OPERAS
open scholarly communication in the european research area for social sciences and humanities

EXPLORATORY STUDIES FOR THE CREATION OF A TECHNOLOGY-AIDED COLLABORATIVE TRANSLATION SERVICE IN OPEN SCHOLARLY COMMUNICATION

General report

Susanna Fiorini
DECEMBER 2023

OPERAS
open scholarly communication in the european
research area for social sciences and humanities

DISCLAIMER

The ideas and views expressed in the exploratory reports only reflect those of the experts involved in the studies and may not be representative of the opinions or policies promoted by any specific organization, institution, or government entity. The present report is therefore only intended for informational purposes.

Abstract

Since the Helsinki Initiative in 2019, language diversity and multilingualism have become key concerns in scholarly communication. Among the initiatives working towards a sustainable multilingual science, the Translations and Open Science project explores the potential of technology-aided translation to help produce and disseminate research in multiple languages.

This report presents an overview of the four exploratory studies conducted as part of the Translations and Open Science project in order to lay the foundations of a technology-aided collaborative translation service for open scholarly communication. More detailed information can be found in the specific deliverables of each study, cited in the present report.

Exploratory studies for the creation of a technology-aided collaborative translation service in open scholarly communication (10.5281/zenodo.10972986)

Project title	:	Translations and Open Science
Funding Scheme	:	French National Fund for Open Science
Project's coordinator Organisation	:	OPERAS
E-mail address	:	susanna.fiorini@operas-eu.org
Website	:	https://operas-eu.org/projects/translations-and-open-science/
Dissemination level	:	public
Delivery date	:	31.12.2023
Authors	:	Susanna Fiorini

Acknowledgments

We would like to thank the scientific committee and all the experts who took part in the interviews, workshops and evaluation tasks carried out as part of the four exploratory studies.

We also thank Académie des sciences, Agence nationale de la recherche, Cochrane, Éditions Quæ, Érudit, and Istex for their contribution to the study *Mapping and collection of scientific bilingual corpora*, and the LIDILE research lab of Rennes 2 University for their help in the study *Machine translation evaluation in the context of scholarly communication*.

Special thanks to OpenEdition for hosting the workshops held as part of the study *Operating model for a technology-aided, collaborative translation service*, and to Campus Condorcet for hosting the final workshop of the current phase of the project.

Contents

Introduction to the four exploratory studies

General context and goals

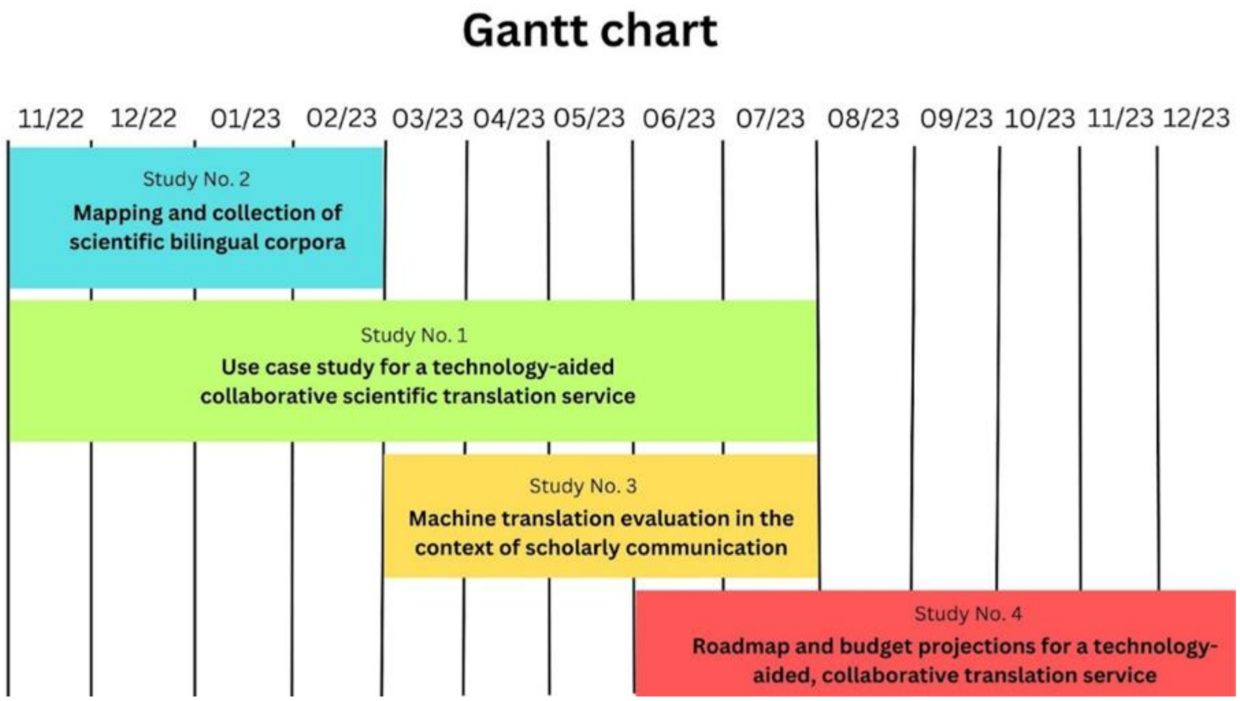
In 2019, the Helsinki Initiative formalised for the first time the importance of language diversity in scholarly communication. A large campaign was launched to raise awareness on the challenges to multilingualism in the current research landscape and to urge for a new paradigm allowing for science to be shared “In All Languages”. Since then, language diversity and multilingualism have become key concerns in scholarly communication. Several actions and solutions have been pointed out, leading the way towards a sustainable multilingual science.

The Translations and Open Science project is one of these initiatives. Following one of the commitments of the National Plan for Open Science of the French Ministry of Higher Education and Research, the project was launched in 2020 with the creation of a first working group made up of experts in natural language processing and translation. The same year, the working group published a report suggesting recommendations and avenues for experimentation with a view to establishing a scientific translation service, combining language resources, assisted-translation tools and human skills.

In order to follow up on these recommendations and lay the foundation of the translation service, a series of four exploratory studies was launched in 2022. The studies were conducted by OPERAS between November 2022 and December 2023.

The present report provides an overview of the four studies. More detailed information can be found in the specific deliverables of each study, cited in this report.

Timeline of the exploratory studies



Information and communication initiatives

More information about the project and the four exploratory studies can be found on the following pages:

- Translations and Open Science page on OPERAS website (in English): https://operas-eu.org/projects/translations-and-open-science/
- Project blog (in French): https://tradso.hypotheses.org/
- Translations and Open Science page on the French National Fund for Open Science website (in French): https://www.ouvrirlascience.fr/traductions-et-science-ouverte/
- Translations and Open Science page on the French National Fund for Open Science website (in English): https://www.ouvrirlascience.fr/report-by-the-translations-and-open-science-working-group/
- Report of the first Translations and Open Science working group (in French): https://hal-lara.archives-ouvertes.fr/OUVRIR-LA-SCIENCE/hal-03640511

Moreover, the project was presented at the following events during 2022-23:

- Workshop *La Fabrique des humanités : Traduire la pensée critique* (Palermo, Italy)
- International Encounter on Multilingualism: Language Policy and Strategies in the EU (Lisbon, Portugal)
- TRIPLE final conference (Bonn, Germany)
- Meeting at *Collège International des traducteurs littéraires* - Goldschmidt programme (Arles, France)
- Workshop *Mercredis de la traduction de l'ISIT* (Paris, France)
- Acfas annual conference (Montréal, Canada)
- Conference *L'universel à l'épreuve de la traduction : Actualités de la traduction des SHS* (Toronto, Canada)
- Workshop at *Centre d'études de la traduction* (Paris, France)
- Conference of the *Chambre nationale des entreprises de traduction* (Paris, France)
- International Conference on Human-Informed Translation and Interpreting Technology (Naples, Italy)
- Translations and Open Science Days (Paris, France)

I. Study No. 1: Use case study for a technology-aided, collaborative translation service in scholarly communication

Study overview

This section presents an overview of the study *Use case study for a technology-aided, collaborative translation service in scholarly communication*. The aim of the study was to map translation needs, practices and tools in scholarly communication in order to suggest possible workflows and features for a technology-aided, collaborative scientific translation service.

In addition to the overview, more information and results from the study are available in the study reports *Overview of translation needs, practices and tools in scholarly communication*[1] and *Suggested features and workflows for a scientific translation service*[2] (reports in French).

1.1 Scope of the study

English is by and large considered as the lingua franca of scholarly communication. Such a generalised use has certainly the advantage of facilitating exchanges in an increasingly internationalised research landscape. However, this linguistic dominance also generates inequalities among researchers and marginalises research productions in languages other than English[3], while preventing research knowledge from spreading to non-English speaking communities[4]. In order to help eliminate language barriers to knowledge production and dissemination, translation could be promptly identified as a possible solution. However, due to the highly specialised nature of scholarly texts, scientific translation requires multiple advanced skills, which can be reasonably hard to find. Moreover, if human resources are limited, so are the financial resources available to

[1] X. Auffret, C. Lapassat, P. Lacour, 2023, Technologies, outils, pratiques et enjeux actuels de la traduction scientifique. https://doi.org/10.5281/zenodo.10972812

[2] X. Auffret, C. Lapassat, P. Lacour, 2023, Propositions de fonctionnalités pour un service de traduction scientifique outillée. https://doi.org/10.5281/zenodo.10972812

[3] See for example:

V. Ramírez-Castañeda et al., 2020, Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. PLoS ONE 15(9): e0238372. https://doi.org/10.1371/journal.pone.0238372

Tatsuya Amano et al., 2023, The manifold costs of being a non-native English speaker in science. PLoS Biol 21(7): e3002184. https://doi.org/10.1371/journal.pbio.3002184

Di Bitetti, Mario S., and Julián A. Ferreras, 2017, Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications, Ambio 46.1: 121-127

[4] Z. Taşkın, G. Doğan, E. Kulczycki, A. Zuccala, 2020, Science needs to inform the public. That can't be done solely in English, LSE blog. https://blogs.lse.ac.uk/covid19/2020/06/18/long-read-science-needs-to-inform-the-public-that-cant-be-done-solely-in-english/

support traditional translation processes. As a result, translation is not a systematic activity in scholarly communication and in such a situation there is little room for large-scale optimisation.

