The OPTIMICE Project: Optimising Translation Quality of Metadata in the Editorial Chain of Academic Journals

Franck Barbin
Rennes 2 University
LIDILE

Katell Hernández Morin
Rennes 2 University
LIDILE

The OPTIMICE project has developed a method that combines neural machine translation and human post-editing to enhance the quality of article metadata when translating from French to English as part of the journal editorial process. In partnership with the LIDILE research unit, the PUR (a French publisher) and the MSHB (French Centre for Human Sciences), we comparatively assessed the quality of human and machine translations of the metadata of 32 articles using our proprietary quality assessment grid and professional translators. The aim was to precisely determine the qualitative elements and limitations of each output, and to design the most appropriate translation method. We then formulated recommendations for writing and translating metadata to complement guidelines for authors, and improve the acceptance, referencing and international visibility of papers in journals. The method was finally tested on 2021 issues of the 4 selected journals, focusing on history, archaeology, education and geography respectively. The objective is to develop a methodology that can be reproduced and transferred to other journals, languages and disciplinary fields.

Keywords: metadata, translation, quality, machine translation, journals, post-editing, editorial chain, HSS

INTRODUCTION

In Humanities and Social Sciences (HSS) as in other scientific fields, the domination of the English language in journal publications is no longer an issue to be debated. English largely remains the academic lingua franca (Mauranen et.al., 2016; Rowley-Jolivet, 2017). In France (and some other countries), the fact that all researchers are not able to publish directly in English without having their paper professionally translated is a problem (Garnier, 2020). In this context, the French Department of Higher Education and Research (MESRI) launched a bid for research projects on the use of machine translation in HSS. OPTIMICE, which stands for Optimising Machine Translation of Metadata and its Integration into the Editorial Chain, was one of the awarded projects. Article metadata consists of titles, abstracts and keywords.

The project aimed to raise awareness about the use of machine translation (MT). Specifically, when writing and translating their metadata, authors must demonstrate their linguistic and disciplinary expertise, and “added value” over such systems (Loock, 2018, p. 787). They should not be deceived by the apparent fluency of the MT output. Koehn (2020, 19) rightfully reminds us that “[t]he goal of current machine translation research is not to achieve perfect translation but to drive down error rates of machine translation systems”. Our method is in line with this principle, and simply seeks to improve the overall quality of metadata produced in English by French researchers, and increase their acceptability rate and referencing in international journals.

Our paper is divided into three main parts. We first describe the OPTIMICE project. We then present our assessment grid (the TRASILT grid) and show how we used it to compare human and machine translations in the project. We finally introduce the guidelines for writing and translating metadata that we devised for authors and editorial teams.

THE OPTIMICE PROJECT

Description

Our study is based on a partnership with the TRASILT team (Translation, Linguistic Engineering and Terminology) within LIDILE research unit (Linguistics, Language Engineering and Education), the Maison des Sciences de l’Homme de Bretagne (MSHB, French Centre for Human Sciences), and the Presses universitaires de Rennes (PUR, one of the major French publishers). The OPTIMICE project devised a method for researchers and editorial teams that combines neural machine translation (NMT, more precisely DeepL) and human post-editing (i.e., correction of machine translation to reach higher standards) to improve the quality of article metadata from French to English in the editorial process of journals. The objective is to develop a methodology for translation that can be reproduced and transferred to other journals, languages and disciplinary fields.

Our intent was to meet the publishers’ need to optimize the quality of metadata in English, publish more articles by French researchers on international platforms such as Cairn.info, and give their works more visibility.

