

02/11/2021

# CybergeonNet

## 2019-2021

### SYNTHESE DU RAPPORT FINAL

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



# Description du projet

## Acteurs/équipe

Le projet est porté par *Cybergeo : Revue européenne de Géographie* en partenariat avec deux revues internationales *Brussels Studies* et *ViaTourism* et l'encyclopédie numérique de géographie *Hypergeo*.

### Responsable du projet

Christine Kosmopoulos, Ingénieure de recherche au CNRS, responsable éditoriale de *Cybergeo*

### Participants

- Natacha Aveline, Directrice de recherche au CNRS
- Colette Cauvin-Reymond, Professeure d'université à la retraite
- Bernard Elissalde, Professeur émérite de l'Université de Rouen et directeur d'*Hypergeo*
- Maria Gravari-Barbas, Professeure à l'Université Paris 1 Sorbonne et directrice de *ViaTourism*
- Margaux Hardy, Chercheuse, Université Saint-Louis – Bruxelles et éditrice de *Brussels Studies*, Belgique
- Nathalie Lemarchand, Professeure à l'Université de Paris 8 et vice-résidente de l'UGI
- Francisco Maturana, Professeur à l'Université de Santiago, Chili
- Denise Pumain, Professeure émérite à l'Université Paris 1 Sorbonne et directrice scientifique de *Cybergeo*
- Benjamin Wayens, Professeur à l'Université Libre de Bruxelles et rédacteur en chef adjoint de *Brussels Studies*, Belgique
- Gloria Zamorano, Professeure à l'Université de Mendoza, Argentine
- Liubing Xie, traducteur chinois

## Langues et type de corpus (métadonnées, résumé, texte intégral)

Les tests ont porté sur la traduction des métadonnées en anglais, en espagnol et en chinois pour les articles de *Cybergeo*, d'*Hypergeo* et de *ViaTourism*. Des tests de traductions de textes intégraux reproduits dans le rapport final complet ont aussi été entrepris en espagnol et en anglais pour des articles de *Cybergeo*, ainsi qu'en arabe et en chinois pour *ViaTourism*.

## Outils et logiciels employés

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

## Liste des documents produits

- Glossaire cartographique – Colette Cauvin-Reymond
- Glossaire plurilingue sur le tourisme de *ViaTourism*
- Guide de bonnes pratiques pour la rédaction et la traduction des métadonnées
- Index plurilingue *Cybergeo*
- Rapport final sur le projet *CybergeoNet* (64 pages) avec ses annexes (323 pages)

## Déroulement des étapes du projet en précisant :

La première réunion de travail début décembre 2019 a réuni l'ensemble des partenaires. Elle a été l'occasion de structurer l'organisation par groupe de travail selon les cinq axes du projet. Courant 2020, les groupes de travail se sont réunis mensuellement pour discuter de l'avancée des travaux. Fin 2020 un rapport à mi-parcours a été remis. Courant 2021, une étude d'impact a été lancée sur l'amélioration de la qualité des métadonnées et de la traduction d'articles intégraux qui s'est accompagnée de la rédaction du rapport final. Novembre 2021, le rapport détaillé du projet et ses annexes ont été finalisés. Un rapport synthétique de 6 pages a également été rédigé.

## Démarche scientifique suivie

*Le projet se décline selon cinq axes de travail :*

- 1) Améliorer la qualité des traductions des métadonnées des articles déjà publiés dans *Cyberge* en faisant vérifier et corriger les traductions existantes par des spécialistes et un recours ponctuel à des traducteurs professionnels ;
- 2) Evaluer l'apport des méthodes de traduction-automatique 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.