1.1.1 Scientific translation and scholarly communication: between past and future

Since the Second World War and the invention of computers, translation has become a key subject of interest for engineers and technology developers. In addition to well-known machine translation tools, many others have been designed in order to support translation processes through the search of translated and monolingual texts, in-context translation analyses, terminology databases, as well as collaborative features, to name just a few examples.

In recent years, interest has focused mainly on machine translation, which has virtually become a mainstream tool and accounts for a large proportion of research on language technologies. Within this context of general excitement, the research community also began to look at machine translation as a possible tool for promoting multilingualism in scholarly communication, and in particular addressing the needs relating to multilingual content discoverability, foreign language writing aid and translation assistance. If the progress made in machine translation development and the resulting usage possibilities are undeniable, it is important, however, to consider the consequences of a potential massive deployment of the tool: for example, machine translation is quite demanding in terms of energy and invisibilised human work, it can perpetuate linguistic and cognitive biases and lead to a loss of human skills.

How to find the right balance in order to take advantage of machine translation, translation technologies and innovative processes, while preserving and recognising the value of human practices and skills in multilingual scholarly communication? The present use-case study is intended to be a starting point to answer this question and lay the foundations of a technology-aided, collaborative scientific translation service, combining language resources, assisted-translation tools and human skills.

Given the ambition to disseminate multilingual research publications on an international scale, two main operational objectives are set for the service: on the one hand, the members of the research community and society in general should be given the opportunity to discover and access scientific productions in all languages, and not only in English; on the other hand, all researchers should be put in the conditions to write in their preferred language, be it their mother tongue or another language according to their training background, their field of specialisation or their writing habits, without consequences on the visibility of their work. The goal is therefore to design a service of general interest, efficiently responding to actual needs and practices observed in scholarly communication. To this end, user-centred design and co-design methodologies were implemented as outlined in Section 1.2.

1.2 Methodology of the study

The study consisted of two main phases: the first phase was aimed at mapping translation needs, practices and tools in scholarly communication; the second phase was intended to make suggestions on the possible features and workflows of the technology-aided, collaborative translation service. A total of 37 potential users and stakeholders of the future service were involved in this study.

1.2.1 Mapping translation needs, practices and tools in scholarly communication

5 formal interviews + 12 informal interviews + 1 focus group workshop

The activities above were organised as follows:

- 3 formal one-to-one interviews with **3 translators** representing each of the three macro-domains of scholarly communication according to the European Research Council panel structure: 1) Life Sciences (LS); 2) Social Sciences and Humanities (SH); 3) Physical Sciences and Engineering (PE).
- 1 formal interview with members of **a scientific platform** disseminating a variety of publications mainly in Social Sciences and Humanities, but also from a more interdisciplinary perspective (links with LS and PE).
- 1 formal interview with members of **a scientific institution** disseminating research publications mainly in Life Sciences, Physical Sciences and Engineering.
- Informal interviews with **9 additional translators** mainly specialising in Social Sciences and Humanities, **1 translation professor and researcher**, **1 researcher in physical sciences**, **1 research engineer**.
- 1 focus group workshop on scientific translation needs, practices and tools with **4 researchers**, **4 translators**, **3 publishers**, **2 dissemination platforms**, **1 research engineer**.

1.2.2 Suggesting possible features and workflows for a technology-aided, collaborative translation service

1 co-design workshop + 1 focus group

The activities above were organised as follows:

- 1 **co-design workshop** on the expected workflows and features of the future service with **4 researchers**, **4 translators**, **3 publishers**, **2 dissemination platforms**, **1 research engineer**, **1 librarian**.
- 1 **focus group** with members of the steering committee and the scientific committee of the project.

Examples of questions for translators

- What types of documents do you translate (papers, monographs, metadata, etc.)? For what type of medium and audience?
- What is your standard translation process? Who are the other parties involved in the workflow? What tools do you use at each stage? What are the critical points in this process and the challenges to overcome?
- How does translation fit in with the overall process of scientific publishing?
- In your opinion, what are the specific characteristics of your discipline? How important is disciplinary expertise in order to translate texts from your discipline?
- What are your favourite translation tools, and why? What features (existing or missing) are important to you?
- In your opinion, what makes a “good” translation? What could be the stages and criteria for a quality control process?
- Do you use one or more machine translation tools? If so, in what context? What is your general perception of these tools? What are their limitations, their advantages and the associated challenges?
- How do you see the future of your profession?

Examples of questions for publishers and dissemination actors

- What is the positioning of your journal? What types of publications do you disseminate, and for what audience? What languages do you translate?
- What is important to your contributors and readers?
- What is your standard publishing process? Who are the other parties involved in the workflow? What tools do you use at each stage? What are the critical points in this process and the challenges to overcome?
- How do you ensure that your publications are disseminated as widely as possible? How does translation fit in with the overall process of scientific publishing? What is your role in scientific translation, and what are the challenges?
- What are the specific translation issues in your discipline(s)?
- In your opinion, what makes a “good” translation? What could be the stages and criteria for a quality control process?
- What is your opinion on machine translation technologies?
- How do you see the future of your sector, in general and in relation to translation? Which languages will be important for you to translate in the future?

Examples of questions for researchers

- What is the standard writing language in your discipline? Are you required to express yourself in a language that is not your mother tongue or your preferred language? If so, how do you manage to do it?
- What tools do you use when writing a publication? Do you use any writing aid? If so, what advantage do you get from them?
- In your opinion, what are the specific characteristics of your discipline? How important is disciplinary expertise in order to translate texts from your discipline?
- In your opinion, what makes a "good" translation? What could be the stages and criteria for a quality control process?
- What is your opinion on machine translation technologies?

1.3 Conclusions from the study

The current scientific translation landscape proves to be quite fragmented in terms of needs, practices, leveraged skills and tools. The study confirmed a variety of existing productions, formats and dissemination frameworks, authoring practices and writing standards according to scientific domains or even subdomains. This diversity leads to specific translation requirements that must be considered in the design of the future service. Based on this assumption, a "one-size-fits-all" approach is clearly not viable. On the contrary, building a scientific translation service that is flexible and modular is the key recommendation.

In *The landscape of multilingual scholarly communication* presented in Section 1.3.1, for example, the column Available tools indicates a range of existing tools that are leveraged - or can be potentially leveraged - to support multilingual scholarly communication. This means that, in the current context, the degree of adoption of each tool can considerably change according to a number of factors, including disciplinary standards, user profiles and resource availability. Therefore, the only way to comply with such a variety of standards is to focus on interoperability in order to provide tools and features which can be used upon request according to existing practices and needs. As an example, the study revealed that machine translation and CAT[5] tools are frequently used to produce multilingual publications in several domains of life sciences or physical sciences, while in other domains - especially in the humanities and social sciences - a more reluctant attitude towards translation technologies is observed. While these differences can be explained by the very specific characteristics of disciplinary content and writing standards, it might be worth exploring if a more suitable usage of these tools can be promoted in some domains by carrying out comprehensive experiments and training programs. Another example of specificity can be found in translation expectations: a dissemination platform might consider a translation as "good" if it makes a content discoverable, for a life-science researcher a good translation leaves no room for misinterpretation, while in the humanities a greater emphasis is put on style which conveys meaning as much as specialised terms.

[5] Acronym for Computer Assisted Translation

Diversity in practices and needs, however, does not mean isolation. According to the study, collaboration is a key factor in fostering and improving translation processes in scholarly communication. Given the nature of the texts to be translated, both language and disciplinary skills must be leveraged and, in most cases, different expert profiles need to be involved in the translation workflow and in language resource management to ensure quality and efficiency. Ease of use, modularity and interoperability are therefore all the more important because they can establish collaborative dynamics in multilingual scholarly communication.

The following Section 1.3.1 offers an overview of multilingual scholarly communication, the profiles involved, their respective needs and available tools. This overview was used as a starting point for discussing the features and workflows suggested for the future translation service.

1.3.1 The landscape of multilingual scholarly communication

Profile category	Category members	Category needs	Available tools
TRANSLATION AND CONTENT PRODUCERS	• Freelance translators • Non-professional translators (researchers, PhD candidates, students) • Traditional translation agencies • Translation agencies offering post-editing services • "Internal" translation agencies (associated with a given publisher or dissemination platform)	• Produce content in English for greater impact and visibility of research • Produce content in languages other than English for wider access to research • Translate or revise content in specialised language • Interact with the authors of a publication, domain specialists who are native speakers of the source language, a reviewer who is a native speaker of the target language • Follow editorial rules and style guides established by scientific publishers	• Computer assisted translation tools • Machine translation • Writing assistants • Language resources, including multilingual corpus and glossaries • Quality assurance tools • Word processors and desktop publishing tools • Collaborative tools • Citation and reference management tools
READERS OF RESEARCH PUBLICATIONS	• Researchers • PhD candidates • Interns • Students • Teachers and professors • Professional experts • Journalists • Science broadcasters • "Unexpected readers" from the general public	• Discover, access and understand research publications in multiple languages according to specific needs • Access academically-validated, specialised language resources • Popularise research publications	• Machine translation • Language resources, including multilingual corpus and glossaries

PUBLISHERS, DISSEMINATION PLATFORMS, ARCHIVES	• Public entities • Private organisations • Variety of funding schemes, economic and operational models	• Produce and disseminate research publications • Develop international discoverability, audience and traffic for research publications, potentially through translation • Optimise translation processes and costs	• Machine translation • Computer assisted translation tools • Word processors and desktop publishing tools • Style guides
LIBRARIES, DOCUMENTATION SERVICES, RESEARCH-SUPPORT INFRASTRUCTURES	• Public entities • Non-lucrative organisations • Private organisations	• Support the dissemination of research knowledge (production, discoverability, access…) • Coordinate and federate knowledge and research resources, potentially on an international scale	• Machine translation • Computer-assisted and collaborative translation tools • Language resources, including multilingual corpus and glossaries • Knowledge repositories and management tools
DECISION-MAKERS	• National and international organisations, government and decision-making bodies	• Ensure fair and effective language policies in research production and scholarly communication	• Research policies • Coordination frameworks • Funding schemes