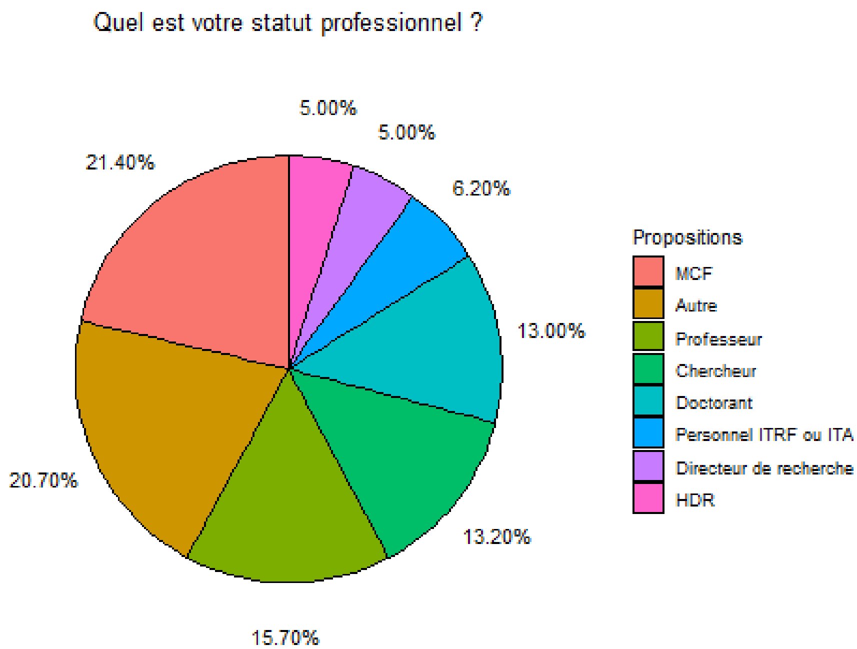
The project consisted of four main stages. Based on the English metadata of 16 articles published in 2017, it was decided to first compare their existing human translations with the NMT-generated translation of the metadata. The analysis was conducted by TRASILT researchers who are also translators. Second, the metadata of 16 other articles published in 2017 were post-edited by professional translators. It thus provided additional feedback on MT quality. The TRASILT team devised a post-editing protocol, designed for researchers and editorial teams in HSS, that was then tested on the 2021 issues of the four selected journals. Those tests enabled us to refine a general method for translating metadata that can be reproduced to other journals. An additional step was also added during the project, as we felt that we needed a better knowledge of the conditions of internationalization in the publishing industry: a national survey on translation practices among HSS researchers was carried out (see 2.2).

Our corpus was based on the metadata of 75 articles published in 2017 in four journals edited by the PUR. The journals were *Annales de Bretagne et des Pays de l’Ouest*, *ArcheoSciences*, *Éducation & Didactique*, and *Norois*, focusing respectively on history, archaeology, education and geography, all freely available on OpenEdition Journals, Persée, and the Cairn.info platform.

Survey

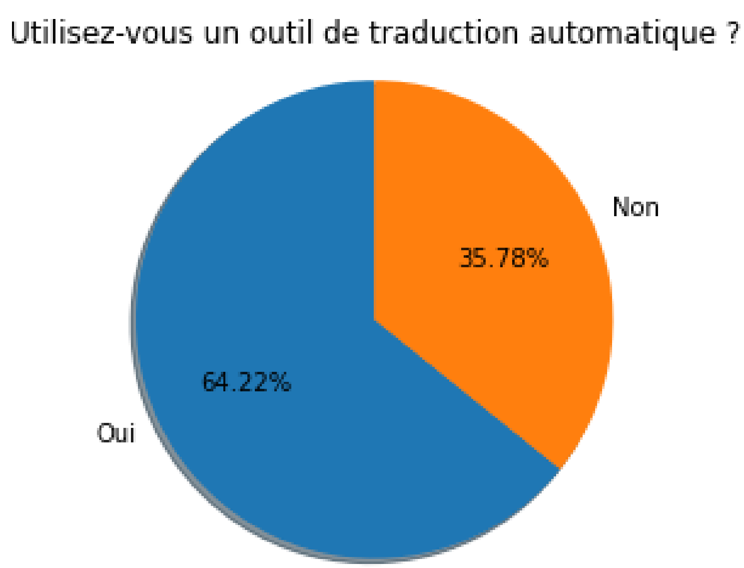
To better understand the translation practices and expectations of researchers in HSS studies, we decided to conduct a national survey that was spread over a month (May-June 2021). The respondents were people working in France who were likely to publish articles or metadata in English. They were distributed in 4 large categories as follows (Figure 1): 42.10% were professors (full, associate and assistant professors), 18.20% full-time researchers (senior and directors), 13% PhD students, and 6.20% engineers. The number of complete responses was 866.

