clés décrivant ses articles.

[L'index plurilingue des mots clés de *Cybergeo*](#) et les glossaires spécialisés en [cartographie](#) et sur le [tourisme](#) pourront s'enrichir de la liste de termes spécifiques acceptés fournis par les auteurs et seront partagés entre les équipes éditoriales.

Elaboration d'un guide de bonnes pratiques de rédaction et traduction des métadonnées

Contributeurs : Natacha Aveline-Dubach, Colette Cauvin-Reymond, Margaux Hardy, Christine Kosmopoulos, Denise Pumain

Dans la perspective d'assurer une meilleure visibilité internationale à leurs auteurs, les revues de *CybergeoNet* partagent l'ambition de fournir des métadonnées de qualité, que les auteurs d'articles rédigent la plupart du temps de manière empirique sans consignes particulières.

La première réunion de *CybergeoNet* en décembre 2019 a fait ressortir la nécessité de constituer un groupe de travail qui réfléchisse à des recommandations portant sur la rédaction des métadonnées (titre, résumé et mots clés). Ces recommandations répondent aux exigences d'une publication scientifique tout en garantissant une diffusion la plus large possible sur le web.

Le groupe de travail a été mis en place en janvier 2020 avec pour objectif de rédiger un guide de bonnes pratiques à l'attention des auteurs et des équipes éditoriales. Composé des membres des trois revues (*Cybergeo*, *Brussels Studies* et *ViaTourism*) concernées par ces questions (l'encyclopédie *Hypergeo* n'utilisant pas de métadonnées de ce type dans son contenu éditorial), le groupe s'est réuni à plusieurs reprises par visioconférences pour discuter des problématiques et décider collectivement des préconisations.

Les moteurs de recherche généralistes ou spécialisés, les bases de données professionnelles, institutionnelles, les sites d'archives ouvertes, etc. collectent les publications essentiellement à partir des métadonnées. Ces dernières ont donc un statut à part du reste de l'article et un rôle déterminant dans sa diffusion et son impact.

De fait, la qualité des métadonnées, tant dans la langue originale que dans la/les langue(s) traduite(s) est essentielle pour une bonne diffusion des publications sur le net. C'est par elles que les internautes vont prendre connaissance des publications, tout comme les sites de référencement. En effet que les sources soient publiques ou privées, les métadonnées sont la seule partie de la publication qui est dans tous les cas librement accessible.

Un soin tout particulier doit donc y être apporté qui prenne en compte plusieurs exigences : une structure et un langage clairs faisant ressortir les points saillants de l'article, une syntaxe et un vocabulaire facilement analysables par les moteurs de recherche sans toutefois en appauvrir le sens, une traduction scientifiquement de qualité.

Les trois types de métadonnées (titre, résumé et mots clés) remplissent chacun une fonction spécifique et il convient de réfléchir à un choix de mots et d'écriture propre à chaque catégorie, tout en leur reconnaissant une forme de complémentarité. En effet, l'insertion d'un mot important dans le contenu du titre et des résumés a autant d'impact que dans la liste des mots clés qui en principe est fermée. On peut introduire le(s) mot(s) spécifique(s) de l'article dans le titre et/ou le résumé qui seront également analysés par les moteurs de recherche, c'est le cas par exemple pour les toponymes rares qui ne décrivent que très peu d'articles dans l'ensemble de la revue.

Le but de ce petit guide est d'aider les auteurs à adopter une écriture rigoureuse avec des concepts facilement analysables pour le repérage et l'accès aux métadonnées sur le web, sans toutefois proposer de formes et de structures qui seraient trop stéréotypées et sans réduire excessivement la richesse et la diversité de la langue.

Le travail du groupe s'appuie sur les tests de traduction des métadonnées en chinois, anglais et espagnol décrits dans la deuxième partie de ce rapport ([Sur la traduction automatique des métadonnées](#))

Les recommandations de ce guide sont intégrées à la note aux auteurs affichée sur le site des revues. Lorsqu'un article sera accepté pour publication, il sera de nouveau demandé aux auteurs d'appliquer rigoureusement les conseils du guide de bonnes pratiques.

Le texte du guide est reproduit en [annexe](#).

Tests de la qualité des logiciels de traduction automatique gratuits

Contributeurs : Colette Cauvin-Reymond, Bernard Elissalde, Nathalie Lemarchand, Denise Pumain

La traduction automatique repose sur *des probabilités*. En multipliant les sources, les algorithmes essayent de se rapprocher le plus possible de la traduction parfaite, c'est-à-dire la plus fréquemment observée. Ils font appel à des calculateurs qui vont piocher sur internet plusieurs milliers de textes déjà traduits et les recourent entre eux. L'ordinateur choisit, en fonction des données qu'il trouve, la combinaison qui lui semble la plus probable, mais probable ne veut pas dire juste. Il repère les cooccurrences les plus fréquentes entre langue-source et langue-cible, et en infère des traductions potentielles au niveau des mots¹⁵.

Chaque correspondance lexicale possède, au début du travail des algorithmes, la même probabilité, mais le fait que deux mots apparaissent régulièrement en regard l'un de l'autre va progressivement renforcer leur probabilité d'être traduction l'un de l'autre. Les algorithmes de traduction automatique tentent de répondre au défi d'optimiser la méthode permettant d'approcher les différents sens d'un mot donné uniquement en observant son usage dans le vaste corpus des publications en ligne.

Les tests de traduction des métadonnées des articles et de chapitres ont été réalisés en anglais, ainsi qu'en espagnol par des experts, avec pour objectif la conservation du sens des textes traduits. Les tests avec l'anglais comparent la traduction humaine et la traduction automatique des métadonnées. La méthode de retraduction immédiate dans la langue d'origine a été employée comme première vérification du maintien du sens. L'outil de traduction automatique utilisé est le logiciel partiellement libre DeepL¹⁶.

1. Tests en anglais et espagnol sur l'article "Mesurer l'évolution de l'accessibilité à l'aide des "espaces-temps d'action" : l'exemple de Bogotá entre 1993 et 2009"- Nathalie Lemarchand

(Measuring changes in accessibility with the help of "spaces-time of action": the example of Bogotá between 1993 and 2009)¹⁷.

Nathalie Lemarchand a utilisé DeepL pour ces tests. Elle a traduit d'une part les titres, intertitres et titres des figures de l'ensemble de l'article et

¹⁵ Poibeau T., 2016, "Traduire sans comprendre ? La place de la sémantique en traduction automatique", *Langages* /1, N°201, p. 77-90.

¹⁶ DeepL n'est en effet pas à proprement parler un logiciel libre, même s'il est utilisable gratuitement pour un nombre limité de caractères, il reste payant au-delà.

¹⁷ <https://journals.openedition.org/Cybergeog/33638>

paragraphe par paragraphe (ou demi-paragraphe) la 1^{ère} partie de l'article. Elle a rédigé deux fichiers qui rendent compte dans le détail de ces tests comparatifs (traduction humaine et traduction automatique avec DeepL) : i) traduction en anglais ii) traduction en espagnol présentés en [annexes](#) .

Methodologie

À chaque fois, N. Lemarchand a traduit (1) du français vers l'autre langue (français/anglais et français/espagnol), puis à partir de la traduction en anglais ou en espagnol, elle a retraduit en français (2), enfin elle a comparé avec la version française originale.

LesTA avec DeepL en anglais et en espagnol ont été conservées dans leur intégrité, mais elle a pu plus facilement « mesurer » les changements en anglais. De plus avec l'anglais, elle a fait deux tentatives de modifications automatiques d'un mot et DeepL lui a immédiatement proposé la traduction par un mot ou une expression plus exacts.

Dans les deux fichiers en annexes 6 et 7 qui détaillent les traductions, les mots surlignés en jaune sont les transformations les plus probantes et transformant la version originale (significativement entre les deux versions françaises).

Dans chaque fichier, ont été ajoutées des remarques intitulées « bilan » pour souligner les problèmes relevés avec la TA [DeepL](#).

Une synthèse des observations faites des effets des traductions est fournie, en 1^{ère} page de ces deux fichiers, dont voici l'essentiel avec les mêmes constats en anglais et en espagnol :

- changement du pluriel en singulier
- disparition de certains mots et/ou transformation du sens par effet de « correspondance »
- contresens liés à des expressions tels que « faire les frais »
- mélange des sigles anglais (!) et français (ex : ETA EA ; UGI IGU).

2. Sur le chapitre d'ouvrage : Cartographie thématique, Volume 5, Le son – Colette Cauvin-Reymond

Un autre test avec DeepL a été réalisé par C. Cauvin sur un chapitre d'ouvrage sur le son, issu d'un ouvrage publié¹⁸ . Il montre que les textes très techniques dont le vocabulaire est connu ou fourni par l'auteur sont traduits correctement avec la TA, mais toujours un peu littéralement. L'intervention humaine ne modifie pas, la plupart du temps, le texte, mais parfois elle intervient au niveau de la structure de la phrase en coupant, par exemple, les phrases un peu longues.

Lors de la rédaction du livre dont est extrait le passage sur le son (cf annexe 8), il est apparu que fournir un glossaire des termes d'une spécialité était indispensable pour le traducteur. Ce

¹⁸ Cauvin, Escobar, Serradj, 2008, version française, Vol.5, et 2010, version anglaise, Ch.1.2.4.2., *Ouvrage op. cit.*

glossaire peut se construire au fur et à mesure de la production du texte, mais nécessite toujours de nombreux échanges et des vérifications.

Les termes très techniques par ex. sur le son sont traduits de la même façon dans les deux cas alors que les termes généraux peuvent changer. Les difficultés principales se rencontrent avec les mots intermédiaires entre le très technique et le vocabulaire courant, ou lorsque le terme est usuel mais a un sens spécifique dans le domaine concerné (ex. généraliser, généralisation).

3. Tests comparatifs de traduction humaine et automatique en espagnol sur 3 articles de Cybergeo – Francisco Maturana

Le processus de traduction et de comparaison a été également réalisé sur 3 articles de Cybergeo. Bien que possédant des thématiques variées, ils sont en relation avec la géographie. Les difficultés de traduction rencontrées sont diverses et peuvent se résumer ainsi :

- Le style d'écriture plus ou moins complexe des auteurs.
- L'équivalence des mots techniques à l'espagnol.
- Certains textes présentaient en langue originale des problèmes de rédaction tels que des expressions redondantes et/ou autres, ce qui a rendu le processus de traduction plus difficile en devant prendre des décisions sur la traduction à choisir afin de ne pas faire les mêmes erreurs dans la version en espagnol, mais en maintenant la structure de ce qui est écrit. Cela se produit plus fréquemment dans les résumés.
- Parfois, il fut nécessaire de prendre des repères pour savoir s'il fallait traduire les noms de villes, de régions, de pays ou autres. Néanmoins, il est difficile d'établir un critère sur ceci, car certaines villes, villages ou lieux ont une traduction, d'autres non.

En ce qui concerne le processus de traduction, si on compare celui-ci aux possibilités qu'offrent les plateformes automatiques, on rencontre les problèmes suivants :

- Problèmes de reconnaissance du vocabulaire technique.
- Changement de certains mots qui modifient le sens ou l'intensité de la phrase.
- Les longs paragraphes sont moins bien traduits que les phrases courtes.
- Le titre des articles sont généralement bien traduits par les programmes automatiques.
- La traduction du résumé généralement est bien traduit, mais ils emploient une rédaction en espagnol moins élégante ou scientifique. La traduction correcte est attribuée au fait qu'en général des phrases courtes et précises sont faites dans cette section. Mais il faut remarquer que depuis que nous traduisons des articles, nous avons observé que les logiciels de traductions automatiques présentent une amélioration.
- L'hors d'une traduction automatique du texte complet, nous pouvons remarquer des nombreux problèmes dans le style d'écriture, même si elle reste compréhensible pour un hispanophone, il peut y avoir des changements dans l'intensité ou dans les expressions spécifiques qui ne clarifient pas le contenu proposé. Cela est attribué au fait que les phrases sont souvent plus longues l'utilisent des connecteurs, des pronoms ou des adjectifs, qui facilitent les erreurs de la part des traductions automatiques.

- Effectuer des traductions automatiques puis les comparer avec celles effectuées personnellement contribue à constater les erreurs. Des erreurs ou parfois des appréciations différées peuvent être observées après une quatrième voire cinquième relecture et contribuer à clarifier et à peaufiner la traduction.

Des exemples sont présentés de manière détaillée dans l'annexe 9.