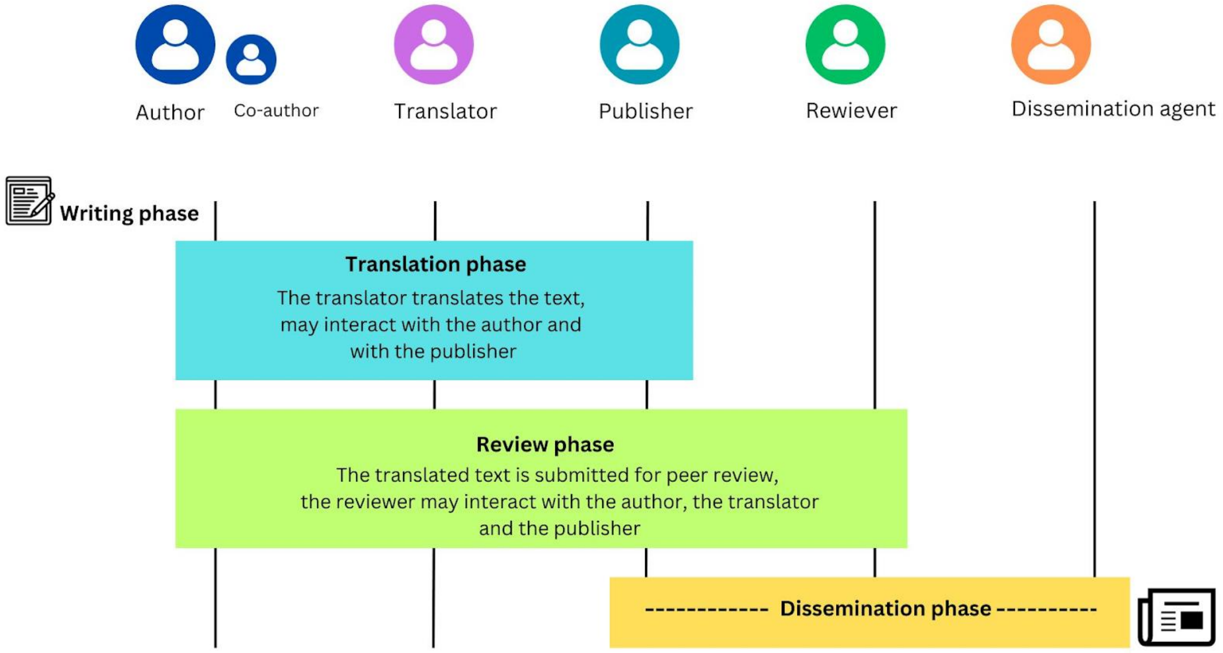

Author
Co-author
Translator
Publisher
Rewiever
Dissemination agent
Writing phase
Translation phase
The translator translates the text, may interact with the author and with the publisher
Review phase
The translated text is submitted for peer review, the reviewer may interact with the author, the translator and the publisher
------------ Dissemination phase ----------

Figure 1: Standard translation workflow in scholarly communication

1.3.2 Suggested features and workflows for the future translation service

Based on the findings and the considerations above, three main features and several workflows were suggested for the future translation service. It should be noted that this proposition does not exclude further developments based on actual practices and needs, as well as the implementation of a training and community management programme in relation to the service.

Suggested feature No. 1: a specialised machine translation engine

A feature that emerged from most discussions is a specialised machine translation engine. The idea would be to provide the community with a more transparent and respectful solution compared to the existing alternatives. The engine would be primarily used to automatically translate content, but it could also be offered as a writing assistant or a summary generator, subject to possible developments and convincing assessments of generative AI in the future.

Based on the opinions collected, the engine should be:

- **open**, in order to provide improvement and fine-tuning capabilities;
- **transparent**, in particular with regard to data collection and traceability;
- **specialised**, i.e. trained on qualitative scientific data;
- **based on a multilingual architecture**, in order to overcome the role of English as the obliged pivot language.

Also, access to the engine should be provided:

- **through a simple, direct machine translation interface**. To stand out from the competition, this interface could offer comparative features between multiple languages or several existing engines, as well as the possibility of loading specialised glossaries;
- **through the integration into the following** environments:
  - **computer assisted translation tools** → This kind of integration would offer an interactive resource to support the users of such solutions (today mainly professional translators, without excluding other possible users in the future);
  - **dissemination platforms** → The engine would be available for use both at the publishing step - in order to help researchers translating at least their abstract[6] into other languages and thus improving the discoverability of their publication - and for the readers, so that they could use raw machine translation to gist the content that has not been translated at the publishing step;
  - **word processors** → This kind of integration could be leveraged by authors as a writing aid and by the translators who do not use CAT tools as a translation assistant.

Suggested feature No. 2: a curated database of language resources

The creation and curation of specialised language resources was identified as a priority by all the stakeholders involved in the project. Indeed, quality data is not only the foundation of Artificial Intelligence (AI) training, but it is also a valuable aid for different users, for example translators who will use these resources to improve their productivity and the consistency of their translations, or students who might need to validate the use of specialised terms.

These resources could be created and enriched through different actions:

- **Collection of existing data when possible**. For more details on this point, see the chapter of this report dedicated to the study No. 2 *Mapping and collection of scientific bilingual corpora*.
- **Public calls for projects addressed to publishers, networks of professional translators and translation university programmes**. Through *ad hoc* funding schemes, publishers could plan translation projects aimed at creating and sharing specialised language resources. Such an approach would make it possible to involve the key players of scholarly communication and raise their awareness on specialised language-resource creation and management.
- **A platform allowing users to submit and access language resources**. A system to ensure the curation of this platform should be implemented in order to guarantee its scientific quality.

[6] Most of the experts who participated in the study agreed that the systematic translation of abstracts and metadata would be a reasonable starting point for the deployment of the service. However, technology-aided translation practices relating to full texts and other formats already exist and should be further explored.

- **Collection and correction of data through the use of the future translation service**. The use of the service would allow users to correct and complete the data over time, provided that they give explicit consent to data collection. Specific licences could also be applied in order to clarify the terms and conditions for using and sharing the language resources.

Suggested feature No. 3: collaborative translation tools

The study identified a variety of collaboration needs that could be met with *ad hoc* tools in order to improve scholarly translation processes. Such needs - which involve all the stakeholders, from publishers to researchers and translators - are not emphasised in this report since they were addressed as part of the OPERAS-PLUS project, Deliverable 5.5 *Design components architecture to support the translation platform*[7].

[7] Leão, D., Silva, B., & Ferreira, N. H. S. (2023). Design components architecture to support the translation platform (OPERAS-PLUS Deliverable 5.5) (v1.0 Draft). Zenodo. https://doi.org/10.5281/zenodo.8289219

II. Study No. 2: Mapping and collection of scientific bilingual corpora

Study overview

This section presents an overview of the study *Mapping and collection of scientific bilingual corpora*. The aim of the study was to identify, analyse and - when possible - collect bilingual scientific publications in the French-English language pair and in three pilot domains, in order to produce specialised language resources.

Redistributable samples of the bilingual data collected as part of this study are available for download on the Ortolang platform[8].

2.1 Scope of data mapping and collection

As highlighted in study No. 1, bilingual datasets can be useful to support translation processes, regardless of the workflow. This study focused on scholarly publications - such as papers, thesis, monographs, conference materials, institutional research reports and associated metadata - in order to build language datasets specific to scholarly translation. In particular, the data collection was intended to create in-domain translation memories[9] and terminology databases, as well as to fine-tune and evaluate machine translation engines.

In this four-month study, priority was given to the English-French language pair and three pilot domains. The choice of the three domains was made according to the criteria below:

- volume of in-domain publications available, especially in open access;
- compatibility between the domain writing standards and translation technologies;
- disciplinary diversity and cross-disciplinarity between the selected domains;
- conceptual accessibility of content for a wide audience;
- proven translation needs within the research community or for society in general;
- strong link to the challenges of the contemporary world.

For a more precise definition of the domains, the **European Research Council's panel**

[8] Corpora:
https://www.ortolang.fr/market/corpora/mob-env-esp-corpus
https://www.ortolang.fr/market/corpora/neurosciences-corpus
https://www.ortolang.fr/market/corpora/climatologie-corpus
Terminology:
https://www.ortolang.fr/market/terminologies/mob-env-esp-termino
https://www.ortolang.fr/market/terminologies/neurosciences-termino
https://www.ortolang.fr/market/terminologies/climatologie-termino

[9] Databases containing previously translated segments in the form of a parallel corpora. In general, a segment stored in a translation memory corresponds to a sentence (unlike terminology databases, which contain terms and concept-level information).

structure was used as a reference. Based on the panel structure and the criteria above, the following pilot domains were selected:

- **Climatology and Climate Change** (PE10_3)
- **Neuroscience and Disorders of the Nervous System** (LS5)
- **Human Mobility, Environment, and Space** (SH7)

The study therefore covered each of the three macro-domains of the panel:

- Physical Sciences and Engineering;
- Life Sciences;
- Humanities and Social Sciences.

The sources of bilingual scientific publications that were identified during the study include academic publishers, dissemination platforms, international research journals, scholarly and academic networks, international organisations, university repositories, as well as popularising publications. These sources are either based in French-speaking countries, or publish materials in multiple languages regardless of the country in which they are based. Examples of sources are: Theses.fr, HAL, Académie des sciences, Éditions Quæ, Érudit, Cochrane Library, OpenEdition, The Conversation, The Shift Project, Public Health Agency of Canada, Climate Change Impacts and Adaptation Division of Canada, World Health Organization.

Once the sources were identified, they were submitted for an analysis of licences and terms of use in order to determine if and how data could be collected, transformed into bilingual datasets and redistributed. When legally possible, data was therefore collected, aligned, cleaned and submitted for terminology extraction. The legal and technical steps are described in Sections 2.2 and 2.3.