FIGURE 1
PIE CHART ON THE PROFESSIONAL STATUS OF THE RESPONDENTS



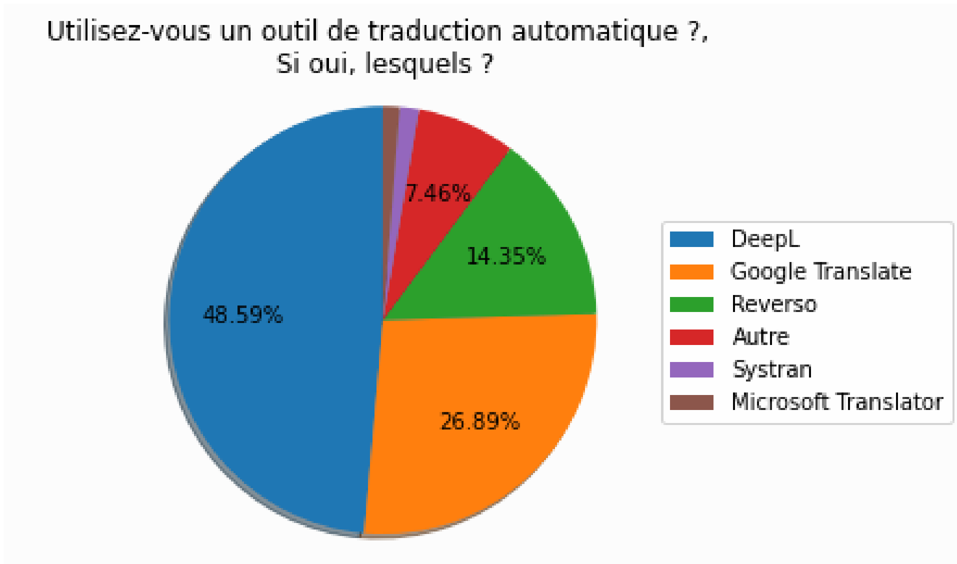
As the OPTIMICE survey is not the main focus of the paper (Hernández Morin and Barbin, forthcoming), we only present four key questions on the use of MT among researchers in HSS. First, when asked if they use any MT tool, 64.22% of respondents declared that they use MT tools (Figure 2).

FIGURE 2
PIE CHART ON THE USE OF MT TOOLS



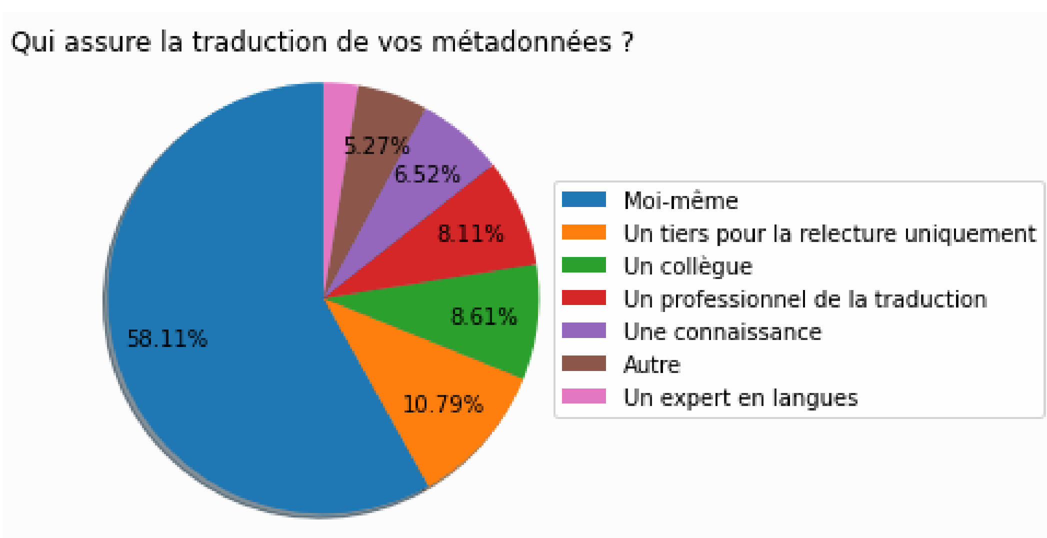
We then asked which MT tool(s) they used, 48.59% of respondents said that they use DeepL, 26.89% Google Translate, and 14.35% Reverso (Figure 3). DeepL is clearly the most common machine translation tool in the HSS community in 2021. It confirmed our initial tool choice for this project: it seemed adapted to researchers' needs (free version, overall recognized quality, user-friendly interface).