4. Tests dans les articles de l'encyclopédie *Hypergeo* – Bernard Elissalde

Tout au long de ces quelques lignes, nous avançons dans l'analyse des éventuelles modifications de sens provoquées par la traduction automatique ainsi que le devenir de certaines figures de rhétorique observées dans des articles de l'encyclopédie. *Les parties en italiques sont celles qui ont posé problème.*

Correspondance et équivalence

Nous partons de la distinction classique entre correspondance et équivalence¹⁹ (cf. Lederer, 2002).

Correspondance

Dans le cas de la "correspondance", le traducteur passe d'une langue à l'autre, introduisant dans son texte des correspondances lexicales ou syntaxiques préexistantes et validées par des dictionnaires ou pour les termes techniques, dans les lexiques ou glossaires spécialisés. Cette simplicité apparente de la traduction par "correspondance" est remise en cause par la situation fréquente de différences des champs sémantiques couverts par un concept en passant d'une langue à une autre : ex : territoire. Il s'agit d'un concept polysémique et objet de débats en français mais dont un seul sens est reconnu en anglais académique. Pour autant, le sens élargi commence à apparaître dans les textes en anglais concernant le « spatial planning » dans l'Union européenne (programmes ESPON). Idem pour « paysage » dont les champs sémantiques en français, anglais, mais aussi espagnol ou allemand, sont fortement marqués par les traditions des « écoles » de pensée nationales.

Un autre cas de figure apparaît dans le domaine « correspondance ». Les choses se compliquent pour la traduction scientifique lorsqu'il existe un correspondant linguistique dans l'autre langue, mais qu'il est nécessaire d'inventer pour la traduction un équivalent « scientifique » rendant compte du champ théorique de référence de l'auteur. Exemple : problème posé par le titre à trouver en anglais pour traduire l'article « habiter » d'O. Lazzarotti, intégrant ou suggérant tout le bagage théorique que le corps de l'article va développer.

¹⁹ Lederer J., "Correspondances et équivalences - Faits de langue et de discours en traduction", in *La traduction comme relation, Identité, altérité*, Editors: Israël F., Didier Érudition, Paris,

Equivalence

Sur le volet "équivalence", le traducteur après avoir trouvé le 'sens' du texte, crée des termes pertinents qui transmettent des sens identiques à ceux de l'original. Il l'exprime par des équivalences, de façon idiomatique dans la langue d'arrivée. Tel est le cas du concept de « déguerpissement » qui est utilisé dans les recherches en français consacrées aux villes des pays pauvres et très spécifique à la langue française. Ce terme est traduit par « eviction » considéré comme un équivalent en anglais mais qui réduit le sens initial. En effet si on refait la traduction de l'anglais vers le français, on obtient alors la traduction "éviction".

D'autres problèmes sont soulevés par cette approche

- La fabrication d'un néologisme dans la langue d'arrivée pour rendre compte d'un équivalent du terme de la langue de départ. Ex : un concept objet de débats chez les chercheurs en français : anthroposystème ou encore comment traduire « géographicité » : par geographicity ?
- Problème avec le terme "citoyen" qui peut produire des hésitations entre *city-dweller* et *citizen*.

Caractéristiques du discours et traduction

Glissements de sens et contresens

Le texte "déguerpissement" tiré d'Hypergéô, ainsi que quelques autres petits extraits ont servi de base aux tests sur la TA (DeepL) et les problèmes posés lors de la réversibilité. "Déguerpissement" a été choisi en raison de son ancrage idiomatique, politique et conceptuel dans le vocabulaire des sciences humaines francophones, et de sa trajectoire depuis le sens commun vers la recherche universitaire. Dans le texte déguerpissement : *réhabilitation* et *renovation* sont employés l'un pour l'autre, tout comme *relogement* et *réinstallation*. Propension à fabriquer de la synonymie entre les versions, en lieu et place des subtilités notionnelles et sémantiques de l'auteur.

Certaines expressions ont des équivalences mais ces équivalences modifient le sens dans la chaîne de traduction. Exemple : « *acteurs de politiques néo-libérales* » devient avec la traduction humaine : *neoliberal political agents* et aboutit avec la traduction automatique inverse à : « *agents politiques néolibéraux* ». Ici le contresens est-il uniquement imputable à la TA ?

Dans le texte « territoire selon Di Méo », la traductrice n'a pas trouvé de correspondant à l'expression « espace vécu ». Cette expression qui renvoie à toute une tradition dans la géographie francophone, a été traduite par l'équivalent « lived-in space ». Néanmoins lorsque l'on utilise la traduction automatique (Google ou DeepL) pour retraduire le texte de Di Méo en français, l'expression « lived-in space » devient « l'espace habité », ce qui, pour le moins, appauvrit le concept d'espace vécu.

Traduction Automatique et modification de sens

"Constructivisme"

Texte original en français : Le terme de constructivisme désigne une posture philosophique pour laquelle *toute* réalité n'est connaissable qu'à travers des catégories *préalables*.

Traduction humaine : The term constructivism refers to a philosophical stance according to which *any reality* can only be known through *pre-defined* categories.

Traduction DeepL : The term constructivism refers to a philosophical posture for which *all reality* can only be known through *previous* categories.

Traduction reverse DeepL vers français : Le terme constructivisme fait référence à une posture philosophique pour laquelle *toute la* réalité ne peut être connue qu'à travers les catégories *précédentes*.

Commentaire : Ici la traduction automatique aboutit à une erreur de traduction et même à un contresens (toute la réalité). Il en va de même pour la confusion entre « idéal » et « idéel » dans l'extrait suivant de l'article « Habiter ».

"Habiter" (extrait 1)

Extrait : Strictement théorie de l'action dans la première édition du dictionnaire des sciences de l'espace, la définition de l'habiter s'est alors, dans la seconde, ouverte aux « *réalités idéelles* ».

Traduction humaine : Finally, the definition of "inhabiting" was strictly a theory of action in the first edition of the Dictionary of the sciences of space, while in the second edition it opened up to "*ideated realities*".

Traduction DeepL : Strictly a theory of action in the first edition of the Dictionary of Space Sciences, the definition of inhabiting it was then, in the second edition, opened to "*ideal realities*".

Traduction reverse DeepL vers français : Strictement une théorie de l'action dans la première édition du Dictionnaire des sciences spatiales, la définition de l'habiter a ensuite été, dans la deuxième édition, ouverte aux "*réalités idéales*".

Traduction et tournure grammaticale

"Habiter" (extrait 2)

Extrait : Habiter : se construire en construisant le « Monde ». Ainsi défini comme résultats autant que processus (Lazarotti, 2014), l'« habiter » conceptualisé désigne la dimension géographique de l'humanité : toute collectivité et chaque humain habitent ici et/ou là, comme ceci ou comme cela pour partager les « lieux » et les « **territoires** » du Monde. Ceux-ci sont

donc, à la fois, une donnée reçue et ce que chaque habitant et toute société transforment en les habitant.

Traduction humaine : *Inhabiting*: forming oneself as one forms the World. Thus defined as a result as much as a process (Lazarotti, 2014) the concept of "*inhabiting*" (or the more archaic "inhabitancy") relates to the geographical dimension of humanity: every collective and every individual living here and/or there, in this way or that, so as to share **places** and **territories** across the World. Thus place and territory are at once a given fact, and also what *each inhabitant* and each society have altered by *living* in them.

Commentaire de la traductrice : pour la question de « l'habiter », j'ai choisi l'emploi d'un gérondif (qui est donc une substantivation) mais pas avec le verbe "to live" (vivre) mais avec "to inhabit". L'Oxford English Dictionary donne même le terme "inhabitancy".

Traduction DeepL : *Living*: building oneself by building the "World". Thus defined as a result as well as a process (Lazarotti, 2014), the conceptualized "*inhabiting*" refers to the geographical dimension of humanity: every community and every human being inhabits here and/or there, like this or like that, in order to share the "places" and "territories" of the World. These are therefore both a received data and what each inhabitant and all societies transform into the *inhabitant*(?)

Traduction reverse DeepL vers français : *Vivre* : se construire en construisant le "Monde". Ainsi défini comme un résultat autant que comme un processus (Lazarotti, 2014), l'"habiter" conceptualisé renvoie à la dimension géographique de l'humanité : chaque communauté et chaque être humain habite ici et/ou là, comme ceci ou comme cela, afin de partager les "lieux" et les "territoires" du Monde. Il s'agit donc à la fois d'une donnée reçue et de ce que chaque habitant et toutes les sociétés transforment *en habitant*.

Commentaire : Cet article a été choisi pour les problèmes grammaticaux qu'il soulève. En français, il est possible de transformer un verbe en substantif. Cela s'appelle la « substantivation infinitive », que l'on retrouve également dans des expressions diffusées dans le grand public et dans les médias, comme "le vivre ensemble". La question se pose sur la manière dont cette tournure est traitée en anglais par la TA.

Avec la traduction automatique, la substantivation infinitive est bien prise en compte mais se fait avec un affaiblissement du sens (« Living ») ce qui dans le retour vers le français débouche sur « vivre » au lieu de « inhabiting »

Capacité à prendre en compte des néologismes et des nouveaux concepts

Article « communauté transnationale »

On est en présence de processus grâce auxquels les immigrants construisent et maintiennent des rapports sociaux multiples entre leur société d'origine et celles de leur installation. Les transmigrants maintiennent ces relations multiples par-dessus les frontières, et leur situation

ne peut être pleinement définie, ni dans le pays d'installation, ni dans celui d'origine, mais dans le champ social formé entre les deux. Dans cette optique, les citoyens d'un État-nation vivent dispersés à l'intérieur des frontières de divers autres États, mais lui appartiennent toujours socialement, politiquement, culturellement, et souvent économiquement. Ces rapports se situent dans un espace associant des États centraux, capitalistiquement hégémoniques, et des États périphériques dominés. Les réseaux en toile d'araignée relient ces transmigrants dans le champ d'une construction de domination hégémonique entre ces divers États. L'identité des différents groupes sociaux doit être reconsidérée, dans la mesure où ils ne sont plus territorialisés ni inclus dans des espaces clairement délimités par des frontières, espaces qui ne sont pas culturellement homogènes. Les concepts de nation et de groupe ethnique dans ce cas ne se réfèrent plus à des entités stables, clairement délimitées. Les identités se définissent de plus en plus par rapport à des pouvoirs dominants et, en opposition à eux, avec des frontières non strictement définies. Elles se configurent en fonction de catégories hégémoniques, telles que la race ou l'ethnicité, et sont profondément impliquées dans les processus de constitution de ces États-nations. Il faut donc se situer dans une perspective de dépassement des catégories nationales et ethniques dominantes, puisque ces populations transmigrantes ont commencé à bâtir des États-nations déterritorialisés, ce qui suppose une construction sociale différente de celle d'une diaspora.

Traduction humaine : Transnational community

This is a process whereby immigrants construct and sustain numerous social relationships between the society from which they originate and that in which they have settled. Transmigrants sustain these many relationships across borders, and their particular situation can be fully defined neither within their country of residence, nor within their home country, it can only be defined within the social sphere formed in between. In this perspective, citizens of a nation-state may live dispersed within the boundaries of various other states, but they still belong to it socially, politically, culturally and often economically. These relationships are positioned in a space that associates central states that are economically hegemonic and peripheral states that are dominated. Spider's web networks link these transmigrants within a construction of hegemonic domination among these various states. The identity of the different social groups then requires reconsideration, since they are no longer territorialised, nor are they included in spaces clearly delineated by «borders», or spaces that are culturally homogeneous. The concepts of nation and ethnic group in this case no longer refer to stable entities that are clearly delineated. Identities are increasingly defined by reference to dominant powers, and in opposition towards them, and boundaries that are not strictly defined. Identities are shaped according to hegemonic categories such as race or ethnic group, and are profoundly implicated in the process of constitution of these nation-states. Thus it can be seen that we have moved on from national categories and dominant «ethnic groups», since these transmigrant populations have started to construct de-territorialised nation-states, which supposes a social construction that is different from that of a «diaspora».

Traduction DeepL : Transnational community

We are in the presence of processes through which immigrants build and maintain multiple social relationships between their society of origin and the societies in which they settle. Transmigrants maintain these multiple relationships across borders, and their situation cannot be fully defined, neither in the country of settlement nor in the country of origin, but in the social field formed between the two. From this perspective, the citizens of a nation state live dispersed within the borders of various other states, but still belong to it socially, politically, culturally, and often economically. These relationships are situated in a space between central, capitalist hegemonic states and peripheral, dominated states. Spider-web networks link these *transmigrants in the field of a hegemonic construction of domination between these various states*. The identity of the different social groups must be reconsidered, insofar as they are no longer territorialized or included in spaces clearly delimited by borders, spaces that are not culturally homogeneous. The concepts of nation and ethnic group in this case no longer refer to stable, clearly demarcated entities. Identities are increasingly defined in relation to and in opposition to dominant powers, with borders that are not strictly defined. They are configured according to hegemonic categories, such as race or ethnicity, and are deeply involved in the processes of constitution of these nation-states. It is therefore necessary to look beyond the dominant national and ethnic categories, since these transmigrant populations have begun to build deterritorialized nation-states, which implies a social construction different from that of a diaspora.