What does a bilingual corpus look like? Structure of an XML bilingual file in Translation Memory eXchange format (TMX) obtained from the alignment process of bilingual publications

```
</header>
 <body>
    <prop type="x-Publication licence information:SingleString">CC-BY</prop>
     <prop type="x-Authors:SingleString">M. Hubert, K. Lebrun, P. Huynen, F.
    Dobruszkes</prop>
     <prop type="x-Title of the publication in French:SingleString">La mobilité quotidienne à
    Bruxelles : défis, outils et chantiers prioritaires</prop>
     <prop type="x-Title of the publication in English:SingleString">Daily mobility in Brussels:
    challenges, tools and priority undertakings.</prop>
     <prop type="x-URL addresses:SingleString">https://journals.openedition.org/brussels/1184 ;
    https://journals.openedition.org/brussels/1188</prop>
    <prop type="x-Disciplines:SingleString">Études urbaines</prop>
    <prop type="x-Publication type:SingleString">Note de synthèse</prop>
    <prop type="x-Publication source:SingleString">Brussels Studies</prop>
    <tuv xml:lang="fr-FR">
     <seg>La seule RBC comptait 714 111 emplois en 2010, contre 658 787 en 2000 (+ 8,4 %),
    dont plus de la moitié occupés par des non-Bruxellois.</seg>
    </tuv>
    <tuv xml:lang="en-US">
      <seg>BCR alone accounted for 714 111 jobs in 2010, compared with 658 787 in 2000
    (+8.4%), more than half of which were held by non-inhabitants of Brussels.</seg>
    </tuv>
    </tu>
    <prop type="x-Publication licence information:SingleString">CC-BY</prop>
    <prop type="x-Authors:SingleString">M. Hubert, K. Lebrun, P. Huynen, F.
    Dobruszkes</prop>
     <prop type="x-Title of the publication in French:SingleString">La mobilité quotidienne à
    Bruxelles : défis, outils et chantiers prioritaires</prop>
     <prop type="x-Title of the publication in English:SingleString">Daily mobility in Brussels:
    challenges, tools and priority undertakings.</prop>
     <prop type="x-URL addresses:SingleString">https://journals.openedition.org/brussels/1184 ;
    https://journals.openedition.org/brussels/1188</prop>
    <prop type="x-Disciplines:SingleString">Études urbaines</prop>
    <prop type="x-Publication type:SingleString">Note de synthèse</prop>
    <prop type="x-Publication source:SingleString">Brussels Studies</prop>
    <tuv xml:lang="fr-FR">
     <seg>Cette croissance est toutefois moins rapide que celle observée dans la périphérie, ce
    qui explique sans doute la légère tendance à l'augmentation de la navette sortante
    [Lebrun<bpt i="1" type="8" x="1" /> et al.<ept i="1" />, 2012 : 19].</seg>
    </tuv>
    <tuv xml:lang="en-US">
     <seg>This growth is however less rapid than that observed on the outskirts, which
    probably explains the slightly increasing trend of the outbound commute [Lebrun<bpt
    i="1" type="8" x="1" /> et al.<ept i="1" />, 2012: 19].</seg>
    </tuv>
 </tu>
```

What does a terminology database look like? Structure of an XML terminology file in TermBase eXchange format (TBX)

```
<text>
  <body>
   <termEntry>
    <note>Discipline::Études urbaines</note>
    <langSet xml:lang="en">
     <tig>
      <term>daily commute</term>
     </tig>
    </langSet>
    <langSet xml:lang="fr">
     <tig>
      <term>navette quotidienne</term>
     </tig>
    </langSet>
   </termEntry>
   <termEntry>
    <note>Discipline::Études urbaines</note>
    <langSet xml:lang="en">
     <tig>
      <term>employee commuter</term>
     </tig>
    </langSet>
    <langSet xml:lang="fr">
     <tig>
      <term>employé-navetteur</term>
     </tig>
    </langSet>
   </termEntry>
   <termEntry>
    <note>Discipline::Études urbaines</note>
    <langSet xml:lang="en">
     <tig>
      <term>urban sprawl</term>
     </tig>
    </langSet>
    <langSet xml:lang="fr">
     <tig>
      <term>étalement urbain</term>
     </tig>
    </langSet>
   </termEntry>
  </body>
 </text>
```

2.2 Analysis of publication licences and terms of use

First, the legal analysis focused on the applicable Intellectual Property (IP) and copyright laws in the countries in which the websites and dissemination platforms are based, or where the texts identified are published. This analysis showed that there are legal exceptions which allow for data collection for text and data mining purposes: it is the case, for example, of the article L. 122-5 in France and similar legal provisions in other European member states, or the Fair Dealing exception in Canada. However, these exceptions do not authorise the redistribution of the data collected.

Then, the publishing licences of the identified journals, books or works were studied. This analysis revealed a wide range of existing publication policies: for example, a number of journals apply journal-specific licences although individual articles remain the property of the authors, while more standardised open licences - such as Creative Commons - are still used by a minority of publishers.

The final step of the analysis focused on the terms of use of the dissemination platforms and websites on which the identified publications are made available. Indeed, it was important to determine whether, in the absence of a legal exception or a licence allowing copying and reuse of the content, this was nevertheless possible under the terms of use of the dissemination platform or website hosting the content.

The analysis was therefore conducted by taking into account the following steps and principles:

- **Identification of the licence under which the individual work is published**. It is crucial to verify for each publication the existence of a specific licence and to determine the related conditions. A rights holder can indeed make their content available under more or less permissive conditions than the hosting website or the law allow.
- **Assessment of the terms and conditions of use of dissemination platforms and websites.** In particular, in case of automatic collection of data (crawling), the hosting sites can put in place legal (and sometimes technical) provisions to prevent such operations. Inversely, they can also expressly authorise them.
- **In the absence of a specific licence or provision, the national intellectual property and copyright laws, including any possible exceptions, determine the use that may be made of the textual data in question**.

2.3 Data collection and processing

After identifying the sources allowing for data collection, bilingual publications and metadata were collected through web scraping, and in particular with the *Scrapy* framework. Web scraping is an automatic process that consists of collecting data from webpages, and *Scrapy* is a versatile framework that allows the user to create their own scraper that will collect data from the desired website. Once provided with a URL address, or a list of URLs, the scraper will extract the information in a structured manner as defined by the user. For instance, it can be programmed to parse the HTML structure of every fetched webpage and extract only textual data in a specific part on the page, while ignoring images, menu bars and unrelated text. Thanks to this functionality, texts and metadata - such as authors, titles and keywords - were automatically collected from each web page selected for data collection. A few variations of web scrapers were designed in order to cover the most common scenarios. In some cases, the platforms of interest provided an API or tools to directly collect the data. Part of the collected data was also manually downloaded from respective sources - for example, reports in pdf format. When dealing with large sources like university and public repositories or dissemination platforms, filtering options were used in order to select publications in relevant domains.

Once the data was collected or extracted, it was processed in an automated manner in order to produce sentence-level aligned segments in the English-French language pair.

This sentence-level alignment pipeline consists of following steps:

1. **Splitting each document to sentence level (*SpaCy*[10] and *Stanza*[11] python libraries)**

   Collected data was published and saved at paragraph or text level. However, in order to create a translation memory useful for machine translation fine-tuning, the data had to be segmented in order to have each sentence in the source text aligned to its translation in the target text. For this purpose, the data first needed to be split into sentences. To do so, tokenizer models by *spaCy* and *Stanza* libraries were used. These libraries are specially designed to process unstructured textual data, for example split text into sentences and tokens, and extract other linguistic properties (such as Parts of Speech, or PoS).

2. **Alignment (*Vecalign*[12] and *Laser*[13] python libraries)**

   Once raw collected data is split into sentences, these sentences are transformed into vector representations. This step is performed with the help of *Laser* library: this library maps sentences in English and French to the same vector space, where each sentence becomes a unique numerical vector. The design of this vector space groups together those sentences that are close in meaning. By measuring pairwise distance between vector representations of candidate English and French sentences, it is possible to determine which two sentences are translations of each other. This kind of calculation is a computationally intensive process, and *Vecalign* library helps to perform it faster thanks to the implementation of an algorithm based on *Fast Dynamic Time Warping*[14].

3. **Cleaning and post-processing**

   Once data is aligned at the sentence level, a cleaning process is performed in order to:

   - Remove empty segments
   - Remove duplicates
   - Check source and target languages, as well as translation directions
   - Measure length and remove segments with drastic length difference
   - Remove segments which contain only numbers, URLs or punctuation

[10] spaCy

[11] GitHub - stanfordnlp/stanza: Stanford NLP Python library for tokenization, sentence segmentation, NER, and parsing of many human languages

[12] GitHub - thompsonb/vecalign: Improved Sentence Alignment in Linear Time and Space

[13] GitHub - facebookresearch/LASER: Language-Agnostic SEntence Representations

[14] Toward accurate dynamic time warping in linear time and space - IOS Press

- Remove potentially misaligned sentences by measuring their semantic similarity (using *LaBSE* model[15], similar to *Laser* model in logic and design)

The bilingual corpora obtained through this process were then used to extract specialised terminology in each of the domains. The PoS tagger *spaCy* was used to tag all the tokens in the segments. Candidate terms - single-word terms and multi-word terms up to 5 tokens - were selected based on combinations of PoS tags. *LaBSE* word embeddings[16] were used to embed the candidate terms and source and target segments, and extract aligned bilingual terms based on semantic similarity. A semantic similarity score was computed between extracted source and target terms in order to help selecting correct translations of source terms. For each candidate term, a rank was calculated in order to determine how relevant they were for the segment. Corpus statistic-based techniques were also leveraged in order to extract the final term lists. Named entity recognition helped filtering out irrelevant term candidates. Lastly, terms were lemmatized and the more common term forms were selected as final term candidates.

The terms extracted using the previous automatic methods were submitted for evaluation to linguists, who were provided with the terms and up to three sentences as a context for each term. In the first step of human evaluation, a linguist who was not a domain specialist selected probable term candidates in English. This helped to get rid of the noise, that is to say extracted terms that were irrelevant. The non-expert linguist selected a short list of 500 term candidates per domain. In the second step, English-French translators who were also domain experts - one expert per domain - annotated each of the 500 terms according to annotation guidelines. They were asked to single-out only domain-relevant technical terms, as opposed to terms that are also part of general language and terms that are technical but not domain specific. The translators also made sure that automatically extracted translations were correct in fr-FR French variety by flagging and correcting non-optimal translations. After the final annotation by domain experts, only the domain-specific, relevant terms with a correct translation were included in the termbase.

2.4 Conclusions from the study

Despite some initial uncertainty about the volume of data available, the study has concluded that, even if translation is not a systematic activity in research publishing, bilingual texts are available in relevant quantities, especially in the case of abstracts and other metadata, at least in the English-French language pair (see Table 1 below). If the feasibility of the project is not compromised by the availability of bilingual data, however, other challenges have emerged.