FIGURE 3
PIE CHART ON THE MT TOOLS USED BY RESEARCHERS



This should be correlated to the question of the person who translates metadata (Figure 4). As expected, 68.90% of respondents translated their metadata themselves (this figure also includes respondents who had their metadata proofread by a third party after translating them on their own). Only 8.11% could afford a professional translator. It clearly shows the lack of resources allocated to translation in HSS, especially for translating metadata.

FIGURE 4
PIE CHART ON THE PERSON TRANSLATING METADATA



With this in mind, we now discuss the quality assessment of translations and how we used our proprietary assessment grid in stages 1 and 2 of the OPTIMICE project.

COMPARATIVE TRANSLATION QUALITY ASSESSMENT

Researcher vs MT Translation Quality Assessment

As mentioned earlier in the article, we expected that translations into English by non-linguist authors in HSS disciplines would globally fall short of quality expectations for the publication of article contents in that language. Consulting advice pages from some editors' websites such as Elsevier reinforced this hypothesis: "According to a statistic shared by Elsevier, between 30 percent and 50 percent of articles submitted to Elsevier journals are rejected before they even reach the peer-review stage, and one of the top reasons for rejection is poor language" (Shaikh, 2016). Scrutiny on metadata is also very high, for "the title and abstract are incredibly important components of a manuscript as they are the first elements a journal editor sees" (ibid.). It can thus be difficult for non-English-speaking authors to reach these standards without a dedicated budget for translation or copy-editing. The first main stage of our project involved verifying our initial assumption that MT (Machine Translation) with DeepL could match or exceed the quality of HSS researchers' translations. This was achieved by comparing the HT (Human Translation) and MT outputs of 16 articles published in 2017. To this end, we chose to use an assessment grid that we developed in 2013 and fine-tuned since: the TRASILT translation assessment human metrics (Toudic, et al., 2014; Hernández Morin, et al., 2015).

The TRASILT Assessment Grid

The TRASILT grid is more or less contemporary with the EU-funded MQM translation assessment metrics (2015). Since the reason for its use was to compare human and machine translation, the choice was made to focus on a human assessment only – the outcome and publishing quality requirements being more important to us than the method used for translation. To try and be as accurate and fair as possible, a quantitative assessment grid was elected by three TRASILT researchers. The three researchers first assessed a sample of the metadata to adjust the weighting criteria. The following metadata were assessed and double peer-reviewed by two of the three evaluators, with a final agreement on each article assessment. The MQM (Multidimensional Quality Metrics) was not chosen for this study, because of the training researchers had received on their own proprietary grid, and because they thought that the TRASILT grid made a clearer distinction between issues (or "error types" in our grid) and quality criteria (or "end-user effects").

The TRASILT grid has three dimensions: it allows to spot 9 error types (Figure 5), but sanctions 4 types of potential effects (Table 1) for each error on translation quality (accuracy, usability, readability and compliance), according to a weight given to each effect (not counted effect, minor, major, or critical).

TABLE 1
THE FOUR END-USER EFFECTS OF THE TRASILT ASSESSMENT GRID

Accuracy	Error prevents the correct conveyance of information in the source document
Usability	Error prevents correct use of the product, process or document
Readability	Error has an impact on the fluency and clarity of the target document
Compliance	Target document does not comply with language-, country-, culture- or client-specific standards, conventions or recommendations

Flawed extracts and their corresponding error types can automatically feed the grid, thanks to a macro based on color identification of errors (Figure 5). Effects on quality can sum up to a maximum of 5-point penalties, according to their estimated severity (with a maximum of two effects per error, and a dominant effect among the two applicable effects).