## Méthodologie et résultats

### *Sur l'amélioration des métadonnées (axe 1)*

Un jeune chinois non francophone spécialiste de langue et littérature anglaise, Liubing Xie, a été sollicité pour faire les traductions en **chinois** à partir de la version anglaise des métadonnées (title et abstract) des articles publiés dans *Cyberge*. Natacha Aveline a pris en charge la relecture en chinois pour garantir une restitution fidèle au champ scientifique. Indépendamment de ses connaissances du chinois, elle s'est aussi appuyée sur *Google traduction* afin de comparer les différentes rédactions/traductions français/anglais/chinois. La qualité de restitution du chinois vers le français s'est avérée assez satisfaisante du fait de la proximité de la structure syntaxique du chinois avec celle du français et de l'anglais.

La traduction des métadonnées par Liubing Xie est apparue dans l'ensemble de très bon niveau, 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. Liubing Xie avait restructuré plusieurs phrases mal énoncées dans le texte d'origine au cours du processus de traduction, et dans certains cas N. Aveline s'est appuyée sur sa traduction en chinois pour modifier l'abstract en anglais ! Liubing Xie a d'ailleurs indiqué que, quoique ne parlant pas français, il recourait parfois au texte français en s'aidant de Google Traduction, pour tenter de comprendre ce que l'auteur avait voulu dire dans la version anglaise. Cependant on a pu constater que les résumés en français posaient également problème, de sorte qu'il a fallu remonter à ce niveau pour contrôler la chaîne de traduction.

**Résultats** : Ils ont alors décidé de procéder à l'inverse de la démarche initiale. Les abstracts en français et en anglais (version de l'auteur) étaient d'abord envoyés à N. Aveline pour vérification, au besoin en s'aidant de DeepL<sup>1</sup> pour l'anglais, sachant que nous disposions par ailleurs d'une ultime relecture par un anglophone natif. La version anglaise de l'abstract était alors transmise à Liubing Xie.

### *Sur l'apport des méthodes de traduction automatique pour la traduction scientifique (axe 2)*

La vérification des métadonnées traduites en **anglais** a conduit à observer des problèmes d'ordre structurels qui 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 rédigés par les auteurs des articles publiés. Dans de tels cas, il a été nécessaire de retravailler en profondeur la version française du résumé et parfois les mots-clés.

Cette version a ensuite été soumise à la traduction automatique (TA) en anglais par DeepL, à laquelle ont pu être faits d'éventuels ajustements. A noter que l'ultime relecture faite par notre correctrice anglophone n'a pas présenté de valeur ajoutée significative à la traduction de DeepL sous notre contrôle, attestant du haut niveau atteint aujourd'hui par ces traducteurs automatiques entre le français et l'anglais.

Des problèmes de formulation/style ont pu être détectés selon les langues. 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. Le choix a donc été fait de la mise aux normes des abstracts aux standards anglo-saxons pour les métadonnées. Il convient de préciser que les traducteurs automatiques n'interviennent pas sur ces différences de formulation.

Les études conduites par C. Cauvin sur les métadonnées d'articles comparent 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. Des confusions de nature sémantique sur un terme (power) ont pu être observées qui incitent à faire un retour au français après la traduction anglaise pour éviter des erreurs involontaires.

F. Maturana et G. Zamorano ont procédé selon la même méthode que celle décrite ci-dessus pour l'anglais. Ils ont utilisé Google Traduction (GT) et DeepL pour la traduction **espagnole** des métadonnées des articles publiés. Très concrètement, ils ont entré le texte français et l'ont traduit automatiquement sur Google traduction et sur DeepL. Globalement peu de problèmes ont été rencontrés, toutefois, il a été souvent nécessaire de modifier des mots inadéquats ou de reprendre certaines expressions. Il pouvait également y avoir des choix à faire entre des suggestions de GT et de DeepL. Dans tous les cas, la relecture de la traduction par un spécialiste de la langue et du domaine s'avère indispensable pour garantir une fidèle traduction. Il arrive d'ailleurs que ces traducteurs géographes en espagnol, 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 ici encore 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.

---

<sup>1</sup> 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.