Commentaire : Limites de DeepL: Si les notions de « communauté transnationale » et de « transmigrants » sont bien pris en compte par la traduction automatique, le fait que les choix de correspondance lexicale soient fondés sur l'ensemble du corpus présent sur le web et non pas aux seules formes langagières académiques peut conduire à des traductions utilisant un vocabulaire plus courant ainsi qu'à l'utilisation d'une syntaxe plus « littérale » et donc moins idiomatique que dans les textes scientifiques.

Traductions et Figures de rhétorique

Nous avons testé quelques figures de rhétorique, parmi les plus usitées, pour observer soit leur non-traductibilité, soit leurs transformations éventuelles suite à une traduction automatique ou selon une traduction humaine.

Acronymie

Article "Géosystème"

Extrait : Cette évolution est formalisée par Georges Bertrand en 1991 à travers le « GTP : géosystème, territoire, paysage ».

Traduction DeepL : This evolution was formalised by Georges Bertrand in 1991 through the "GTP: geosystem, territory, landscape".

Commentaire : faible influence de la technique de traduction (TA ou traduction humaine), seul l'ordre des mots change d'une langue à une autre : OCDE (fr) OECD (eng)

Analogie

Article " Dynamique des systèmes selon Forrester"

Extrait : La construction d'un modèle forrestérien démarre par la réalisation d'un diagramme sagittal. Ce diagramme représente les différents éléments qui composent le système en termes de *stocks et de flux*, et précisent les relations établies entre les différentes variables du système modélisé. C'est ce qu'on appelle le langage Dynamo. Les éléments de base du modèle sont : les variables de niveau ou d'état. Elles fonctionnent *comme des accumulateurs*. Ainsi, les variables d'état représentent des stocks dont la quantité varie à travers le temps en fonction des flux d'entrée et de sortie qui les alimentent et les vident. *Les stocks sont utilisés pour représenter tant des accumulations matérielles* (l'eau, les individus) qu'immatérielles (la connaissance). La valeur de ces stocks renseigne sur l'état du système à chaque instant t. À chaque instant t, un stock représente l'accumulation passée des flux d'entrées et de sorties.

Traduction DeepL : The construction of a forresterian model starts with the construction of a sagittal diagram. This diagram represents the different elements that make up the system in terms of *stocks and flows*, and specifies the relationships established between the different variables of the modelled system. This is called the Dynamo language. The basic elements of the model are the level or state variables. They function like *accumulators*. Thus, state variables represent stocks whose quantity varies over time according to the input and output flows that feed and empty them. *Stocks are used to represent both tangible (water, people) and intangible (knowledge) accumulations*. The value of these stocks provides information on the state of the system at each point in time t. At each instant t, a stock represents the past accumulation of inflows and outflows.

Énumération

Article "environnement"

Extrait : Ce terme recouvre à la fois un milieu physique et un système de relations, un champ de forces physico-chimiques et biotiques en interrelation avec la dynamique sociale, économique, spatiale

Traduction DeepL : This term covers both a physical environment and a system of relationships, a field of physico-chemical and biotic forces interrelated with social, economic and spatial dynamics.

Euphémisme

« Pays en voie de développement », « pays du sud » ou « pays émergents » pour désigner les « pays sous-développés », en opposition aux « pays riches »

Commentaire : Problème de traduction entre euphémisme et vrai concept en passant d'une langue à une autre. Le degré d'euphémisme n'étant pas le même, à un moment donné, en passant d'une « culture » (école ?) géographique nationale à une autre

Explétion

Exemple : Le salaire moyen, *faut-il le préciser*, est aujourd'hui inférieur à ce qu'il était il y a cinq ans → The average wage, *it must be said*, is lower today than it was five years ago (DeepL).

Article « IBN KHALDÛN » (Hypergeo) L'énorme Kitâb al-'Ibar (le Livre des Exemples) qui est une histoire universelle précédée d'une autobiographie et surtout précédée de l'ouvrage le plus important, une véritable anthropologie politique, la Muqqadima (les Prolegomènes) qu'il considère, orgueilleusement (*mais c'est justifié !*) comme une science nouvelle, inventée par lui, « la science de la société humaine ».

Traduction humaine : The vast Kitab al'Ibar (the book of lessons), which is a universal history, preceded by an autobiography and above all by his most important work, a veritable political anthropology the Muqqadima (Prolegomena) which he proudly (*but rightly*) considered to be a new science, his own invention, "the science of human society".

Traduction DeepL : The enormous Kitâb al-'Ibar (the Book of Examples) which is a universal history preceded by an autobiography and above all preceded by the most important work, a true political anthropology, the Muqqadima (the Prolegomena) which he proudly (*but justifiably!*) considers to be a new science, invented by him, "the science of human society".

Hyperbole

Article "J. Gottman"

Extrait : Géographe *hors norme* dans le monde académique de la Guerre Froide

Traduction DeepL: *Outstanding geographer* in the Cold War academic world

Retour vers français par Google translation : *Géographe exceptionnel* dans le monde universitaire de la guerre froide

Article "A. Demangeon"

Extrait : *Son oeuvre est impressionnante* : une douzaine de livres (plus une dizaine de manuels scolaires), une centaine d'articles, un millier de notes et de comptes rendus ! »

Traduction DeepL : *His work is impressive*: a dozen of books (plus a dozen textbooks), a hundred articles, a thousand notes and reviews

Traduction humaine : *His work is impressive*: a dozen books (more than ten school textbooks), and some hundred articles and reports.

Litote

Exemple : Va, je ne te hais point (Corneille) = - Go, I don't hate you

Article "Occident" (H. Chamussy, *Hypergeo*) : La civilisation occidentale est, selon le cas, celle du christianisme qui va évangéliser le monde, ou celle des Lumières, qui va faire régner la liberté et la fraternité ; *du moins c'est ce que l'on dit*.

Traduction humaine : Western civilisation is sometimes seen as that of Christianity with the calling to evangelise the world, and sometimes as that of Enlightenment, supposed to usher in the reign of liberty and fraternity, *whatever the actual truth of the matter* (interprétation de la traduction humaine).

Traduction DeepL : Western civilization is, depending on the case, that of Christianity, which will evangelize the world, or that of the Enlightenment, which will make freedom and fraternity reign; *at least that is what is said...* (DeepL traduction plus littérale)

Métaphore

Exemple : *La vie est un long fleuve tranquille* → Life is a long, calm river

Article "développement" (*Hypergeo*)

Extrait 1 : Après les économistes, les géographes se sont *emparés (to grab)* de la question

Traduction humaine : Following the economists, geographers *took up* the issue.

Traduction DeepL : After economists, geographers *have taken up* the question.

Extrait 2 : Les inégalités sont internes au développement lui-même, lequel *bouscule* les hiérarchies existantes, et en crée d'autres.

Traduction humaine : Inequalities are inherent in development itself, because it *upsets (déranger)* existing hierarchies, creates new hierarchies.

Traduction DeepL : Inequalities are internal to development itself, which *shakes up (secouer)* existing hierarchies and creates new ones.

Article "mondialisation"

Extrait 1 : Auparavant les précédentes formes de l'espèce humaine étaient restées dans le *berceau* africain et l'Eurasie.

Traduction DeepL : Previously, the previous forms of the human species had remained in the African *cradle* and Eurasia".

Extrait 2 : Mais cette diffusion, si elle innove profondément, prolonge également les circuits d'échange qui *irriguaient* l'Ancien Monde depuis le paléolithique.

Traduction DeepL : "But this diffusion, while profoundly innovative, also extends the circuits of exchange that have *irrigated* the Old World since the Paleolithic period.

Extrait 3 : Le sida n'est que le *dernier avatar* de nombreuses pandémies mondiales.

Traduction DeepL : AIDS is only the *latest manifestation* of many global pandemics.

Google Translation : AIDS is just the *latest avatar* of many global pandemics.

Article "banlieue"

Extrait : Dans une société qui se cherche, la banlieue situerait donc l'épicentre des problèmes sociaux et le *lieu* d'application par excellence de la « politique de la ville ».

Traduction DeepL : In a society that is searching for itself, the suburbs would therefore be *the epicentre* of social problems and the place where "urban policy" is applied par excellence » (DeepL).

Commentaire : Aucun problème de traduction si utilisation de la langue courante.

Métonymie

Exemple : L'Occident est alors assimilé à une partie de l'Europe (description d'une métonymie), précisément qualifiée d'Europe occidentale

Traduction humaine : This assimilated the Occident to part of Europe, which came to be known as Western Europe.

Traduction DeepL : The West is then assimilated to a part of Europe), precisely described as Western Europe.

Article "Est"

Extrait : dans le contexte spécifique de la guerre froide, le terme a connu un emploi considérable, et l'affrontement Est-Ouest a marqué les relations internationales pendant des décennies. *Moscou et Washington* ont pris l'Europe comme enjeu central.

Traduction DeepL : In the specific context of the Cold War, the term was used extensively, and the East-West confrontation marked international relations for decades. *Moscow and Washington* have taken Europe as a central issue.

Traduction humaine : in the specific setting of the Cold War, the word East was widely used, and the East-West contrast was part of international relations for decades. From *Moscow and Washington*, Europe was the central focus.

Oxymore

Extrait (Victor Hugo) : Le plus lourd fardeau c'est d'exister sans vivre → The heaviest burden is to exist without living. (DeepL)

Extrait (F.Braudel) : La première partie met en cause une histoire quasi immobile, celle de l'homme dans ses rapports avec le milieu qui l'entoure

Traduction DeepL : The first part questions an almost immobile history, that of man in his relationship with the environment that surrounds him.

Extrait article "Habiter" (Hypergeo) « Muette mais pas silencieuse... » → « Mute but not silent » (DeepL).

Parataxe

La parataxe des mots ou « style télégraphique » consiste à supprimer la subordination entre des propositions, ex : « *Accident voyageur* » pour « *accident survenu à un voyageur* »

Article "géosystème"

Extrait : S'inscrivant dans une démarche systémique, il est utilisé en géographie pour étudier les interactions nature-sociétés.

Traduction humaine : It belongs to systemic approaches generally, and is used in geography to study interactions between the natural environment and societies.

Traduction DeepL : As part of a systemic approach, it is used in geography to study nature-society interactions.

Commentaire : comme déjà observé, la traduction fournie par DeepL est plus « littérale »

Périphrase

Les périphrases dans une langue fonctionnent par équivalence dans l'autre langue et donc introduisent des glissements de sens que seul l'auteur peut contrôler : ex : *Les groupes les plus défavorisés (article déguerpissement)* qui devient *underprivileged groups* ou *disadvantaged groups*. À cela s'ajoutent des connotations de date, d'époque, ou de référence théorique, voire idéologique à l'intérieur d'une même langue (ex : l'ancienne catégorie « *économiquement faible* » devenant « *population défavorisée* »).

Article "Est"

Extrait : d'où l'usage rapidement répandu du terme de Pays de l'Est pour désigner dans la langue courante – depuis les Etats-Unis jusqu'à l'Europe occidentale libérale – les pays européens du camp socialiste. [...] *Pour autant, on ne parlait pas de "pays de l'Ouest" mais d'Europe Occidentale*. Cet usage ancré dans un contexte *géopolitique* précis n'a pas survécu à la disparition du bloc communiste. Les "Pays de l'Est" et "l'Europe de l'Est" ont fait naufrage, laissant la place à l'Europe centrale et Orientale, termes moins chargés des souvenirs de l'affrontement Est-Ouest. On a ainsi vu fleurir l'expression délicieusement paradoxale d'"*ex-pays de l'Est*".

Traduction DeepL : Hence the rapidly spread use of the term "Eastern European countries" to designate the European countries of the socialist camp in everyday language - from the United States to liberal Western Europe. [...] *However, the term "Western countries" was not used to refer to "Western countries" but to Western Europe. This usage, anchored in a precise geopolitical context, did not survive the disappearance of the communist bloc. The "Eastern countries" and "Eastern Europe" were wrecked, giving way to Central and Eastern Europe, terms less charged with memories of the East-West confrontation. We have thus seen the deliciously paradoxical expression "former Eastern countries" blossom.*"

Traduction humaine : Leading very soon to the widespread use of "Eastern Bloc" in everyday language in western Europe and the USA to refer to the European countries belonging to the Soviet Bloc. [...] However, there was not so much reference to the "Western Bloc" as to "Western Europe". The "Eastern" or "Soviet" Bloc foundered and has been replaced by Central and Eastern Europe, which are terms that are less reminiscent of the East-West confrontation. We have also seen the arrival of the equivocal phrase "countries from the former Soviet Union".