[15] Sentence-transformers/LaBSE · Hugging Face

[16] LaBSE: Language-Agnostic BERT Sentence Embedding by Google AI | by Rohan Jagtap | Towards Data Science

Domain	Number of collected segments	Number of extracted terms
Climatology and Climate Change (PE10_3)	100,960	397
Neuroscience and Disorders of the Nervous System (LS5)	103,125	415
Human Mobility, Environment, and Space (SH7)	112,963	299

Table 1: Number of segments collected and terms extracted by domain

First of all, it would certainly be desirable to operate within a more clearly defined legal framework. For the three domains considered, between 24% and 44% of the identified sources did not indicate clear conditions regarding the possibilities of collecting textual data (see Table 2 below). For one of the domains, this percentage reaches 50% when it comes to clarifying whether it is possible to redistribute the data collected, in the form of a shared translation memory for example (see Table 3 below). Some standardisation in terms of licensing, through wider use of Creative Commons for example, could certainly help to use textual data in a more informed way. It is interesting to note that dissemination platforms can play a role in this process of standardisation: platforms generally define terms of use that are applied to all the hosted content, which makes it easier to clarify - both for publishers and users - what can be done with the textual data of a large number of publications. In the present study, the impact of the role of dissemination platforms is particularly visible in the domain of the humanities and social sciences, for which the percentage of sources expressly authorising the collection or even the redistribution of data is the highest (see Tables 2 and 3 below).

Domain	% of sources expressly forbidding the collection of data*	% sources expressly authorising the collection of data*	% requiring further analysis and/or authorisations*
Climatology and Climate Change (PE10_3)	40%	36%	24%
Neuroscience and Disorders of the Nervous System (LS5)	44%	12%	44%
Human Mobility, Environment, and Space (SH7)	13%	43%	44%

Table 2: Statistics on the collection of identified data

* Percentage calculated on the number of relevant sources identified

Domain	% of sources expressly forbidding the redistribution of data*	% of sources expressly authorising the redistribution of data*	% requiring further analysis and/or authorisations*
Climatology and Climate Change (PE10_3)	60%	16%	24%
Neuroscience and Disorders of the Nervous System (LS5)	69%	6%	25%
Human Mobility, Environment, and Space (SH7)	25%	25%	50%

Table 3: Statistics on the redistribution of identified data

* Percentage calculated on the number of relevant sources identified

The study also raised various technical challenges, especially when it comes to automated processes for data collection and processing. The quality of the texts and their translations is heterogeneous, publication formats and indexing keywords are often non-standardised: the collection and processing of textual data can therefore be more complex than expected and require significant manual effort. This concerns the identification of relevant bilingual publications, their alignment for the creation of translation memories and the extraction of terms to feed specialised glossaries and termbases - with automated terminology extraction being already intrinsically complex given the statistical and machine approaches on which it relies. As for the quality of translations, it is rare to find information about the origin of the translated texts and the processes by which they were produced, which makes it almost impossible to automatically filter out translations of questionable quality, or those that have been generated by a translation engine without any human intervention. In the light of all these efforts and constraints, it is legitimate to question the strategy of building language datasets based on existing resources and publications, especially if the dataset obtained cannot be mutually shared for legal reasons. A possible solution could be to produce *ad hoc* translations and language resources, as suggested in the first study (see Section 1.3.2 of the present report).

These initial findings seem therefore to suggest that mapping existing bilingual publications with a view to collecting scientific textual data is not enough to build specialised language resources that meet the expected criteria in terms of volume, quality and redistribution. It seems rather advisable to implement future-oriented initiatives to raise awareness on the legal and technical requirements relating to the creation and the management of shared multilingual textual datasets. Such an approach should also lead to greater recognition of the work done by translation and publishing professionals, who make it possible to produce quality data.

III. Study No. 3: Machine translation evaluation in the context of scholarly communication

Study overview

This section presents an overview of the results of the study *Machine translation evaluation in the context of scholarly communication*. The aim of the study was to assess the performance of a set of machine translation engines in different scholarly communication scenarios.

In addition to the overview, more information and results from the study are available in the study reports *Machine translation evaluation - General Methodology*[17], *Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space*[18], *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*[19], and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[20].

3.1 Engines evaluated

The engines included in the evaluation are the following:

1. **OpenNMT**: engine based on an open-source library with highly customisable, multi-parameter setup. ***Open source: YES.***
2. **ModernMT**: commercial engine allowing for simplified, user-level adaptation. ***Open source: free version of the engine only. The premium version is proprietary.***
3. **DeepL:** commercial engine which is the most used by the project's target community. ***Open source: NO.***
4. **eTranslation**: engine developed by the European Commission. Only submitted for automatic evaluation with a view to a potential use in the future. ***Open source: NO.***

3.2 Fine-tuning with in-domain datasets

One of the main aims of the evaluation was to assess whether fine-tuning can help to produce better machine translation output, especially by taking into account specialised terminology. In order to do so, the engines were fine-tuned with in-domain parallel language datasets in the English-French language pair in the three pilot scientific domains considered in study No. 2 (see Section 2.1 of the present report). The main characteristics of the datasets are presented below for each pilot domain. For further details, please refer to the reports *Machine translation evaluation - General Methodology*, *Machine translation evaluation - Outcome for discipline Human mobility,*

[17] T. Vanallemeersch, S. Szoc, K. Migdisi, L. Meeus, L. Macken, A. Tezcan, 2023, Machine translation evaluation - General Methodology. https://doi.org/10.5281/zenodo.10972872

[18] T. Vanallemeersch, S. Szoc, K. Migdisi, L. Meeus, L. Macken, A. Tezcan, 2023, Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space. https://doi.org/10.5281/zenodo.10972872

[19] T. Vanallemeersch, S. Szoc, K. Migdisi, L. Meeus, L. Macken, A. Tezcan, 2023, Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System. https://doi.org/10.5281/zenodo.10972872

[20] T. Vanallemeersch, S. Szoc, K. Migdisi, L. Meeus, L. Macken, A. Tezcan, 2023, Machine translation evaluation - Outcome for discipline Climatology and climate change. https://doi.org/10.5281/zenodo.10972872

Environment, and Space, *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*, and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[21].

3.2.1 Pilot domains and characteristics of the datasets

CLIMATOLOGY AND CLIMATE CHANGE (Physical Sciences - PE3_10)

100,960 collected segments, 397 extracted terms; translation direction of the bilingual corpus and evaluation task → English to French.

Type of publication	Documents collected
Book	1
Conference paper abstract	134
Journal article	103
Journal article abstract	1677
Publication type not available	61
Report	6
Thesis abstract	3703
Terminology	397

Table 4: Dataset statistics by document type for the PE3_10 domain

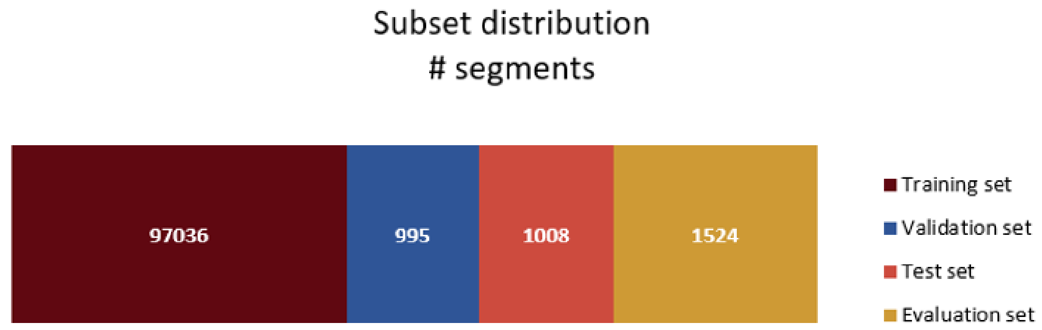


Figure 2: Distribution of training, validation, testing, and evaluation sets for the PE3_10 domain

[21] Ibid. footnotes 17-18-19-20

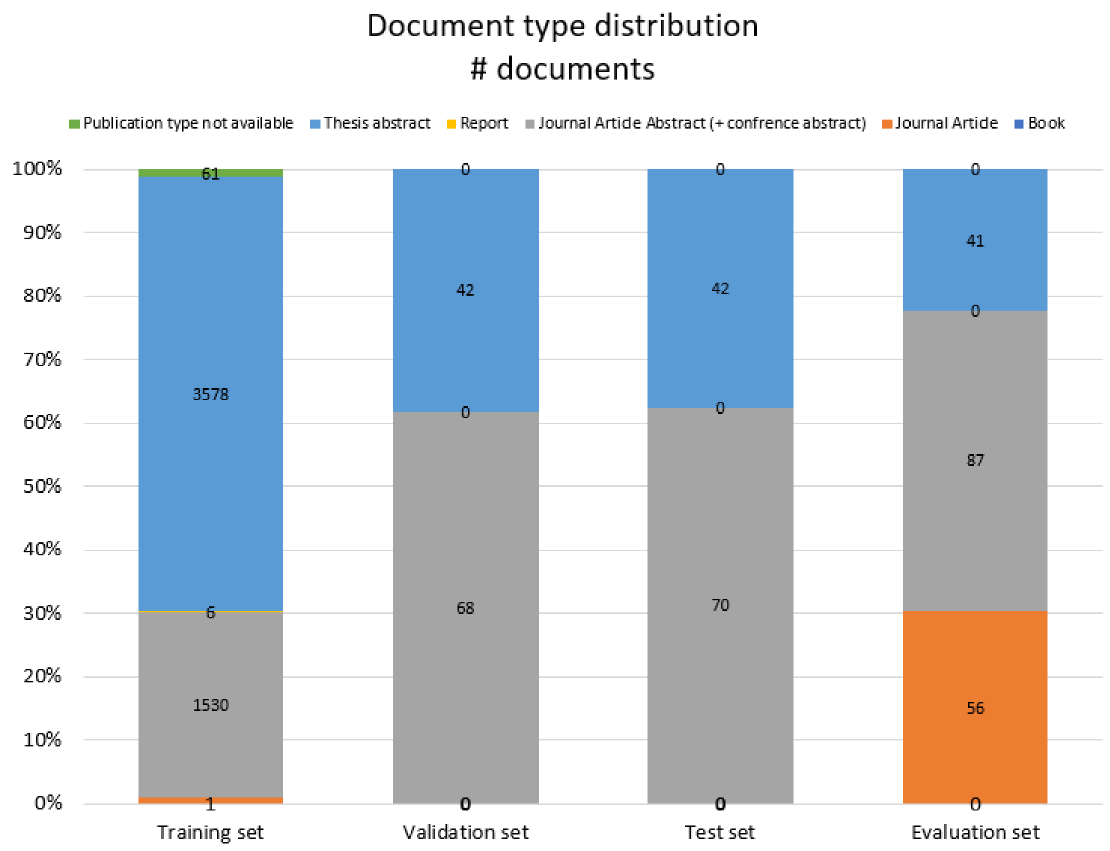


Figure 3: Distribution of publication types for each subset, number of documents for the PE3_10 domain

NEUROSCIENCES (Life Sciences - LS5)

103,125 collected segments, 415 extracted terms; translation direction of the bilingual corpus and evaluation task → English to French.