N. Lemarchand et B. Elissalde ont appuyé leurs observations sur la **traduction de textes intégraux** de *Cyberge* et *d'Hypergeo* et comparé dans les deux sens les **traductions anglaises et espagnoles**. Parce que les algorithmes fonctionnent selon un calcul de probabilité de fréquence d'usage, les textes de TA (DeepL et GT) 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 ? À ce jour, les tests montrent bien le risque d'appauvrissement de la langue par la traduction automatique avec un vocabulaire et une syntaxe simplifiés voire trop simplistes pour ce qui concerne des articles entiers.

**Résultats** : en conclusion, la grande qualité atteinte par les logiciels de traduction automatique permet de restituer les métadonnées en français en anglais et en espagnol de manière satisfaisante à condition de procéder à une relecture humaine avant la mise en ligne. L'effort doit aussi porter sur la qualité de du titre et du résumé originaux afin d'éviter les problèmes structurels, les erreurs de formulation et l'usage de faux-amis. En revanche afin de conserver la richesse de la langue, le style, la construction argumentaire il est essentiel que les auteurs s'appliquent à bien vérifier la traduction du texte intégral.

### *Sur la méthode de traduction innovante (axe 3)*

*CybergeNet* 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 pour la traduction des métadonnées au fil de l'eau 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 *Cyberge*, décrites ci-dessous, sur *l'index plurilingue des mots clés de Cyberge* et sur les *glossaires spécialisés en cartographie et sur le tourisme*.

Dans la note aux auteurs publiée sur le site de la revue *Cyberge*, 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 du comité de rédaction et le *guide de bonnes pratiques sur la rédaction et la traduction des métadonnées* est retourné par les auteurs, 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 *Cyberge*, la revue n'ajoute que de manière très exceptionnelle un mot clé. En effet, même si le contenu de la science se renouvelle continuellement, 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 sont examinés pour intégration. Actuellement l'index plurilingue de *Cyberge* comporte près de 2200 mots clés décrivant ses articles.

**Résultats :** *L'index plurilingue des mots clés de Cybergegeo* et les *glossaires spécialisés en cartographie et sur le tourisme* seront disponibles en accès ouvert et pourront s'enrichir de la liste de termes spécifiques acceptés fournis par les auteurs et seront partagés entre les équipes éditoriales.

#### ***Sur la montée en compétence des auteurs et des équipes éditoriales en matière de traduction et de rédaction des métadonnées (axe 4)***

Dans la perspective d'assurer une meilleure visibilité internationale à leurs auteurs, les revues de CybergegeoNet 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 *CybergegeoNet* 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) qui jouent un rôle déterminant dans la diffusion et l'impact de l'article. Le groupe de travail, composé des membres des trois revues (*Cybergegeo*, *Brussels Studies* et *ViaTourism*), 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. Plusieurs exigences ont été retenues : 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 et enfin une traduction scientifiquement de qualité.

Les trois types de métadonnées (titre, résumé et mots clés) remplissant chacun une fonction spécifique, il a été convenu 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.

**Résultats :** 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 qui leur sera adressé.

#### ***Sur l'impact des améliorations sur la traduction (axe 5)***

Le numéro 16/2019 de *ViaTourism* a comporté pour la première fois des traductions en **chinois** et en **arabe**. Publier les articles dans ces deux nouvelles 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 avec un rebond notable des visites. Sur la période comprise entre les mois d'avril et septembre 2020, le nombre total de visites de la revue hébergée sur OpenEdition passe de 499 en 2019 à 4 270 en 2020.

Colette Cauvin a produit un rapport détaillé à partir des statistiques fournies par OpenEdition sur l'impact de la traduction des métadonnées et de certains articles sur la consultation de *Cybergegeo*, qui conduit à plusieurs constats. Globalement entre 2016 et 2020, la fréquentation du site, partant d'un niveau élevé en 2016 (année de la célébration du 20<sup>e</sup> 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.

## Conclusions et perspectives

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.

## Valorisation du projet

L'ensemble des livrables seront mis à disposition en accès libre sur le site des partenaires et déposés sur un entrepôt d'archives ouvertes dans le respect des principes FAIR de la science ouverte.

Il est également prévu qu'un dossier spécial sera consacré à CybergeoNet dans la revue.