Commentaire : Les habitudes des expressions instituées dans les différents pays et les différentes langues posent problème à la traduction des périphrases et conduisent même à des interprétations erronées.

Personnification

Article "constructivisme"

Extrait 1 : Le constructivisme est devenu depuis deux décennies une sorte de *porte-étendard* pour de nombreux praticiens des sciences sociales (O. Orain, *Hypergeo*)

Traduction humaine : Constructivism has over the last two decades become *a sort of rallying flag (drapeau de ralliement)* for numerous practitioners in the social sciences.

Traduction DeepL : Constructivism has become over the past two decades a sort of *standard bearer* for many social science practitioners.

Extrait 2 : une théorie scientifique est vraie et fonctionnelle dans la mesure où elle obtient des succès par les actions sur le réel qu'elle rend possibles, et ce sont ces succès opérationnels qui fondent sa légitimité.

Traduction humaine : a scientific theory is true and functional insofar as *it achieves success by the action on the real world* that it renders possible. It is these operational achievements that are the basis of its legitimacy.

Traduction DeepL : a scientific theory is true and functional insofar as *it achieves success through actions on the reality* it makes possible, and it is these operational successes that form the basis of its legitimacy.

Prolepse

La prolepse est une figure de style dans laquelle le locuteur fait objection à son propre argument et y répond immédiatement. Ce faisant, il espère renforcer son argumentation en abordant d'éventuels contre-arguments avant que son auditoire puisse les soulever.

Article "banlieue"

Extrait : Le mot « banlieue » est polysémique [...] une signification symbolique relative à la marginalité et au discrédit qui pèsent sur *ceux qu'on qualifie « d'exclus » par réduction hâtive*.

Traduction humaine : [...] a symbolic meaning relating to marginality and a certain discredit cast upon the inhabitants, viewed, *by oversimplification, as "excluded" (from the city)*.

Traduction DeepL : [...] a symbolic meaning relating to the marginality and discredit that weighs down on *those who are described as "excluded" through hasty reduction*.

Article "imaginaire spatial"

Extrait : *Si l'on croit que l'image est ou peut être un reflet du réel*, la singularité et la pertinence de la notion d'imaginaire est discutée, voire niée. Une conception hyper-positiviste de la science, telle qu'elle a prévalu en géographie jusqu'au *milieu* du XXe siècle par exemple, postule que les images produites (cartes et photographies par exemple) donnent à voir le réel tel qu'il est.

Traduction DeepL : *If we believe that the image is or can be a reflection of the real*, the singularity and relevance of the notion of the imaginary is discussed or even denied. A hyper-positivist conception of science, as it prevailed in geography until the middle of the 20th century for example, postulates that the images produced (maps and photographs for example) show the real as it is.

Commentaire : DeepL a tendance à traduire en utilisant les mots du vocabulaire le plus courant sans chercher à produire un vocabulaire adapté au contexte des phrases d'origine.

Auteur(e)s et écritures.

Nécessité de tenir compte non pas uniquement du vocabulaire mais aussi de la complexité stylistique des articles. Exemple avec l'écriture très « riche » et très « figurée » d'Henri Chamussy avec l'article « Orient » dans *Hypergeo*.

Texte original en français : La fameuse « Question d'Orient », paradis des diplomates et terreur des étudiants en histoire est essentiellement la question de la lente décomposition de l'Empire ottoman et les ambitions rivales des Puissances *penchées au chevet* (métaphore) de l'« Homme malade », ce qui, *soit dit en passant* (explétion), étend l'Orient jusqu'à Sarajevo... Et c'est aux XVIIIe et XIXe siècles que se crée le mythe littéraire et artistique de l'Orient (le mythe de l'Extrême Orient viendra plus tard) ; le terme d'Orient joue un rôle important dans la Franc-Maçonnerie naissante (rôle qui perdure aujourd'hui : le Grand Orient de France) ;

après la mode des « turqueries », Voltaire, Montesquieu, puis Victor Hugo donnent à l'Orient des *lettres de noblesse* (métaphore) (qui n'ont pas grand-chose à voir avec l'Orient « réel » : plusieurs poèmes des Orientales de Hugo se situent *en ... Espagne !* (antithèse)) Bonaparte a un grand rêve oriental ; le voyage en Orient devient un *must littéraire* (périphrase) ; Chateaubriand, Lamartine, Byron, Nerval, et bien d'autres écrivains et peintres, feront ce voyage. La conquête de l'Algérie exacerbera cette mode en France : l'Algérie, terre d'Islam, est, bien entendu, « orientale ». Il est piquant de remarquer que l'Algérie est située au Maghreb, et que « Maghreb » en arabe, *veut dire* « couchant », « occident »... !(paradoxe).

Traduction DeepL : The famous "*Question of the Orient*", paradise for diplomats and terror for history students, is essentially the question of the slow decay of the Ottoman Empire and the rival ambitions of the Powers leaning at *the bedside of the "Sick Man"*, which, by the way, extends the Orient all the way to Sarajevo? And it was in the 18th and 19th centuries that the literary and artistic myth of the *East* was created (the myth of the Far East would come later); the term "*East*" plays an important role in the emerging Freemasonry (a role that continues today): the Grand Orient of France); after the fashion of the "turqueries", Voltaire, Montesquieu, then Victor Hugo gave the Orient *letters of nobility* (which have little to do with the "real" Orient: several poems of Hugo's Orientales are in ... *Spain!*). Bonaparte has a great oriental dream; the journey to the Orient becomes a *literary must*; Chateaubriand, Lamartine, Byron, Nerval, and many other writers and painters, will make this journey. The conquest of Algeria will exacerbate this fashion in France: Algeria, land of Islam, is, of course, "oriental". It is pungent to note that Algeria is located in the Maghreb, and that "Maghreb" in Arabic means "setting", "western"... !

Reverse vers français par DeepL : La fameuse "Question de l'Orient", paradis pour les diplomates et terreur pour les étudiants en histoire, est essentiellement la question de la lente décadence de l'Empire ottoman et des ambitions rivales des puissances qui se penchent au chevet du "Malade", qui, soit dit en passant, étend l'Orient jusqu'à Sarajevo ? Et c'est aux XVIIIe et XIXe siècles que se crée le mythe littéraire et artistique de l'Orient (le mythe de l'Extrême-Orient viendra plus tard) ; le terme "Orient" joue un rôle important dans la Franc-maçonnerie naissante (rôle qui se poursuit aujourd'hui) : le Grand Orient de France) ; à *la mode des "turquoises"*, Voltaire, Montesquieu, puis Victor Hugo donnent à l'Orient des lettres de noblesse (qui ont peu à voir avec le "*vrai*" Orient : plusieurs poèmes des Orientales d'Hugo *sont en ... Espagne !*) Bonaparte a un grand rêve oriental ; le voyage en Orient devient un *must littéraire* ; Chateaubriand, Lamartine, Byron, Nerval, et bien d'autres écrivains et peintres, feront ce voyage. La conquête de l'Algérie *va exacerber* cette mode en France : L'Algérie, terre d'Islam, est bien sûr "orientale". Il est piquant de constater que l'Algérie est située au Maghreb, et que "Maghreb" en arabe signifie "*cadre*", "occidental"... !

Commentaire : Cet extrait d'un texte de Henri Chamussy a été choisi parce qu'il illustre assez bien les qualités et les défauts de la TA par DeepL. Si la traduction des effets de style produits par les nombreuses figures de rhétorique qui jalonnent ce texte sont dans l'ensemble bien

rendues, tout comme certaines tournures (ex : « *qui n'ont pas grand-chose à voir avec l'Orient « réel »* » devient : *which have little to do with the "real" Orient*) par contre la richesse de la langue utilisée disparaît avec la traduction reverse de DeepL version anglaise vers le Français, soulignant la tendance de DeepL à proposer une langue courante plutôt qu'une langue académique.

Les expressions idiomatiques

Ex : cf texte de Nathalie Lemarchand : « *les étudiants font les frais d'une localisation résidentielle* » en [annexe](#). La phrase est une impasse pour une traduction automatique

L'opposition classique entre correspondance et équivalence. En prenant des textes relativement courts, on découvre les problèmes posés par l'écriture des auteurs qui vont ressortir avec des synonymies, des expressions idiomatiques, de la polysémie etc. On se rend compte qu'avec la traduction automatique l'écriture est simplifiée et sur des textes de l'encyclopédie, cela pose des problèmes. Ce problème rejoint les choix qui ont été faits dans le [guide des bonnes pratiques pour la rédaction des métadonnées](#).

Emplois de termes courants avec un sens scientifique particulier

La relecture humaine est là encore indispensable par exemple pour des mots de vocabulaires courants mais que l'on emploie avec un usage spécifique comme "approche", par exemple une approche économiste ou politiste.

Omissions et incomplétude

Dans le cas de traductions en réversibilité, on découvre des phrases non achevées et des bouts de phrases non traduits : *Il fait référence à l'expulsion par les forces publiques de tous les habitants dans un quart...* », ainsi que la disparition de certains mots ou expressions dans la phrase traduite automatiquement en anglais et aussi en espagnol, exemple dans les annexes (avec suivi TA par Google Translation et DeepL)

Omissions avec modification de sens : « les analyses conjointes des déplacements *au sein* des ménages » ----→ « Joint analyses of household travel and the evolution of mobility practices ».

Problèmes linguistiques plus généraux de la TA

Nous ne relevons ici que quelques exemples pour des termes assez fréquemment employés dans la discipline géographique ; il est possible de se référer à des ouvrages spécialisés pour éviter les mauvaises surprises à la traduction.

Non distinction du singulier et du pluriel : donnée et données en français, data en anglais.

Flexibilité de la traduction des prépositions : plusieurs tournures sont employées par les auteurs en français avec les prépositions ("pour", "de") qui peuvent être traduits par "for" (ex. dans la traduction réversible [des textes de Nathalie Le Marchand](#)) qui peuvent poser des

problèmes à la TA. Le terme "for" en anglais a un multiple usage, mais il s'agit là d'une catégorie de situation linguistique particulière qui fait partie du langage commun. On ne peut ici qu'appeler à la vigilance à l'égard des faux amis, prépositions, etc. sur lesquels nous ne reviendrons pas puisqu'ils ne sont pas propres à notre domaine d'études.

5. Conclusions

Parce que les algorithmes fonctionnent selon un calcul de probabilité de fréquence d'usage, les textes de TA ont tendance à choisir les correspondances lexicales les plus fréquentes. La traduction automatique tend donc à fabriquer du normatif, ce qui peut être intéressant pour l'enregistrement de métadonnées mais appauvrissant pour l'écriture des auteurs.

Le biais classique de la TA est d'aller vers la plus grande simplicité de la langue, qui a des avantages certains mais pourrait aussi nous conduire à nous demander s'il ne faudrait pas inventer un « Google scholar » de la TA. Doit-on pour autant réifier par l'analyse de discours l'opacité apparente de l'écriture de certains auteurs, dont la profondeur de sens devrait être retranscrite dans la traduction ?

En fait, puisque les méthodes d'IA (deep learning) qui jouent sur les fréquences d'associations des termes dans les textes et qui emploient des méthodes d'apprentissage par réseaux neuronaux sont entraînés sur des données massives, on peut imaginer que plus on soumettra aux algorithmes des textes académiques, plus la langue utilisée conviendra aux scientifiques... On peut notamment penser que les tests de *CybergeoNet* effectués dans DeepL et la vérification humaine (déjà faite dans Hypergeo par exemple) influencent la traduction dans DeepL. Il faudrait étudier de manière plus approfondie comment fonctionnent DeepL et Google Translation. Pour la traduction scientifique, peut-être faudrait-il soumettre l'idée d'un DeepL Scholar ?

À ce jour, les tests montrent bien le risque d'appauvrissement de la langue par la traduction automatique (TA) avec un vocabulaire et une syntaxe simplifiés voire simplistes pour ce qui concerne des articles entiers.

Toutefois les métadonnées sont à traiter différemment du reste de l'article. Des consignes précises sont fournies aux auteurs pour la rédaction de leurs métadonnées (cf. annexe guide des bonnes pratiques pour la rédaction des métadonnées). Le choix est fait par contre de conserver la richesse de la langue, le style, en faisant vérifier la traduction du texte intégral par les auteurs. C'est le cas pour *Cybergeo*, où les auteurs sont sollicités pour vérifier la conformité de la traduction avec le texte original. Il arrive même assez souvent que la traduction du texte intégral nécessite une reformulation des métadonnées et dans la langue originale, et dans la langue traduite. Nous tenons également compte de ces expériences dans [le guide de bonnes pratiques pour la rédaction et la traduction des métadonnées](#) élaboré dans le cadre de *CybergeoNet*.