Type of publication	Documents collected
Article	170
Conference paper abstract	31
Journal article abstract	2211
Report	8
Research journal article	62
Review abstract	947
Thesis abstract	4860
Terminology	415

Table 5: Dataset statistics by document type for the LS5 domain

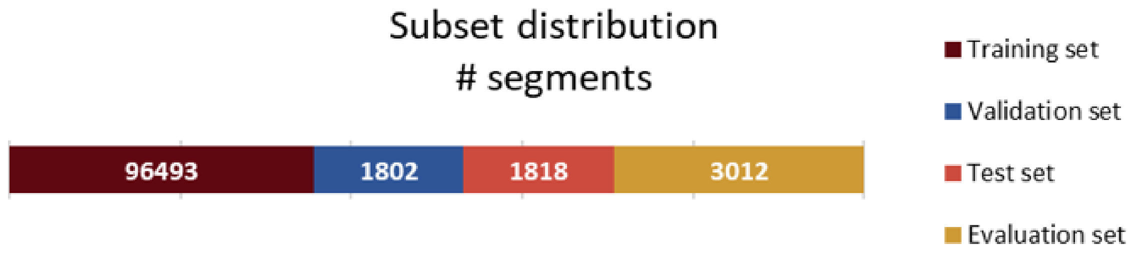


Figure 4: distribution of training, validation, testing, and evaluation sets for the LS5 domain

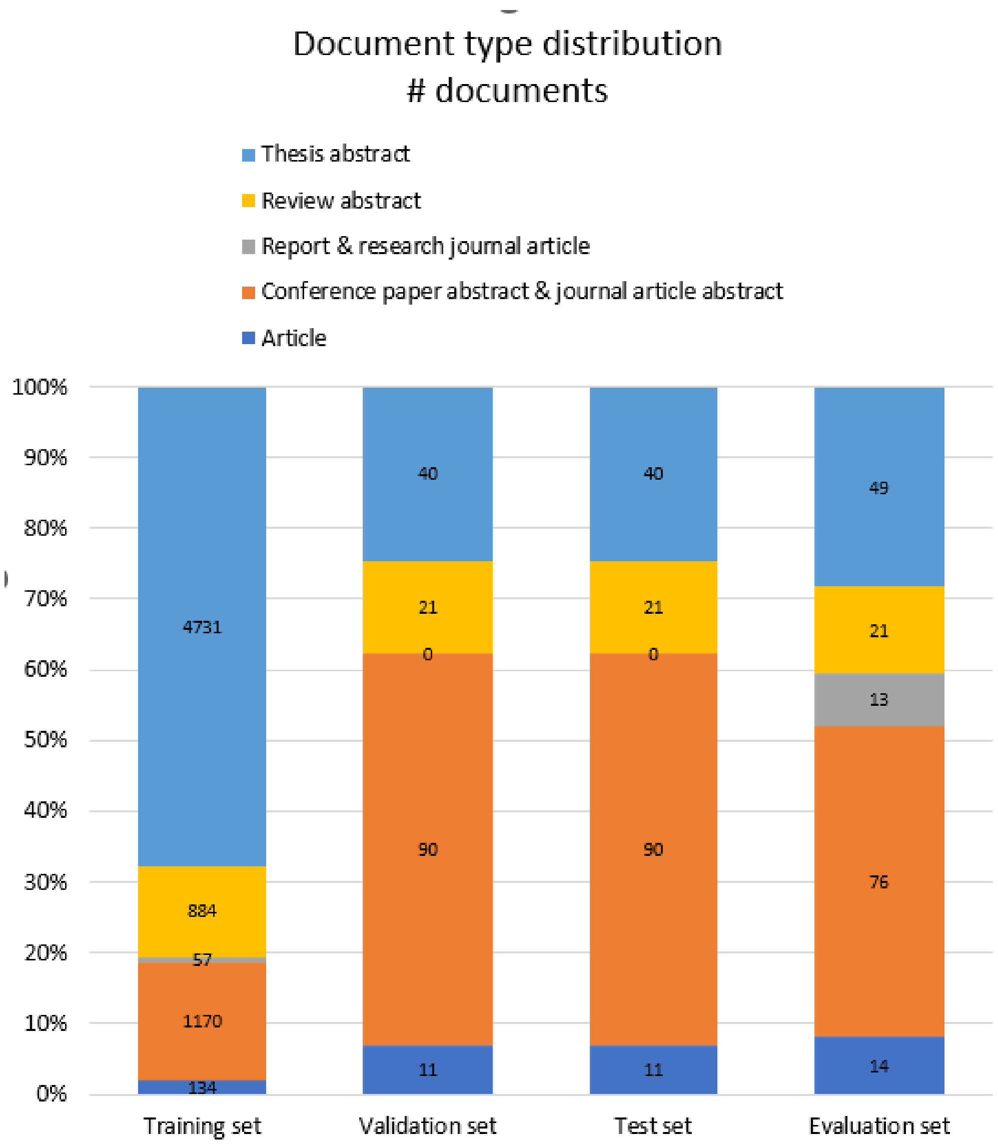


Figure 5: Distribution of publication types for each subset, number of documents for the LS5 domain

HUMAN MOBILITY, ENVIRONMENT, AND SPACE (Social Sciences and Humanities - SH7)

112,963 collected segments, 299 extracted terms; translation direction of the bilingual corpus and evaluation task → French to English.

Type of publication	Documents collected
Journal article	145
Journal article abstract	8886
Thesis abstract	3520
Terminology	299

Table 6: Dataset statistics by document type for the SH7 domain

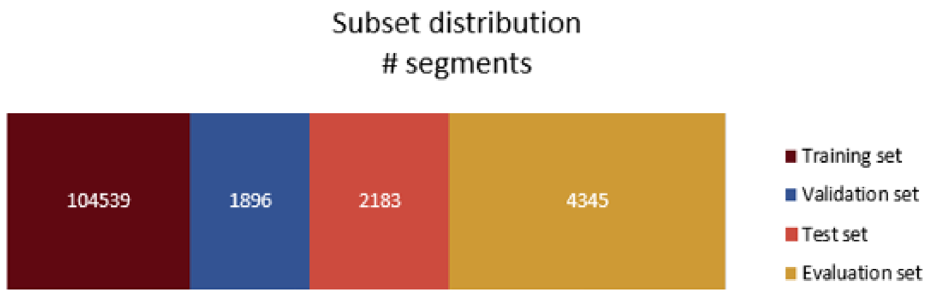


Figure 6: distribution of training, validation, testing, and evaluation sets for the SH7 domain

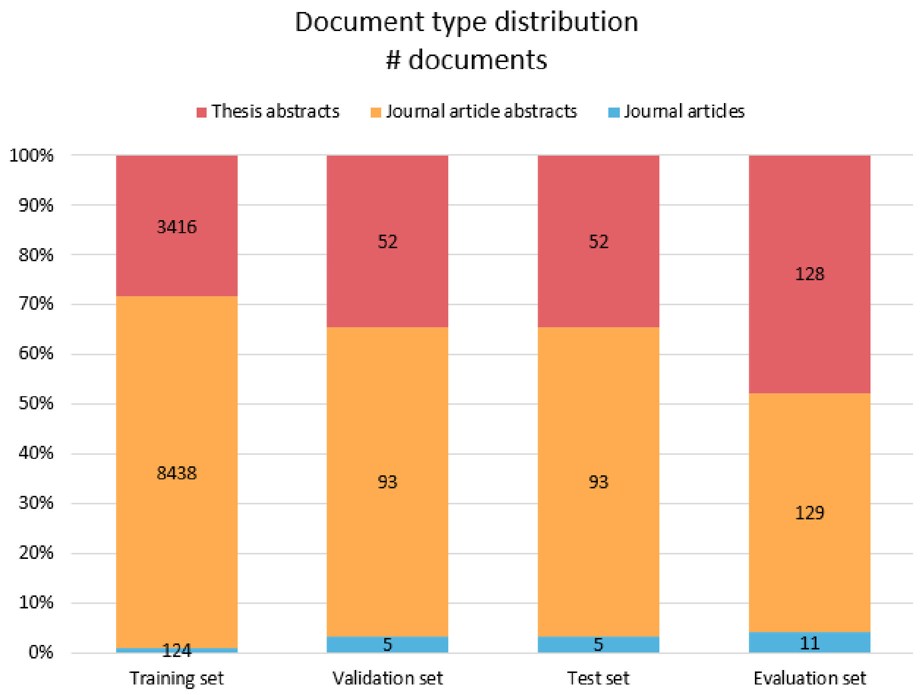


Figure 7: Distribution of publication types for each subset, number of documents for the SH7 domain

3.2.2 Fine-tuning methodology

OpenNMT: the engine based on the OpenNMT library was trained from scratch on open-source parallel datasets provided in OPUS[22]. This resulted in a generic machine translation model, which was then fine-tuned on the specialised datasets collected as part of *Translations and Open Science* project and, in a second step, also on the corpora of the *SciPar* project (~9M segments from scientific abstracts, various domains combined)[23].

ModernMT: the baseline engine was fine-tuned by uploading the specialised datasets of the *Translations and Open Science* project in TMX format *via* the dedicated feature provided in the online user interface.

DeepL: the baseline engine was fine-tuned through the *Glossary* feature for terminology customisation. It should be noted that this feature is not supported yet in all the API configurations available, which could be a potential limitation in the case of a large-scale deployment.

eTranslation: the engine offers no fine-tuning capability.

For a detailed description of the fine-tuning methodology, please refer to the report *Machine translation evaluation - General Methodology*[24].

3.3 Machine translation evaluation

The evaluation consisted of an automatic evaluation task, followed by a human evaluation task involving different profiles of evaluators. For a detailed description of the methodology, please refer to the report *Machine translation evaluation - General Methodology* [25].