FIGURE 5
VIEW OF THE TRASILT ASSESSMENT GRID FOR AN ARTICLE FROM ABPO

LIDILE EA 3874 — The TRASILT Grid: a Three-dimensional Translation Quality Assessment Grid for Training, Scientific, and Professional Purposes

Choose File	Trads_ABPO_1_Lehuédé_GPT_KHM_FR.docx	HUMAN TRANSLATION (HT)					
Context	**Deficiency**	**Error Type**	**Effect on quality (0= no effect/not counted effect, 1= minor, 2 = major, 3 = critical)**				**Correction**
			Accuracy	Usability	Readability	Compliance	
The relief from the East lodging of Suscinio was long considered a simple decorative element of the façade.	Suscinio	Localization		2			Suscinio's castle
was never associated to the symbolism of the dukes of Brittany and was more consistently connected with two contemporary devices belonging to kings Charles VI and Richard II.	hart	Terminology				2	stag
was never associated to the symbolism of the dukes of Brittany and was more consistently connected with two contemporary devices belonging to kings Charles VI and Richard II.	devices	Terminology				1	badges
In this context it is worth questioning whether the hart was really a ducal device or an unsuccessfully attempt of which the relief is the only vestige.]	the relief	Omissions/Additions	1	2			the relief from Morbihan

The Results tab produces totals per error type and functional effect, and total weighted scores. It calculates a quality ratio by comparing the total number of errors and the total of weighted scores (depending on the criticality of the effects). Positive effects of translation can also be individually counted in the quality score. Figures 6 and 7 are simply given to illustrate how the TRASILT grid works. More precisely, they indicate that the metadata of an article generated by DeepL (MT) obtained a higher quality ratio (i.e., a lower score) than the same metadata translated by the author (HT).

FIGURE 6
RESULTS TAB OF THE TRASILT GRID FOR HT-GENERATED METADATA OF AN ARTICLE FROM NOROIS

Assessment Results			
Document Information		**Translator's Information**	
Name of the Document	Trads_Norois_1_Di Pietro_FINALE.docx	Translator	
Type of translation	HUMAN TRANSLATION (HT)	Operator Code	FB/GPH
Batch Number	Norois 242	Team	
Page Numbers		Level	
Number of Words Translated	290		
Number of Words Assessed	290		
Original Language	French		
Target Language	English		

Category	Error Count	Effects: Accuracy	Usability	Readability	Compliance	Total
Meaning	6	5	9	0	0	14
Omissions/Additions	8	7	4	0	0	11
Terminology	4	1	0	1	4	6
Phraseology	6	0	0	0	2	2
Grammar/Syntax	3	0	0	0	3	3
Spelling	3	0	0	0	3	3
Style	3	1	0	2	1	4
Localization	3	0	0	0	3	3
DTP	0	0	0	0	0	0
Sub-total	36	14	13	3	16	46
Bonus	0	0	0	0	0	0
Total weighted score						46

General comment	Quality ratio
	[illegible]

FIGURE 7
RESULTS TAB OF THE TRASILT GRID FOR MT-GENERATED METADATA OF AN ARTICLE FROM NOROIS

Assessment Results			
Document Information		Translator's Information	
Name of the Document	Trads_Norois_1_Di Pietro_FINALE.docx	Translator	
Type of translation	MACHINE TRANSLATION (MT)	Operator Code	FBIGPH
Batch Number	Norois 242	Team	
Page Numbers		Level	
Number of Words Translated	278		
Number of Words Assessed	278		
Original Language	French		
Target Language	English		

		Effects				
Category	Error Count	Accuracy	Usability	Readability	Compliance	Total
Meaning	3	2	3	0	0	5
Omissions/Additions	0	0	0	0	0	0
Terminology	3	0	0	0	3	3
Phraseology	1	0	0	0	1	1
Grammar/Syntax	1	0	0	0	2	2
Spelling	1	0	0	0	1	1
Style	2	0	0	3	0	3
Localization	2	0	1	4	0	5
DTP	0	0	0	0	0	0
Sub-total	13	2	4	7	7	20
Bonus	0	0	0	0	0	0
Total weighted score						20

General comment	Quality ratio
	13,17

Comparative Assessment Results

This model and scheme of assessment was used to compare the metadata translated by HSS researchers and, subsequently, by DeepL. Without going into all the details of our assessment, we can underline a few tendencies observed on the 16 assessed papers (Tables 2 and 3).