Impact des opérations de traduction pour Cybergeo - Colette Cauvin-Reymond

Afin de connaître l'impact éventuel des efforts effectués pour les traductions des articles et des métadonnées, une étude des statistiques de fréquentation de la revue a été menée. À partir des données disponibles générées²⁰ par le serveur d'hébergement du site, elle porte sur la variation temporelle des consultations de 2016 à 2020 ainsi que sur les pays les plus assidus.

Les données retenues sont au nombre de deux : le nombre de visites effectuées et le nombre de pages consultées ; elles ont été utilisées soit brutes, soit transformées (indices, taux, etc.) et ont été traitées afin de dégager des tendances générales et des groupements de pays ayant des comportements de visites similaires. Les procédures²¹ adoptées ont permis dans un premier temps de mettre en évidence les caractéristiques de la fréquentation annuelle et mensuelle du site avant de mettre l'accent dans une seconde étape sur l'importance respective des différents pays dans la consultation de la revue.

1. Au sujet de la fréquentation globale par mois et année

Au niveau de la fréquentation globale, la description portera sur les tendances d'abord uniquement sur les 5 années prises en compte, ensuite sur les années et les mois simultanément, pour terminer par un complément, les durées de consultation.

Fréquentation globale par année entre 2016 et 2020

À partir du tableau 1 qui regroupe les totaux de visites et de pages de 2016 à 2020 et de la figure associée (figure 1), on constate que les années 2016 et 2020 sont de loin les plus fréquentées en termes de visites, encadrant 3 années nettement moins fréquentées avec un creux au centre en 2018. Les taux de variation confirment ces résultats puisque celui de 2017 à 2018, en particulier, atteint -24.84 %. On observe les mêmes tendances pour le nombre de pages, avec un décalage d'un an, simplement un peu moins marquées au niveau du déficit, mais suivies par une reprise beaucoup plus accentuée en 2020 (taux 2019-2020 de 84.67 %).

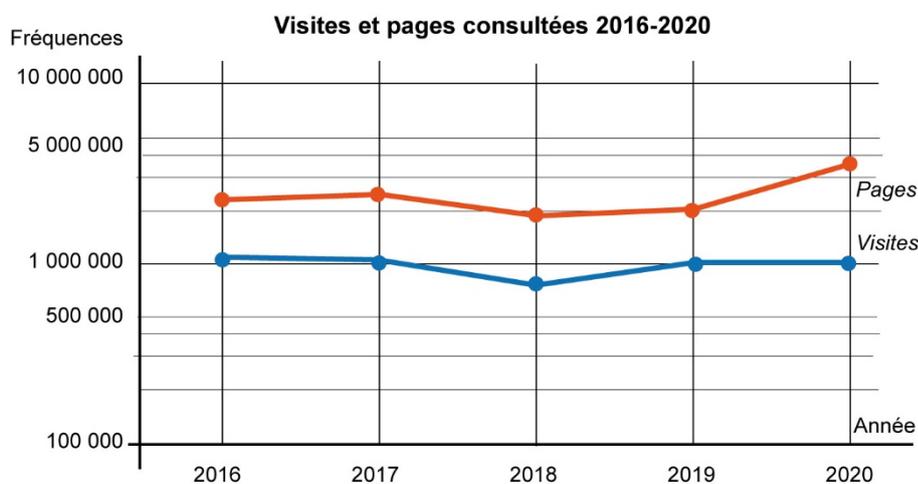
²⁰ Dans le cadre des revues gérées par OpenEdition, des statistiques sont recueillies et calculées à partir des fichiers générés par le serveur d'hébergement des sites, grâce à l'outil d'analyse AWstats. La description des indices établis peut être consultée sur le web aux références électroniques suivantes :
- « Les statistiques de fréquentation », *La maison des revues* [En ligne], Des outils pour *Journals*, Statistiques, mis en ligne le 10 avril 2017, consulté le 07 mai 2021. URL : <http://www.maisondesrevues.org/379>
- « Les statistiques d'usages », *La maison des revues* [En ligne], Des outils pour *Journals*, Statistiques, mis en ligne le 10 avril 2017, consulté le 07 mai 2021. URL : <http://www.maisondesrevues.org/381>

²¹ Les transformations et traitements sont décrits en détail dans la première partie du rapport général sur l'impact qui se trouve en annexe 12.

Tableau 1 : Visites, pages, pays par année

Années	Visites	Pages	Visite_Tx variation %	Pages_Tx variation %	Nb pays	Nbpage/pays
A2016	1 036 170	2 223 833			229	9711
A2017	979 844	2 350 573	-5,44	5,70	233	10088
A2018	736 403	1 858 466	-24,84	-20,94	227	8187
A2019	976 476	1 918 937	32,60	3,25	243	7897
A2020	1 035 805	3 543 723	6,08	84,67	245	14464

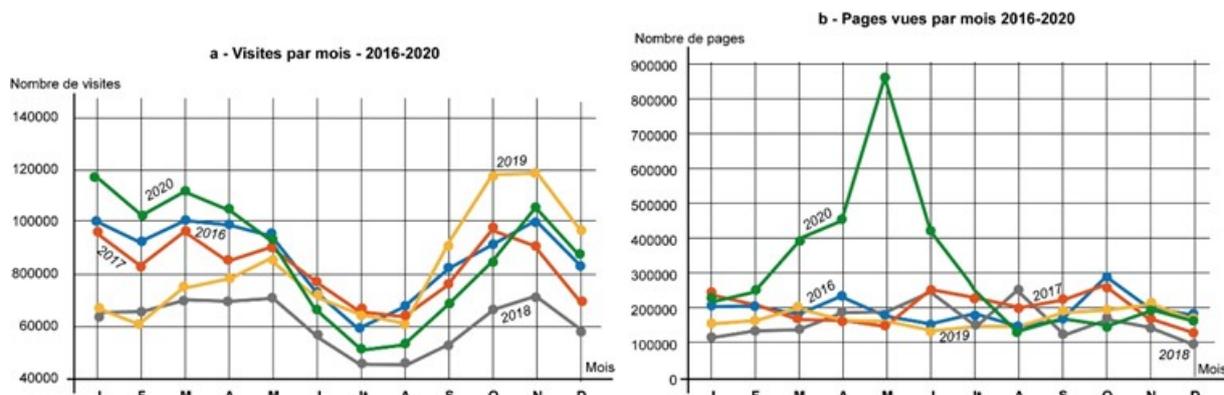
Figure 1 : Visites, pages consultées et pays 2016-2020



Fréquentation annuelle et mensuelle entre 2016 et 2020

Au niveau des fréquences des visites, les variations mensuelles sont très semblables quelle que soit l'année, avec un départ haut, un creux en été, une reprise à l'automne. Les pointes, pour les visites correspondent aux périodes de travail intensif classique (mars-mai) et les baisses aux mois d'examen et aux temps des vacances.

Figure 2 : Visites (3a) et pages (3b) : fréquences mensuelles

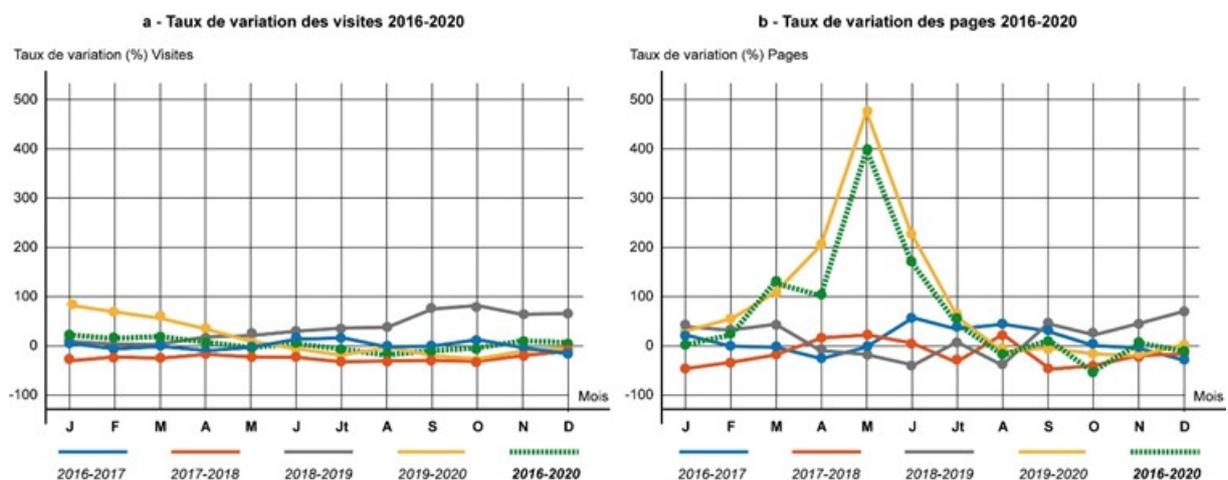


Pour les pages, les variations sont peu marquées mais irrégulières et les années ne montrent pas de grandes différences si ce n'est 2020, avec une montée plus qu'accentuée de mars à mai, confirmée par le rapport du nombre de pages par visite. Ceci traduit, sans doute, le rôle du confinement en raison de la covid 19.

Les taux externes valident les résultats précédents mais mettent mieux en relief, en les accentuant, les petites variations déjà notées pour les pages. Quant aux taux internes, ils ne présentent pas de fortes amplitudes pour les visites contrairement aux pages révélant beaucoup plus d'irrégularités et renforçant encore le comportement particulier de 2020. Concrètement ces deux types de taux n'apportent pas de grosses informations supplémentaires.

Le taux de variation apparaît plus riche (figure 3). La courbe des visites est d'abord assez plate pour 2016-2018, entièrement en négatif, témoignant ainsi d'une tendance déficitaire d'une année sur l'autre. Inversement, 2019 (courbe 2015-2019) se distingue des années précédentes par un taux croissant (proche de 80 %) produisant presque un mamelon en septembre octobre. L'année 2020 repart de ce point haut mais descend régulièrement jusqu'en octobre où elle remonte un peu tout en restant en négatif.

Figure 3 : Visites (3a) et pages (3b) : taux de variation mensuel et annuel



Les courbes des pages ressemblent fortement à ce qui a été trouvé jusqu'à présent avec des variations faibles de part et d'autre de la ligne de changement de signe, et des nuances légères et irrégulières. L'exception de 2020 se constate ici aussi avec une pointe plus que marquée en mai. La courbe pour les 5 années est, quant à elle, quasiment parallèle à celle de 2020 mais avec un pic légèrement moins accentué. Des données pour 2021 seraient importantes pour confirmer ou non ces conclusions.

Un complément : l'analyse des durées de consultation

Cette mesure permet d'analyser les visites en fonction de leur durée par mois et par année. Une majorité des consultations est inférieure à une minute, mais aucune information n'est disponible sur son contenu.

Globalement, à partir du tableau 2, on constate que l'année 2018 est l'année où les durées ont été les plus faibles, et que les deux années extrêmes sont au contraire, celles où les durées ont été les plus élevées. La durée la plus courante est de 0 à 30 secondes, la moins fréquentée est celle d'une heure. Pour les 3 premières catégories l'année 2020 est toujours en tête, comme si en 2020 les visites avaient été plus nombreuses pour la consultation de la revue selon une durée de moins de 5 minutes ; pour les autres catégories, c'est l'année 2016 qui témoigne d'une plus grande fréquentation.

Tableau 2 : fréquences par catégories de durée et par années

Années/ Catégories	0s-30s	30s-2mn	2mn-5mn	5mn-15mn	15mn-30mn	30mn-1h	1h et +	Total
année 2016	870 493	33 570	22 609	29 616	22 992	31 614	25 276	1 036 170
année 2017	841 817	25 231	19 061	25 627	19 447	26 452	22 209	979 844
année 2018	631 054	22 884	16 778	21 616	15 683	19 699	8 689	736 403
année 2019	858 300	28 796	19 207	24 467	17 463	20 417	7 826	976 476
année 2020	880 016	50 355	29 884	28 834	18 374	21 372	6 970	1 035 805
Total	4 081 680	160 836	107 539	130 160	93 959	119 554	70 970	4 764 698

Mais il est difficile d'aller plus loin car, à partir de ces données, il semble impossible de savoir quels sont les visiteurs. Une analyse détaillée des irrégularités mensuelles par catégorie n'apporte pas d'informations spécifiques au niveau de l'impact des traductions sur la revue. Pour aller plus loin un approfondissement par pays s'avère indispensable.