3.3.1 Automatic evaluation

As a first step, the following engines were submitted for automatic evaluation by producing output for in-domain test datasets with baseline and fine-tuned engines:

1. **OpenNMT** - **baseline**

2. **OpenNMT** - **fine-tuned** with **Translations and Open Science dataset**

3. **OpenNMT** - **fine-tuned** with **Translations and Open Science dataset** + **SciPar dataset**

4. **ModernMT** - **baseline**

5. **ModernMT** - **fine-tuned** with **Translations and Open Science dataset**

6. **DeepL** - **baseline**

7. **DeepL** - **fine-tuned** with Translations and Open Science dataset (**terminology only**)

[22] J. Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), 2214-2218
[23] D. Roussis et al., 2022, SciPar: A Collection of Parallel Corpora from Scientific Abstracts, Proceedings of the Thirteenth Language Resources and Evaluation Conference
[24] Ibid. footnote 17
[25] Ibid. footnote 17

8. **eTranslation - baseline** (General text domain)

The comparison between the baseline and fine-tuned engines was intended to provide further insight into fine-tuning needs, and in particular to bring additional information about the relevance and the required level of fine-tuning effort in order to improve machine translation output.

The outputs produced by the eight engines were compared to reference translations using automatic evaluation metrics such as the statistical metrics BLEU and TER (Translation Edit Rate) and the neural (deep learning based) metric COMET.

The automatic evaluation results for the three domains tend to converge towards the following conclusions:

- **DeepL** → there is hardly any difference between the DeepL baseline and DeepL fine-tuned using the Glossary feature. In some cases, the baseline DeepL engine performed even better than the fine-tuned one. However, given that fine-tuning in DeepL only covers terminology, this could be due to terminology inconsistencies in the reference translations. In general, DeepL is the engine with the best performances, even if the gaps with other engines are not always very significant.
- **ModernMT** → the disparity between ModernMT baseline and ModernMT fine-tuned is slightly larger compared to DeepL. Except for the review abstracts and thesis abstracts in the Neurosciences domain, ModernMT tends to obtain the best scores after DeepL.
- **OpenNMT** → there is a more pronounced difference between baseline and fine-tuned OpenNMT engines, with the engine performance further improving after adding the SciPar dataset. This could suggest that in order to ensure efficient fine-tuning for scholarly communication, data collection should not be strictly narrowed to in-domain texts only. Except for the review abstracts and thesis abstracts in the LS5 domain and the thesis abstracts in the SH7 domain, OpenNMT tends to obtain lower scores than DeepL and ModernMT. It should be noted, however, that a small overlap between the test set segments and the SciPar segments went initially unobserved during the automatic evaluation stage, which can explain some higher scores. On the other hand, it should be also taken into consideration that DeepL and ModernMT were potentially trained on SciPar data as well.
- **eTranslation** → its scores are slightly lower than or comparable to OpenNMT fine-tuned without SciPar data.

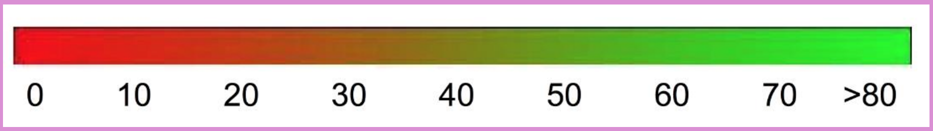
These results can be observed in the BLEU scores below. Similar observations are made with other metrics (TER, ChrF, METEOR and COMET). The TER, METEOR and ChrF scores are generally in line with the ones from BLEU, while the picture for COMET scores is more variable. For further details, please refer to the reports *Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space*, *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*, and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[26].

[26] Ibid. footnotes 18-19-20

Note on the BLEU scores

The BLEU score measures the similarity between a machine-translated text and a reference translation. The lower the score, the less the machine-translated output overlaps with the reference translation. Low scores are therefore symptomatic of low quality. Inversely, the higher the score, the more the machine-translated output overlaps with the reference translation. High scores are therefore symptomatic of good quality.

The colour gradient below can help to give a rough interpretation of the results [27].



Despite being one of the most widely used machine translation metrics, BLEU scores have recently raised some criticism because they rely on a superficial comparison between translations that does not take into account semantics (for example, synonyms are considered as differences from the reference translation, while it is perfectly acceptable to use a different word with the same meaning). The variable quality of reference translations could also require further interpretation. For this reason, as part of this study, BLEU scores were mainly considered in order to assess fine-tuning relevance.

[27] Evaluating models, AutoML Translation Documentation in Google Cloud

CLIMATOLOGY AND CLIMATE CHANGE (PE3_10) - BLUE SCORES

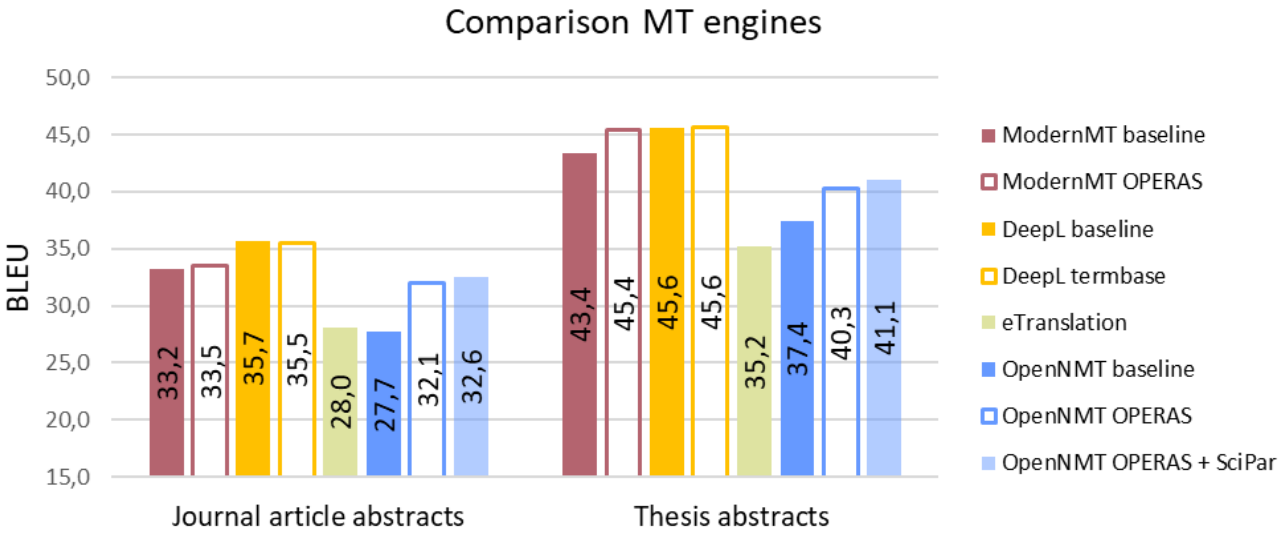


Figure 8: Comparison of MT engines, using BLEU score, for each text type for the PE3_10 domain

NEUROSCIENCES (LS5) - BLUE SCORES

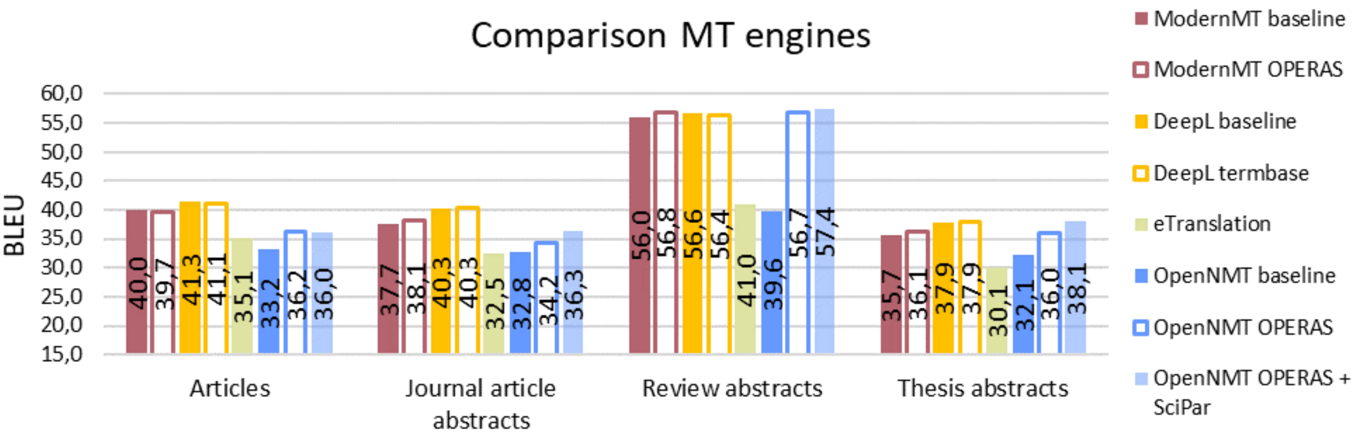


Figure 9: Comparison of MT engines, using BLEU score, for each text type for the LS5 domain

HUMAN MOBILITY, ENVIRONMENT, AND SPACE (SH7) - BLUE SCORES

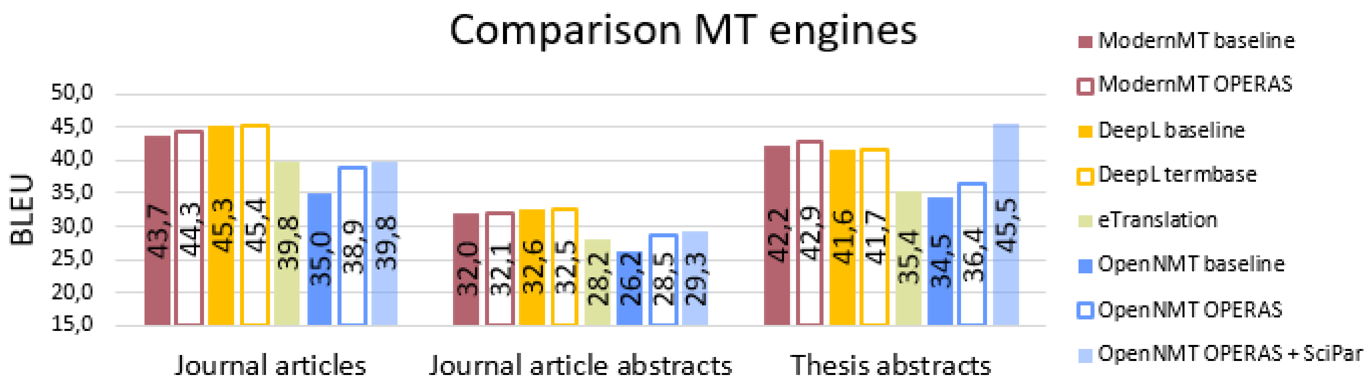


Figure 10: Comparison of MT engines, using BLEU score, for each text type for the SH7 domain