TABLE 2
GLOBAL QUALITY SCORES BY METHOD AND JOURNAL

Journals	Human Translation scores	Machine Translation scores
ABPO		
Article 1	23	23
Article 2	42	31
Article 3	17	12
Article 4	13	8
ABPO mean score	**23.75**	**18.5**
ARCHÉOSCIENCES		
Article 1	16	7
Article 2	76	21
Article 3	38	35
Article 4	65	16
Archéosciences mean score	**48.75**	**19.75**
ÉDUCATION & DIDACTIQUE		
Article 1	21	11
Article 2	26	20
Article 3	21	26
Article 4	16	25
Éducation & Didactique mean score	**21**	**20.5**
NOROIS		
Article 1	46	20
Article 2	18	16
Article 3	50	18
Article 4	48	9
Norois mean score	**40.5**	**18**
Mean score of the 4 journals	**33.5**	**19.18**

The mean score obtained for human translation (HT) on all metadata was 33.5 against 19.18 for machine translation (MT). Globally, HT mean score almost doubles MT score (the higher the score, the lower the overall quality obtained). The level of human quality was heterogeneous throughout the journals, as shown in Table 2 above. Metadata translations in one journal, *Éducation & Didactique*, resulted in similar scores (21 versus 20.5) with HT and MT. This can be explained by several factors, such as the great use of idiolects and neologisms in this journal (features that are usually badly handled by MT). The varying command of the English language among translating authors is another explanation for the differences in HT quality among the four journals. Nevertheless, HT of metadata performed a little better than MT in only two papers out of 16, and the MT quality score was most of the time lower, or much lower than the HT score. It seems that MT in this context can provide a minimum level of overall quality throughout all the papers and disciplines selected (scores ranging between 7 and 35 with MT, versus 13 to 76 with HT).

TABLE 3
GLOBAL QUALITY SCORES FOR HT AND MT

	Human Translation (HT)	**Machine Translation (MT)**
Most frequent errors per method	Grammar/syntax (64) Omissions/additions (59)	Terminology (32) Meaning (25)
Most weighted effects per method	Compliance (178) Usability (103)	Usability (133) Compliance (59)

But beyond those results that were largely expected, we were above all interested in the type of recurrent errors found with both methods: the most frequent errors observed in human translations were grammar and syntax errors, and omissions or additions of information. With machine translation, the main flaws were terminology and meaning issues – to a lesser extent. In terms of effects on quality, the errors affected mostly compliance (to standards, language conventions or target culture, according to the grid definition) in human translations. With MT, the most weighted effect impacted the usability of the paper.

To summarize our main conclusions, HT logically reflects a better field command by authors, but the linguistic quality of translations is uneven from one author to another, with authors sometimes adding or omitting elements in their translations. MT offers higher linguistic and content accuracy, and a better readability or fluency, but translations are often literal and terminology is handled erratically. Faced with these observations and with the help of professional translators and scientific editors, we established a series of recommendations to authors and editing teams, completing the existing instructions to authors.

Guidelines for Writing and Translating Metadata

There is often no guidance on how to write or translate when submitting an article. The following recommendations seek to improve the overall linguistic quality and fluency of the translation of article metadata.

Our guidelines (Barbin, et al., 2022a) are divided into three phases: writing metadata in French for machine translation, implementing DeepL, and optimizing the quality of the translation in English.

Examples of Recommendations

We will quote four simple examples of recommendations to illustrate the kind of problems and solutions encountered. First, during the writing phase in French, it is good practice to avoid synonyms for the same notion (Table 4). The golden rule of one term per concept should prevail. This recommendation is in line with the standardization of keywords and terminology, which is a major issue for journals.

TABLE 4
AVOID USING SYNONYMS

✗	l'**emblème** du cerf la **devise** du cerf	*deer **emblem*** *deer **motto***
✓	l'**emblème** du cerf	*deer **emblem***

In this example from *Annales de Bretagne et des Pays de l'Ouest*, the historic term "emblème" (emblem or symbol in English) should not be equally replaced in French by "devise", for it could trigger a different MT translation. Even if the use of the French word "devise" is legitimate in the field, this technical understanding of the term may mislead the MT, which could easily translate it by "motto" or even "currency".

TABLE 5
IMPORTANCE OF CAPITAL LETTERS

✗	les SCIC comme **bois bocage énergie**	*SCICs like* ***wood, bocage énergie***
✓	les SCIC comme **Bois Bocage Énergie**	*SCICs such as* ***Bois Bocage Énergie***

Second, the presentation of the metadata (upper/lower case, spaces, line breaks, etc.) also plays a part in their understanding (Table 5). As "Bois Bocage Énergie" in the example is the name of a French company, each word should be capitalized, otherwise MT will consider that they are separated words, and logically miss the fact that they should not be translated separately and put in lower case. When writing their abstracts, authors should indicate when it is not a random series of words and they refer instead to a unique concept.