2. Composante géographique de la fréquentation

Pour la fréquentation selon les pays²², seule la variable "pages" est disponible. Le nombre de pays ayant consulté le site est de l'ordre de 200-245 selon les mois et les années (tableau 1), mais seuls les 30 premiers d'entre eux, pour chaque mois et chaque année, ont été retenus car ils "détiennent" au total plus de 90 % de l'information. En combinant les 30 premiers pays de chaque catégorie, 43 pays ont finalement fait l'objet d'une étude approfondie.

En analysant le tableau 1, on constate que le nombre de pages par pays fait apparaître des variations tout à fait similaires à celles du nombre global de pages par année avec un léger décalage temporel des hautes valeurs se rencontrant plutôt en 2017 qu'en 2016 et une baisse accentuée plus marquée en 2019 ; la remontée en 2020 est, quant à elle, plus prononcée puisque ce nombre double presque.

Une approche d'une part des quantités de pages regardées par pays, par année et par mois et, d'autre part, des rythmes de consultation correspondant, permettront de mieux mettre en évidence les pays, et donc les langues associées, vraiment concernés par la revue.

²² Le terme pays a été retenu ; il correspond aux entités spatiales identifiées par les statistiques qui ne sont pas nécessairement des États reconnus au sens strict. Ceci explique que le nombre de pays puisse atteindre 250.

Nombre de pages consultées par pays et par année : une typologie nette des entités spatiales sur les 5 années

Le total des pages consultées par pays, par année et sur les 5 années fait apparaître des constances que révèlent bien les cartes associées. 24 pays en tête du classement sont présents tous les ans et retiennent à eux seuls plus de 92 % de l'information ; une exception, la Tchéquie, qui, malgré un total important de pages vues la plaçant au 4^e rang parmi les 24, n'est présente qu'en 2018 et 2020. En étudiant le total des consultations sur les 5 années, une typologie évidente apparaît qui distingue 4 grands types²³ (carte 1).

Type 4 : Pays de tête avec plus d'un million de pages consultées sur 5 ans et un accroissement plus ou moins accentué en 2020²⁴

Trois pays se partagent toujours les premières places, leur position variant de 1 à 4 : la France, l'Allemagne et les États-Unis. Leurs lecteurs consultent respectivement de 2016 à 2020, 75 %, 73 %, 69 %, 60 % et 49 % du nombre des pages. On constate que la part de la France diminue progressivement (53 à 15 %) alors qu'inversement celle de l'Allemagne croît fortement passant de 7 à 27 % en 2020. Pour les USA, après un démarrage élevé et accentué l'année suivante, sa part retombe autour de 7 %.

Le 4^e pays est la Tchéquie dont le comportement est très surprenant : quasiment invisible jusqu'en 2018 où elle pointe au 4^e rang, elle recule en 2019 (46^e) pour prendre la 1^{ère} place en 2020 avec 29 % des pages de cette année.

Type 3 : Pays avec plus de 100 000 pages consultées sur 5 ans et des remontées plus ou moins prononcées en 2020 à l'exception de la Russie.

L'importance quantitative de ce type est évidente puisque, avec le 12^e pays, on atteint déjà 86 % des pages regardées entre 2016 et 2020. En tête le Canada, la Chine et l'Algérie montrent tous une baisse en 2018 ou 2019 mais une remontée en 2020 ; cette dernière est bien plus sensible pour la Chine qui double ce nombre de pages vues entre 2019 et 2020. *On peut supposer que la traduction des métadonnées en chinois à partir de fin 2018 ait pu jouer un rôle.*

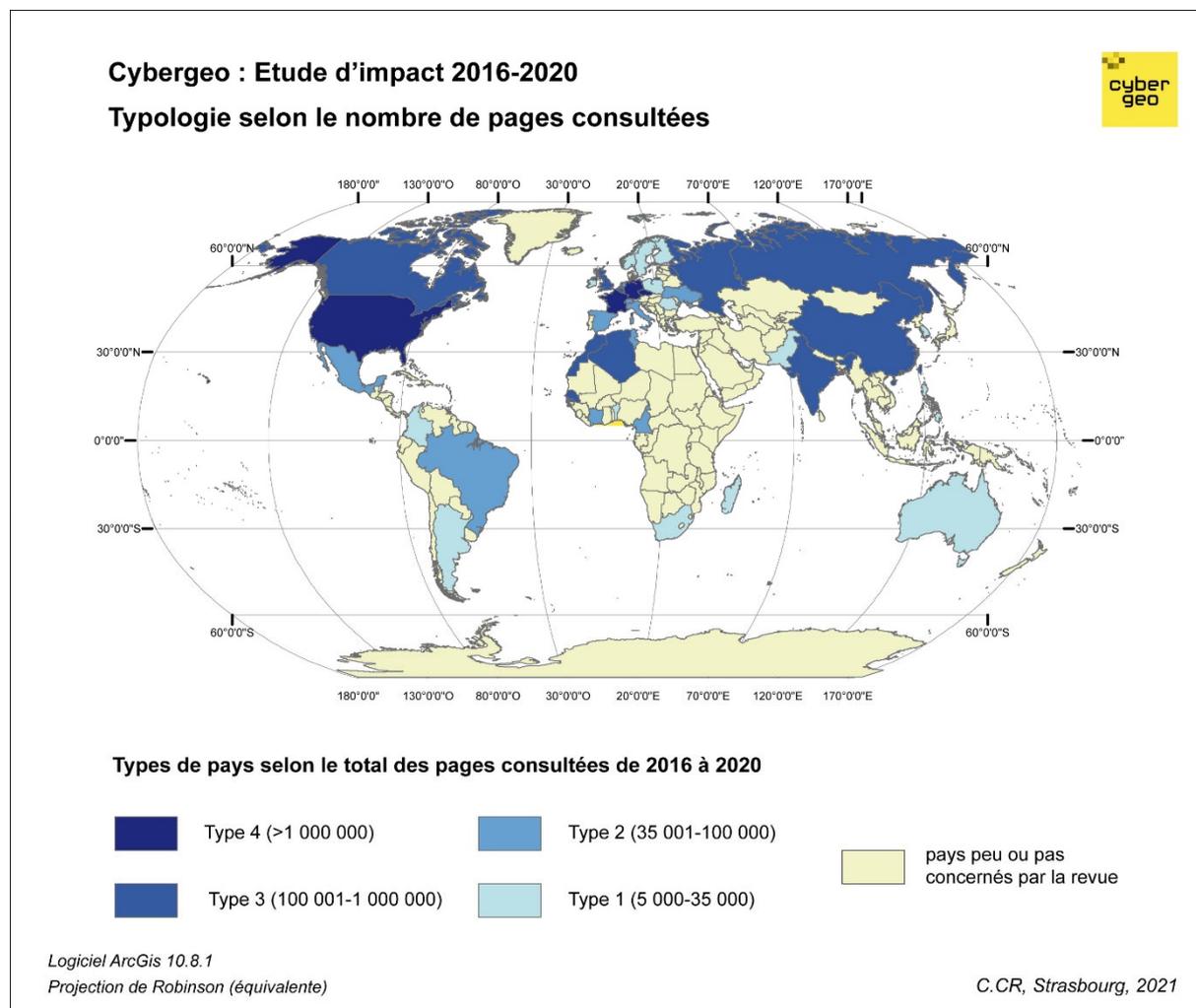
De l'Algérie à l'Inde, ce groupe comprend des grands pays comme la Russie, l'Inde, des pays européens tels la Grande Bretagne et la Belgique et des pays d'Afrique du Nord et de l'Ouest : le Maroc et le Sénégal. Le rythme de la Russie décroît nettement passant de 48 000 pages par an à 11 427. La Grande Bretagne et la Belgique sont relativement stables (autour des 20 000 pages par an), le Maroc (21 492 - 43 564) et le Sénégal (11 436 - 32 993) remontent à partir de 2019, et surtout en 2020. L'Inde, quant à elle, augmente fortement son nombre de

²³ Le terme "type" sera réservé à cette première classification (nommée ici, typologie) établie en fonction du nombre total de pages vues pour les 5 années retenues, les pays étant alors ordonnés dans l'ordre décroissant correspondant.

²⁴ Notons que la rubrique "inconnu", qui apparaît à partir de 2018, appartient à ce groupe avec une place variant entre la 3^e et la 4^e.

consultations (de 9 209 en 2018 à 43 409 en 2020) : *là aussi on peut supposer que les traductions ont accentué cet accroissement.*

Carte 1 : Répartition des types de pays déterminés par le nombre de pages consultées de 2016 à 2020



Type 2 : Pays sans grandes variations interannuelles avec cependant des pays présents toutes les années.

Cet ensemble est formé de pays d'Europe (Pays-Bas, Suisse, Ukraine, Espagne, Italie + l'Union européenne), deux pays d'Afrique (Cameroun et Côte d'Ivoire) et des pays d'Amérique latine, le Brésil et le Mexique. Présents toutes les années, ces pays consultent entre 35 000 et 100 000 pages par an sans qu'il y ait de fortes variations à l'exception du Mexique qui remonte en 2020 (traduction en espagnol ?) et les Pays-Bas qui diminuent nettement.

Type 1 : Pays présents de manière intermittente, parfois même seulement une année, avec plus de 5 000 pages consultées sur 5 ans

Géographiquement, on rencontre des pays de tous les continents : l'Afrique (Bénin) et pays proches avec Madagascar (+ *African Regional Property Organization*) et la Réunion ; l'Asie avec Singapour, les Philippines, le Pakistan et la Corée du Sud ; l'Amérique avec la Colombie et l'Argentine + Haïti ; et pour l'Europe des pays d'Europe de l'Est (Roumanie et Pologne), trois

pays du Nord (Norvège, Suède, Finlande) + l'Irlande ; enfin l'Australie. Tous ces pays ont regardé, pour les 5 années, plus de 5 000 pages.

Singapour, la Colombie, le Bénin, consultent des pages à partir de 2018 ou 2019, au contraire des pays du Nord qui ne ressortent qu'aux années entre 2016 et 2018. *Les pays d'Amérique latine comme Haïti noté en 2018 et 2020, la Colombie en 2019 et 2020, et l'Argentine qui n'est visible qu'en 2020, semblent donc se détacher en fin de la période étudiée, peut-être en relation avec les traductions en espagnol métadonnées commencées fin 2018, et faites systématiquement depuis janvier 2020.* Les pays qui ne se sont manifestés qu'une année sont répartis de manière tout à fait aléatoire sur les 5 années et se rencontrent sur tous les continents.

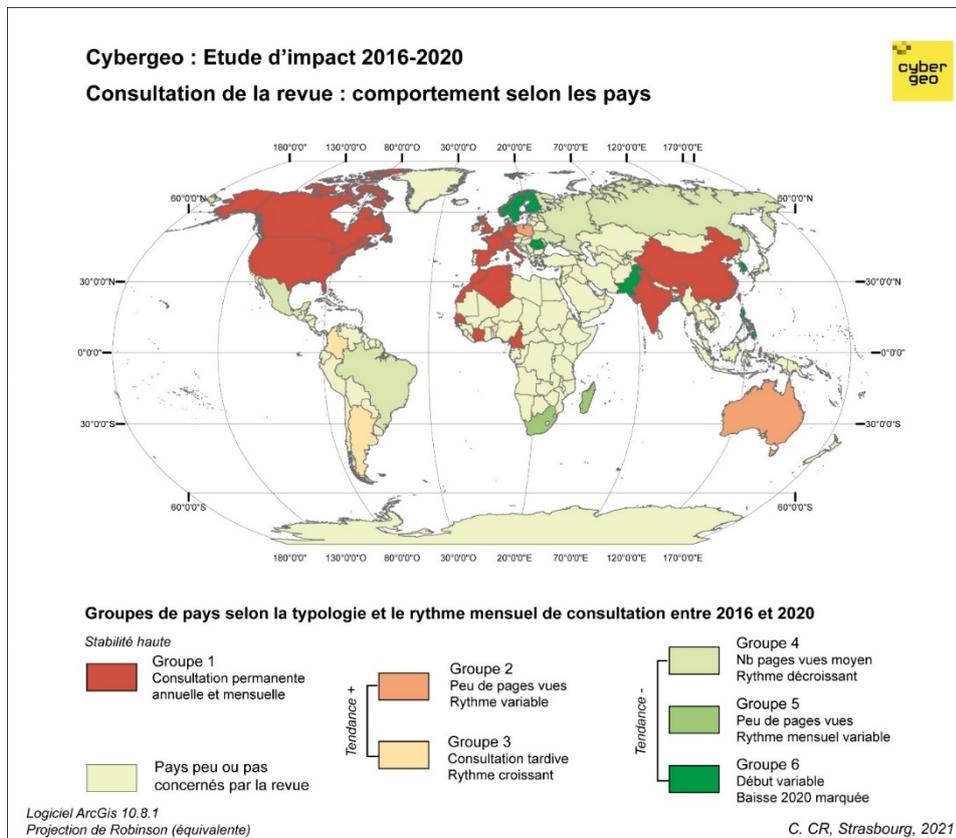
Cette première typologie, intéressante certes, peut être affinée et précisée en abordant les rythmes annuels et mensuels de consultation du site.

Une classification appuyée sur les rythmes annuels et mensuels (analyse du total des pages consultées par pays, par année et par mois).

Un tableau, qui intègre la présence de chaque pays pour chaque mois de l'année, leur rythme de présence par mois (classes) et leur place dans la typologie (types), a servi de base pour rechercher les groupements de pays ayant un même comportement de consultation. Après divers traitements, un scalogramme a abouti à l'identification de 6 groupes²⁵ visualisés sur la carte 2.

²⁵ Au niveau de la classification finale, on retiendra le terme de "groupe" pour identifier les différents sous-ensembles reconnus. Pour les associations concernant seulement les rythmes, on parlera de classe.

Carte 2 : Classification selon le nombre de pages regardées sur 5 ans et les rythmes mensuels et annuels de consultation (Attention ! À l'échelle mondiale, la Réunion et Singapour ne se voient pas).



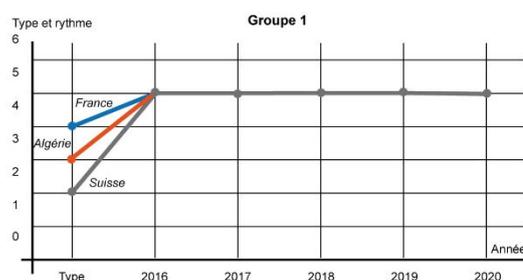
Le premier groupe montre une stabilité haute, soit une présence tous les mois de toutes les années ; deux groupes, avec des tendances initiales variées, révèlent plutôt des tendances à l'accroissement de leur présence vers 2020, et trois autres ensembles témoignent au contraire d'une propension à la baisse pour la même année.

Groupe 1 : Présence permanente annuelle et mensuelle

Ce groupe correspond à des pays présents les 5 années et tous les mois de ces années, indépendamment de tout rythme mensuel. Au nombre de 17, ils appartiennent tous aux pays de la typologie globale qui étaient présents toutes les années dans les types 4, 3, 2 ne se différenciant que par l'importance du nombre de pages consultées. On retrouve tous les pays de la catégorie haute (type 4) à l'exception de la Tchèque, et de la catégorie suivante (type 3), moins la Russie. Quant aux pays de la 3^e catégorie (type 2), le groupe 1 n'en conserve que 6 : les 5 pays "exclus" diffèrent de ce premier groupe par leur rythme mensuel qui ne témoigne pas d'une présence permanente.

Figures 4a et 4b : Groupe 1 : tableau (4a) et profil (4b)

Groupe	Nom	Type	2016	2017	2018	2019	2020
1	1-France	4	5	5	5	5	5
1	2-Germany-	4	5	5	5	5	5
1	3-United States of Ame	4	5	5	5	5	5
1	5-Canada	3	5	5	5	5	5
1	6-Chine	3	5	5	5	5	5
1	7-Algérie	3	5	5	5	5	5
1	9-Morocco	3	5	5	5	5	5
1	10-Great Britain	3	5	5	5	5	5
1	11-Belgique	3	5	5	5	5	5
1	12-Senegal	3	5	5	5	5	5
1	13-Inde-	3	5	5	5	5	5
1	14-Tunisie	2	5	5	5	5	5
1	16-Suisse	2	5	5	5	5	5
1	17-Cameroun	2	5	5	5	5	5
1	18-Côte d'Ivoire	2	5	5	5	5	5
1	20-Espagne	2	5	5	5	5	5
1	22-Italie	2	5	5	5	5	5



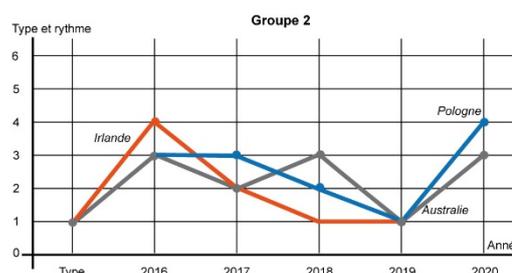
Ce bloc de 17 pays, qui révèlent tous un accroissement plus ou moins marqué en 2020, comprend les pays d'Europe qui entourent la France, le Luxembourg non compris, des grandes puissances (USA, Canada, Chine, Inde) à l'exception de la Russie, et des pays d'Afrique du Nord et de l'Ouest (Algérie, Tunisie, Maroc, Sénégal, Cameroun, Côte d'Ivoire). Ils dessinent sur la carte (carte 2) une bande Est-Ouest légèrement arquée vers le Sud, interrompue entre l'Inde et l'Europe de l'Est. **Ces pays constituent en quelque sorte le bloc de base des pays consultant régulièrement des pages et, par suite, visitant régulièrement le site. Ils peuvent être concernés par les traductions des articles et des métadonnées tant en anglais qu'en espagnol ou en chinois.**

Groupes à tendance "montante" vers 2020 (groupes 2 et 3)

Le groupe 2 (figure 5) est formé de 3 pays, différents politiquement et économiquement, très distants spatialement : la Pologne, l'Irlande, l'Australie. Ces trois pays appartiennent à la partie basse (1) de la typologie en termes de nombre de pages regardées (entre 5 000 et 35 000 sur 5 ans), et se caractérisent par des variations du rythme mensuel. Bien présents en 2016, ils voient leur occurrence mensuelle baisser plus ou moins régulièrement passant de 10-11 mois par année à 1 ou 2 mois en 2019 (classe 1 pour tous), pour revenir, l'année 2020, à une présence mensuelle presque régulière (classes 2, 3 et même 4). De plus, cette dernière année, l'Irlande et l'Australie, la Pologne à un moindre degré, consultent un plus grand nombre de pages. **Pour l'Irlande et l'Australie, les traductions en anglais ont pu jouer un rôle.**

Figures 5a et 5b : Groupe 2 : tableau (5a) et profil (5b)

Groupe	Nom	Type	2016	2017	2018	2019	2020
2	38-Poland	1	3	3	2	1	4
2	40-Irlande	1	4	2	1	1	3
2	43-Australie	1	3	2	3	1	3

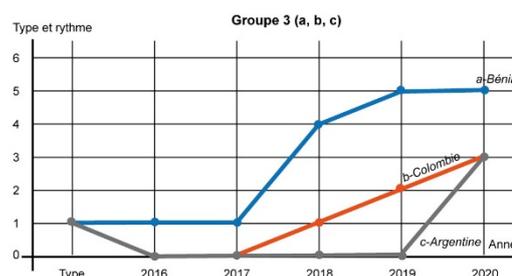


Le groupe 3 (figure 6) comprend trois sous-ensembles qui révèlent chacun une remontée en 2020, à l'exception de Singapour. Formé de 6 pays, ce groupe s'identifie principalement par une situation de consultation nulle ou quasi nulle (classes 0 et 1) les premières années, c'est-à-dire que les pays concernés ne regardaient pas du tout le site ou seulement 1 ou 2 mois par an. Des changements apparaissent soit en 2018 (Bénin, Haïti, Tchèque), soit en 2019 (Colombie, Singapour) ou même seulement en 2020 (Argentine), qui introduisent des différences justifiant les sous-ensembles selon l'importance de la fréquence de consultation et sa précocité. Le nombre de pages consultées confirme une certaine amélioration, plus ou moins accentuée, entre 2018 et 2020, mais plus marquée en 2020 (excepté pour Singapour).

Le Bénin et Haïti, formant le premier sous-groupe (3a), débutent avec une présence d'un ou deux mois par an qui s'accroît par la suite, nettement pour le Bénin qui devient présent tous les mois, un peu moins pour Haïti. La Tchèque et la Colombie, deuxième sous-groupe (3b), sont absentes en 2016 et 2017 ; en 2018 et 2019, on note une apparition irrégulière et en 2020, ces deux pays atteignent les classes 3 ou 4. Ceci témoigne d'une présence sur le site qui se renforce. L'Argentine et Singapour (sous-groupe 3c) partent de zéro, et ne se manifestent ensuite qu'en 2019 pour Singapour et 2020 pour l'Argentine en classe 3. Singapour semblerait même baisser un peu en 2020 contrairement aux autres pays.

Figures 6a et 6b : Groupe 3 : tableau (6a) et profil (6b)

Groupe	Nom	Type	2016	2017	2018	2019	2020
3a	28-Bénin-	1	1	1	4	5	5
3a	32-Haïti	1	1	2	5	3	3
3b	4-Czech Rep.	4	0	0	3	1	4
3b	27-Colombia	1	0	0	1	2	3
3c	25-Singapore	1	0	0	0	3	1
3c	39-Argentine	1	0	0	0	0	3



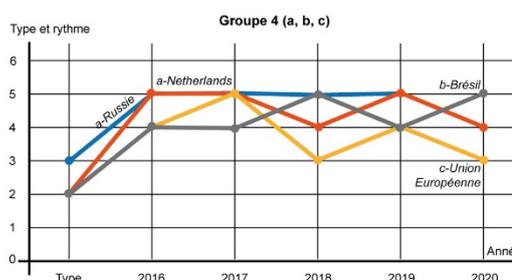
La présence accrue de l'Argentine et de la Colombie en 2020 peut s'expliquer par la traduction des métadonnées en espagnol. Pour les autres pays, il est difficile d'affirmer que les traductions ont joué un rôle particulier. Quant à la Tchèque son rythme est assez étonnant, comme cela a été souligné à plusieurs reprises, compte tenu de sa place (4) dans la typologie générale du nombre de pages regardées pendant les 5 années étudiées. **Mais le trait essentiel qui caractérise ce groupe est un démarrage très lent, suivi d'une présence de plus en plus marquée en 2020.**

Groupes tendant à une baisse de la consultation en 2020 (groupes 4, 5, 6)

Le groupe 4 (figure 7) correspond, à l'exception de la Russie, aux pays avec le nombre de pages le plus faible parmi les pays présents toutes les années. Ce groupe est identifié par le fait que, bien que présents toutes les années, les 6 pays qui le composent ne ressortent pas nécessairement tous les mois ; le rythme mensuel est varié passant de la classe 5 à 3 et même 1. À l'exception du Brésil et du Mexique qui voient leur rythme s'améliorer légèrement en 2020 et forment un premier sous-groupe, les autres pays ont une tendance à la baisse pour la même période. Ces pays manifestent une présence mensuelle forte (classes 4 ou 5) en 2016, 2017, encore en 2018 et 2019, et une tendance à une diminution plus ou moins prononcée en 2020, mais une baisse pour l'Union européenne et l'Ukraine dissocie ces 2 pays en un 3^e sous-groupe. Ces fléchissements ressortent également dans les tableaux des totaux annuels, avec la même exception pour le Brésil et le Mexique. **On peut penser que les traductions en espagnol ont joué pour ces deux pays (même si le Brésil est de langue portugaise, mais les traductions en anglais ont pu aussi intervenir).**

Figures 7a et 7b : Groupe 4 : tableau (7a) et profil (7b)

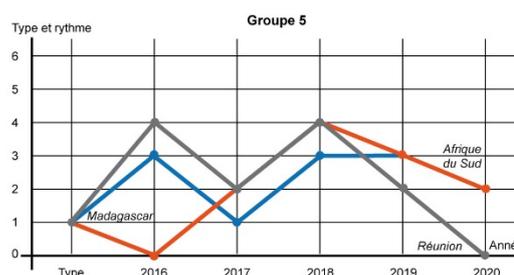
Groupe	Nom	Type	2016	2017	2018	2019	2020
4a	8-Russian Federation	3	5	5	5	5	4
4a	15-Netherlands	2	5	5	4	5	4
4b	23-Brésil	2	4	4	5	4	5
4b	24-Mexico	2	4	3	4	5	5
4c	19-Ukraine	2	4	4	4	3	1
4c	21-European Union	2	4	5	3	4	3