TABLE 6
RELEVANCE OF FIELD TERMS

✗	Exemplification par la comparaison de deux épisodes mettant en scène des dyades entraîneur / **lanceur** [sport javelot]	*Exemplification by comparing two episodes featuring coach/**pitcher** dyads* [sport base-ball]
✓		*Exemplification by comparing two episodes featuring coach/**thrower** dyads*

Third, authors have to check the relevance of machine translated terms in their very specific context (Table 6). In this example (in sports education), the term "lanceur" (in the context of javelin training) was translated by the engine as "pitcher", which corresponds to a baseball term, whereas "thrower" should be used in the context of javelin.

TABLE 7
ACCURACY OF PROPER NOUNS

✗	site de Saint-Georges-de-Rouelley (**Manche**)	*the Saint-Georges-de-Rouelley site (**Channel**)* [Manche, mer]
✓		*the Saint-Georges-de-Rouelley site (**Manche, Normandy**)*

Another frequent feature in HSS translation is proper nouns (Table 7). Toponymy and cultural references are a well-known issue with MT. This funny example illustrates the problem. Here, the county of “Manche”, in Normandy, was translated as “Channel”, as in the name of the sea between France and England. Those aspects have to be systematically checked by authors after translation. Also, it is best to specify the location of a French site for an English-speaking audience (here, “Normandy” to better locate the county of “Manche”).

Integration into the Editorial Chain

We are well aware that integrating those recommendations is not self-evident. Authors and editing teams may be somewhat reluctant to adopt the OPTIMICE method, leading to changes in their publishing habits and processes. Each journal has its own ways, and guiding and collaboration are key to integrating those recommendations into the editorial chain.

The experiment enabled us to determine at which stage in the chain it was most relevant to integrate the MT phase. If publishers implement the method, it will be more efficient for them to work at the end of the chain rather than trying to anticipate the journal schedule, which can vary from one journal to another. If authors carry out MT and post-editing, it should be supervised by the editing teams of journals, and not by publishers.

PERSPECTIVES OF DEVELOPMENT AND EVOLUTION

Implementing MTPE (Machine Translation and Post-Editing) in the editorial chain is a good way to raise awareness on scientific copy editing and translation practices. This project has led to a better mutual understanding between translators and field specialists, who shared common quality objectives. Improving the linguistic and semantic quality of paper abstracts, titles and keywords gives greater international visibility to articles. The PUR publisher saw the value of our translation optimization methodology and decided to spread it to other journals. Also, the project has led to the creation of three bilingual glossaries on the basis of metadata from the studied domains (archeology, education and geography). These glossaries will be fed by the editing teams of the corresponding journals (and integrated into DeepL Pro for teams who invest in the tool), bearing in mind the importance of keyword and term coherence in scientific research. Finally, a guide (Barbin, et al., 2022a) describing our method was made publicly available, together with a video (Barbin, et al., 2022b), which are regularly used in the context of training sessions for researchers and editing teams in various contexts.

On a wider perspective, this project is only exploratory, and the use of generic commercial MT tools like DeepL showed some limits: training the tool is limited to feeding it with glossaries and translations, when training specific engines by discipline would probably yield better results in the long run. Generative AI could be also added to this technology to further improve results, provided the chosen tools comply with copyright regulations. This short-term project focused on the existing tools used by researchers in 2021, but there is a strong concern for open science in the French scientific sector, and building collaborative open tools would help achieve that goal. Bigger projects are already under study, or even development on certain platforms such as Cairn Mundo (training data on the French to Latin American Spanish combination). Optimizing translation of metadata in other languages than English is also very important to foster linguistic diversity and multilingual access to research. We intend to adapt and spread our own method to other languages, starting with Spanish and Breton (our regional language), in the context of our selected journals. For this kind of efforts to be productive, close collaboration is needed between research teams and publishing sector stakeholders, at the national and international levels.