Le groupe 5 (figure 8), en bas du classement par le nombre de pages, rassemble trois pays (Madagascar, la Réunion et l'Afrique du Sud) caractérisés surtout par des rythmes mensuels irréguliers mais similaires, même si le départ en 2016 est différent. On constate un début assez haut (4 et 3) pour Madagascar et la Réunion en 2016, l'Afrique du Sud ne commençant à consulter qu'en 2017. À cette date, les trois pays se retrouvent en position 1 ou 2, traduisant une baisse pour les deux îles mais un départ pour l'Afrique du Sud. 2018 est l'année de la consultation la plus importante atteignant même la classe 4, donc avec une présence presque tous les mois de l'année. Cette montée ne semble pas se poursuivre puisque les 3 pays appartiennent ensuite à des classes de faible présence (2 et même 0). En fait ces pays n'ont guère que 2 années de "vraie" fréquentation, ce que confirment les totaux annuels des pages consultées. **Les traductions en anglais ont pu toucher l'Afrique du Sud mais sans qu'elles aient eu un effet à long terme.**

Figures 8a et 8b : Groupe 5 : tableau (8a) et profil (8b)

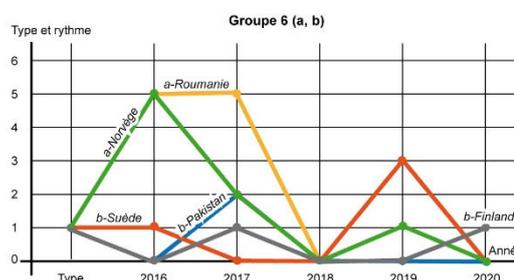
Groupe	Nom	Type	2016	2017	2018	2019	2020
5	33-Madagascar	1	3	1	3	3	2
5	30-Reunion	1	4	2	4	2	0
5	37-Afrique du Sud	1	0	2	4	3	2



Le groupe 6 (figure 9) fait ressortir des pays dont la présence sur le site diminue ou disparaît à partir de 2018 ou 2019. On peut la diviser en deux sous-ensembles en fonction de leur place en 2016 et 2017. Le premier sous-ensemble (6a) part d'une bonne présence très régulière en 2016 (classe 5) qui s'atténue de plus en plus en plus surtout à partir de 2019 pour disparaître totalement en 2020 (Roumanie et Norvège). Le deuxième (6b) est quasiment toujours dans les classes "absence", "présence un ou deux mois par an", et "présence irrégulière". Les pays de ce groupe ne regardent que très peu la revue, même si on peut constater dans certains cas une légère remontée en 2020 (Philippines, Corée du Sud, Suède, Finlande, Pakistan). **On note, cependant, que tous les pays d'Europe du Nord se trouvent dans ce groupe et donc qu'ils consulteraient très peu (!) la revue, même avec les traductions en anglais.**

Figures 9a et 9b : Groupe 6 : tableau (9a) et profil (9b)

Groupe	Nom	Type	2016	2017	2018	2019	2020
6a	26-Roumanie	1	5	5	0	1	0
6a	31-Norway	1	5	2	0	1	0
6b	35-Philippines	1	1	3	1	0	2
6b	42-Korea, South	1	2	2	0	1	1
6b	41-Pakistan	1	0	2	0	0	0
6b	34-Suede	1	1	0	0	3	0
6b	36-Finland	1	0	1	0	0	1



3. Conclusion

Cette approche de l'impact des traductions d'articles et des métadonnées sur la consultation de la revue conduit à faire un certain nombre de constats qui demanderaient à être confirmés en prolongeant cette étude au moins dans les deux années à venir pour être plus certain des tendances qui semblent se dégager.

Le premier fait à souligner est le suivant : à partir de 2016, année avec des chiffres élevés, la fréquentation et la consultation de la revue baissent jusqu'en 2018-2019, et remontent ensuite nettement en 2020 ; ceci se vérifie que l'on analyse les visites ou les pages, par année ou par mois, ou que l'on prenne en compte les fréquences ou les différents taux.

Les variations mensuelles se révèlent assez similaires entre les années, avec des pointes correspondant aux périodes de travail intensif classiques (mars-mai) et des creux aux vacances et aux temps d'examen. Une exception pour les pages en 2020 : les mois d'avril-mai dessinent

un pic très marqué, repris partiellement en octobre-novembre. On peut supposer que la covid 19 ait joué un rôle.

Le second fait concerne les pays et leur consultation globale. Le nombre total de pages visitées fait apparaître 4 types de pays distincts.

Le premier avec 3 pays d'Europe (France, Allemagne, Tchéquie) et les USA, consulte sur 5 ans plus d'1 million de pages et la question *de la traduction intervient sûrement pour les USA et sans doute aussi pour les pays d'Europe cités précédemment.*

Le deuxième type avec les grands pays du monde (Canada, Russie, Inde, Chine, Grande Bretagne) et des pays d'Afrique du Nord et de l'Ouest, regardent encore bien la revue avec plus de 100 000 pages regardées sur 5 ans et ont tendance à être stables selon les années ou à remonter en 2020 comme l'Inde. Seule la Russie manifeste une baisse. *Pour les grands pays, les traductions soit en anglais soit en chinois sont sans doute importantes.*

Le troisième ensemble, avec des consultations encore supérieures à 5000, regroupe des pays contigus à la France (PB, Suisse, Espagne, Italie), l'Ukraine, deux pays d'Afrique (Côte d'Ivoire et Cameroun) et des pays d'Amérique latine (Brésil, Mexique). Seuls les Pays-Bas réduisent leur présence. *L'espagnol peut intervenir pour le Mexique et l'Espagne.*

Le dernier type est très éparpillé et comprend d'une part les pays du Nord, peu concernés semble-t-il malgré les traductions en anglais qui ont commencé tôt, d'autre part des pays pour qui les traductions en anglais ou en espagnol peuvent être importantes comme l'Australie, la Corée du Sud, l'Afrique du Sud, l'Irlande, le Pakistan ou l'Argentine et la Colombie.

Mais en s'arrêtant à 2020, il n'est possible, en tenant compte seulement du nombre de pages, que d'indiquer des tendances.

L'analyse des rythmes mensuels, jointe aux résultats précédents, permet d'approfondir ces conclusions tout en restant sur des tendances.

Un groupe stable, permanent, de pays constitue le socle des lecteurs de la revue, formant une bande est-ouest partant des US et du Canada auxquels s'associe une partie de l'Europe de l'Ouest et de l'Afrique du Nord et l'Ouest, suivie d'un grand vide spatial avant d'être rejointe par l'Inde et la Chine. *Pour certains de ces pays, l'anglais et le chinois sont importants.*

Les groupes 2 et 3 témoignent d'une tendance à une amélioration en 2020 et les consultations plus régulières peuvent être liées à l'anglais bien sûr (Irlande, Australie, Singapour) mais aussi à l'espagnol (Colombie, Argentine). Pour la Pologne, le Bénin, la Tchéquie, il est impossible de conclure sur ce point.

Les trois autres groupes ont un comportement inverse puisqu'ils tendent à regarder de moins en moins le site. Il s'agit, plus particulièrement pour la classe la moins consommatrice de pages, des pays de l'Europe du Nord, du Pakistan, de la Corée du Sud, des Philippines et de la

Roumanie : *les traductions n'ont pas encore en 2020 joué un rôle*. Les autres pays des groupes 4 et 5 présentent les mêmes particularités mais moins accentuées.

En conclusion, plusieurs résultats soulignent le rôle probable des traductions dans les variations de consultation de la revue, mais très inégalement selon les pays du monde. Les traductions en anglais attirent bien sûr les pays anglophones (Inde, Asie du SE, Australie, etc.) ainsi que certains pays d'Europe de l'Est. Les métadonnées en espagnol concernent surtout des pays d'Amérique latine (Argentine, Colombie) et les textes en chinois semblent avoir un impact réel sur la consultation en Chine. Cependant il serait important de continuer pendant 1 an ou 2 l'analyse des données de visite du site pour savoir si les pays se référant fréquemment à la revue demeurent fidèles et si les pays qui ne semblent pas touchés par la revue conservent leur attitude actuelle ou la modifient.

Impact des traductions pour ViaTourism

Le numéro 16/2019, a comporté pour la première fois des traductions en chinois et en arabe. Publier les articles dans ces nouvelles deux langues a permis d'observer à travers les statistiques si le lectorat est amplifié dans les pays arabes et la Chine. De manière générale, l'examen comparatif des statistiques entre les années 2019 et 2020 (de septembre à septembre), révèle une augmentation significative des visites du site depuis la Chine (1 772 pour 2019, 5 085 pour 2020). Après la sortie du numéro 16/2019, le nombre de visites a connu un rebond notable. Dès le mois d'avril 2020, 449 visites sont enregistrées, à mettre en perspective avec les 62 visites d'avril 2019. La tendance se poursuit dans les mois qui suivent, avec un nombre de visites atteignant 1 967 au mois de mai (pour 50 visites l'année précédente). Sur la période comprise entre les mois d'avril et septembre, le nombre total de visites passe de 499 en 2019, à 4 270 en 2020.

Concernant les pays arabophones, le cas de l'Égypte se distingue des autres pays par une augmentation conséquente des visites sur la période de référence mise en comparaison (avril-septembre) : alors que le nombre total des visites pour cette période en 2019 atteignait le chiffre de 238, il est de 775 en 2020. En particulier, on note un rebond sur la période d'avril à juin : 502 visites en 2020, pour 114 en 2019.

Conclusion Générale

Plusieurs résultats intéressants émergent de la mise en commun des problèmes de traduction rencontrés par des revues à petit budget spécialisées en géographie. Le résultat le plus important est que l'opération de traduction exige toujours un contrôle humain, en raison précisément de la spécificité du vocabulaire scientifique. Le risque d'erreur se renouvelle avec l'émergence de nouveaux concepts ou d'expressions détournées de leur usage commun pour construire et affiner les notions travaillées en sciences humaines.

Cela dit, un deuxième résultat tout aussi important souligne la relative fiabilité des traductions automatiques, qui font gagner un temps précieux aux bénévoles traducteurs. A condition de procéder à une relecture finale, l'opération de traduction se trouve grandement simplifiée par l'emploi de DeepL ou Google translation notamment. La démarche est validée, du français vers l'anglais ou l'espagnol, mais aussi vers le chinois. Une initiative intéressante est de signaler l'intérêt d'une rétro-traduction dans la langue d'origine afin de vérifier que des contresens n'ont pas été introduits.

Un troisième résultat intéressant et non attendu est l'attention que l'exercice de traduction fait porter sur la qualité des métadonnées originales, en obligeant à purger les titres et résumés de leurs incohérences, imprécisions ou formes d'écriture qui seraient des sources d'ambiguïté pour la traduction. La qualité de la communication scientifique en milieu international ressort bien améliorée du fait de cette prise de conscience de l'importance des métadonnées à destination des moteurs de recherche. En revanche, l'alignement qu'elle implique sur des normes rédactionnelles de type anglo-saxon porte le risque d'un appauvrissement certain des subtilités d'écriture qui peuvent se former dans d'autres langues, y compris en anglais !

Un autre résultat intéressant, et à plus long terme de cet exercice, est dans le rapprochement entre les pratiques et l'interconnaissance qui résultent de la collaboration entre des revues francophones. La mise en discussion des problèmes et la mise en commun des solutions, sous forme de guides de bonnes pratiques et de glossaires interactifs multilingues et évolutifs sont des résultats qui montrent la fécondité d'une approche coopérante pour l'édition scientifique.

En ce qui concerne l'impact des traductions des métadonnées et de certains articles sur la consultation de la revue, plusieurs constats sont à souligner. Globalement entre 2016 et 2020, on remarque que la fréquentation du site, partant d'un niveau élevé en 2016 (année de la célébration du 20e anniversaire de la revue), a baissé fortement en 2018 pour remonter nettement en 2020 mais cette reprise n'est pas nécessairement liée aux traductions. L'approche spatiale, plus adaptée à une mise en évidence de l'impact des traductions, a permis d'identifier des pays intéressés par la revue, qui ont progressivement accru leur consultation. Deux classifications, l'une d'ordre quantitatif, l'autre liée aux rythmes des consultations, ont

révélé le rôle fondamental de l'anglais qui attire aussi bien les pays anglophones (Inde, Asie du Sud-Est, Australie, etc.) que des pays d'Europe (Allemagne, Tchéquie, Pologne). Les traductions en chinois ont eu, semble-t-il, un réel impact en Chine, pays qui se trouve au rang 5 en 2020. Les métadonnées en espagnol concernent non seulement l'Espagne mais aussi des pays d'Amérique latine (Argentine, Colombie) ou d'Amérique centrale (Mexique). Cependant ces conclusions doivent être nuancées car les traductions sont encore récentes et il faudrait prolonger cette étude pour savoir, si dans les années à venir, les pays intéressés restent fidèles et si d'autres pays émergent.