REFERENCES

Barbin, F., Hernández Morin, K., & Phuez-Favris, G. (2022a). *Rédaction et traduction des métadonnées. Guide à destination des auteurs et comités éditoriaux de revues*. NAKALA. https://doi.org/10.34847/nkl.1b145pz7

Barbin, F., Hernández Morin, K., & Phuez-Favris, G. (2022b). *Rédiger et traduire des métadonnées pour un article : Suivez le guide !* [Vidéo], Canal-U. Retrieved from https://www.canal-u.tv/134177

Barbin, F. (2022). Neural MT and human post-editing: A method to improve editorial quality. *Theory and practice of translation as a vehicle for knowledge transfer. Colección lingüística*, *71*, 15–36.

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH/German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad. (2015). *Multidimensional Quality Metrics (MQM) Definition.* Retrieved from http://www.qt21.eu/mqm-definition/definition-2015-12-30.html

Fiorini, S., Garnier-Rizet, M., Hernández Morin, K., Barbin, F., & Humphreys, F. (2020). *Rapport du groupe de travail Traductions et science ouvert*. Comité pour la science ouverture. Retrieved from https://hal-lara.archives-ouvertes.fr/hal-03640511/document

Garnier, L. (2020). Traduire les sciences humaines et sociales du français vers l'anglais. Enjeux, défis et apports. *Biens Symboliques / Symbolic Goods*, *7*. https://doi.org/10.4000/bssg.498

Hernández Morin, K., & Barbin, F. (2023). Le projet OPTIMICE : Une optimisation de la qualité des traductions de métadonnées par la collaboration entre acteurs du monde scientifique et traduction automatique. *Journal of Data Mining & Digital Humanities, Robotique du traduire*. Retrieved from https://hal.archives-ouvertes.fr/hal-03583645

Hernández Morin, K., Barbin, F., Moreau, F., Phuez-Favris, G., Efraim, O., & Quéniart, C. (2015). The TRASILT Grid: A Three-Dimensional Translation Quality Assessment Grid for Training, Scientific, and Professional Purposes. *New perspectives in assessment in translation training: bridging the gap between academic and professional assessment*. University of Westminster. Retrieved from https://hal.archives-ouvertes.fr/hal-02186887.

Hernández Morin, K., Barbin, F., Moreau, F., Toudic, D., & Phuez-Favris, G. (2017). Translation technology and learner performance: Professionally-oriented translation quality assessment with three translation technologies. *Translation in Transition. Between cognition, computing and technology*, pp. 207–232. Amsterdam/Philadelphia: Benjamins. https://doi.org/10.1075/btl.133.08mor

Koehn, P. (2020). *Neural Machine Translation*. Cambridge: Cambridge University Press.

Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4791–4796. Brussels, Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1512

Loock, R. (2018). Traduction automatique et usage linguistique : Une analyse de traductions anglais-français réunies en corpus. *Meta : le journal de traducteurs*, *63*(3), 785–805. https://doi.org/10.7202/1060173ar

Mauranen, A., Hynninen, N., & Ranta, E. (2016). English as the academic lingua franca. *The Routledge Handbook of English for Academic Purposes*. London: Routledge. https://doi.org/10.4324/9781315657455

O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *JosTrans*, *The Journal of Specialised Translation*, *17*, 55–77. Retrieved from http://www.jostrans.org/issue17/art_obrien.php

Rowley-Jolivet, E. (2017). English as a Lingua Franca in research articles: The SciELF corpus. *ASp*, *71*. https://doi.org/10.4000/asp.4987

Shaikh, A.A. (2016, updated 2021). *7 steps to publishing in a scientific journal.* Elsevier Connect. Retrieved from https://www.elsevier.com/connect/7-steps-to-publishing-in-a-scientific-journal

TAUS. (2020). Measuring the Impact of Machine Translation. *DQF BI Bulletin*, *7*. Retrieved from https://info.taus.net/taus-business-intelligence-bulletin-september-2020

Toudic, D., Hernández Morin, K., Moreau, F., Barbin, F., & Phuez-Favris, G. (2014) Du contexte didactique aux pratiques professionnelles : Proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée. *ILCEA*, *19*. https://doi.org/10.4000/ilcea.2517

Wisniewski, G., Singh, A.K., Yvon, F. (2013). Quality Estimation for Machine Translation: Some Lessons Learned. *Machine Translation*, *27*(3–4), 213–238. Retrieved from https://www.jstor.org/stable/42628816