3.3.2 Human evaluation

Based on the observations from the automatic evaluation task, 3 engines were selected to perform human evaluation:

1. OpenNMT - fine-tuned with Translations and Open Science dataset + SciPar dataset

2. ModernMT - fine-tuned with Translations and Open Science dataset

3. DeepL - baseline

The human evaluation was set up to assess machine translation output usability for the usage scenarios and personas below:

- **"Translator" persona:** a professional translator who uses machine translation in a computer-assisted translation environment. The translator masters the source language, is a native speaker of the target language, and has a good knowledge of the domain in question. 6 translators took part in the evaluation (2 for each domain). They performed an adequacy assessment task, as well as a post-editing task in a dedicated evaluation tool.
- **"Expert" persona:** a researcher specialised in the domain in question, who uses machine translation to (a) translate their publication, (b) write an article in the target language (writing aid), or (c) gist texts that are not written in their native language (reading aid). The expert has a good to native knowledge of the source and target languages, as well as a perfect command of specialised terminology in both languages. 8 experts took part in the evaluation (2 for the PE3_10 domain, 2 for the LS5 domain, 4 for the SH7 domain). They performed the same evaluation tasks assigned to the "Translator" persona: adequacy and post-editing.
- **"Layperson" persona:** a person who has at most basic knowledge in the domain (e.g. a non-academic reader or a researcher in a different scientific domain). This persona has good to excellent knowledge of the target language and makes use of machine translation to gist educational scientific texts. The participants to this task read text excerpts of 100-200 words, drawn from the evaluation set, in a cumulative self-paced reading view. Based on text characteristics - such as the origin of the excerpt (abstract or full text), sentence length, and lexical variety - the texts were classified into different sets which were submitted to different user groups. The human reference translation was used as a benchmark. Reading time was measured. After reading each excerpt, the evaluators were asked to answer multiple-choice comprehension questions, as well as a YES-NO question about translation quality and usefulness.
- **MQM error annotation:** errors found in the machine translation outputs were annotated by an annotator who assigned error categories and severity. The scores relating to critical and terminology errors in machine translation output are leveraged to understand whether raw machine translation can be useful to automatically translate publication metadata and therefore improve the discoverability of research in multiple languages.

For more information on the human evaluation setup, please refer to the report *Machine translation evaluation - General Methodology*[28].

3.3.2.1 Adequacy assessment task performed by translators and researchers

The adequacy assessment task consisted of judging the adequacy of the machine translated sentences of research publications, by assigning a score between 1 and 5 (1 being the worst - see the note on the adequacy user ratings below). The aim of this task was to assess how adequately machine translation expressed the meaning of the source sentence.

The user ratings obtained for the three domains show with significant confidence that DeepL is on average higher rated than ModernMT, which is in turn higher rated than OpenNMT. It is also worth noting that researchers rate the translations on average higher than the translators.

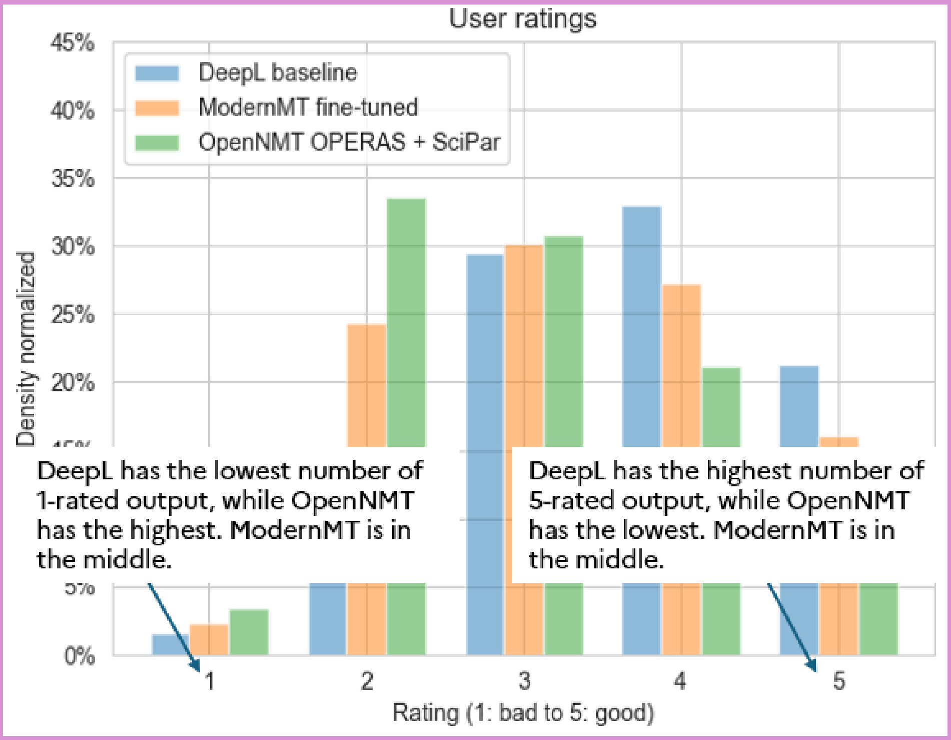
An overview of the user ratings is presented in the following charts. For further details, please refer to the reports *Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space*, *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*, and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[29].

[28] Ibid. footnote 17

[29] Ibid. footnotes 18-19-20

Note on the adequacy user ratings

As in the example below, the charts in the following pages (Figures 11, 12 and 13) show that **DeepL (light blue bar)** is the engine with the lowest number of cases of low-rated adequacy, while **OpenNMT (green bar)** has the highest number of cases of low-rated adequacy (see "Rating 1" column). Inversely, OpenNMT has the lowest number of cases of high-rated adequacy, while DeepL has the highest number of cases of high-rated adequacy (see "Rating 5" column). **ModernMT (orange bar)** is virtually always in the middle.



DeepL has the lowest number of 1-rated output, while OpenNMT has the highest. ModernMT is in the middle.

DeepL has the highest number of 5-rated output, while OpenNMT has the lowest. ModernMT is in the middle.

A positive correlation is observed between the data collected from the different evaluators. However, given the very subjective nature of these judgements, their reliability could be further improved by extending the panel of evaluators.

CLIMATOLOGY AND CLIMATE CHANGE (PE3_10) - ADEQUACY USER RATINGS

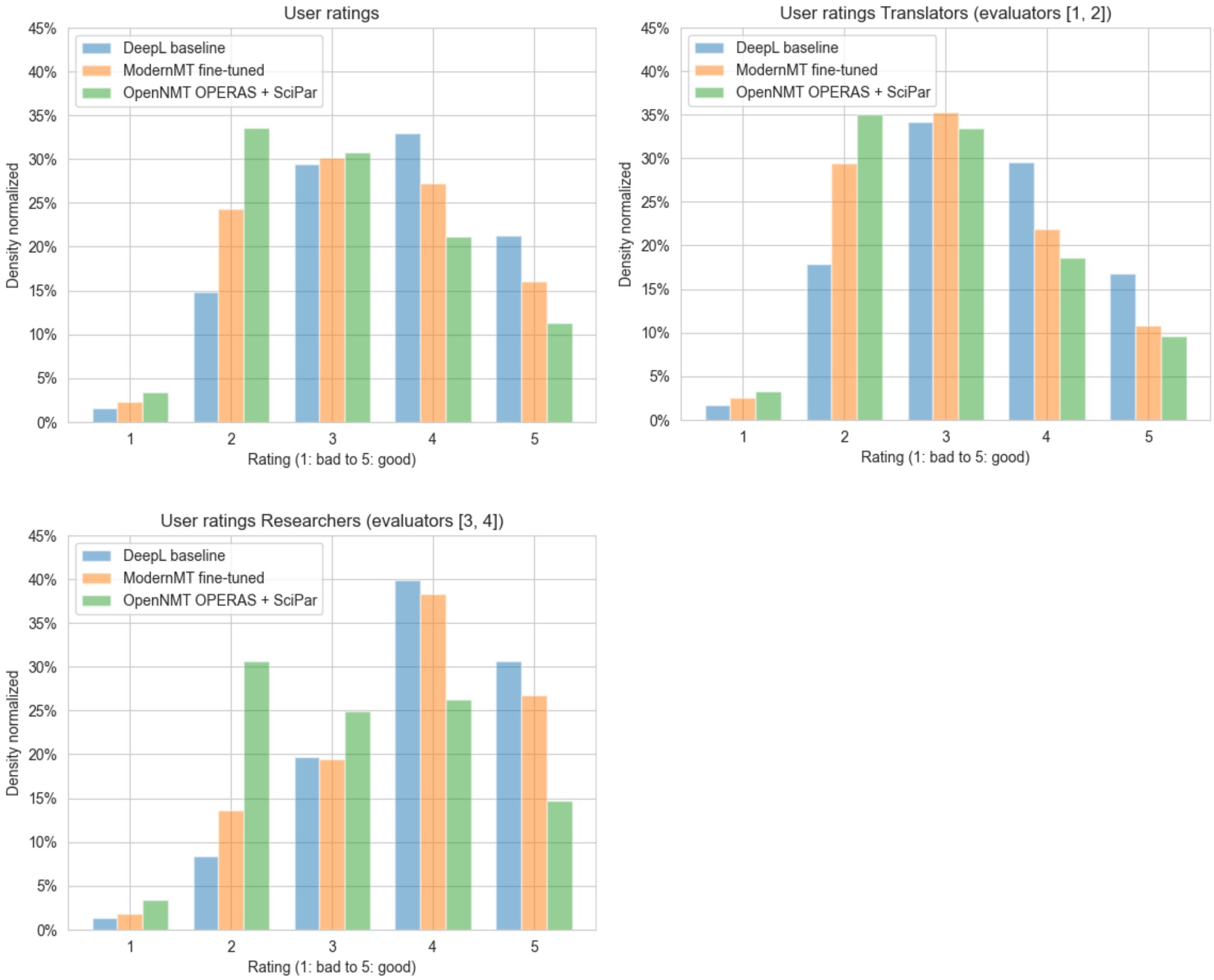


Figure 11: Adequacy user ratings for the PE3_10 domain according to user profile and all user profiles combined

NEUROSCIENCES (LS5) - ADEQUACY USER RATINGS

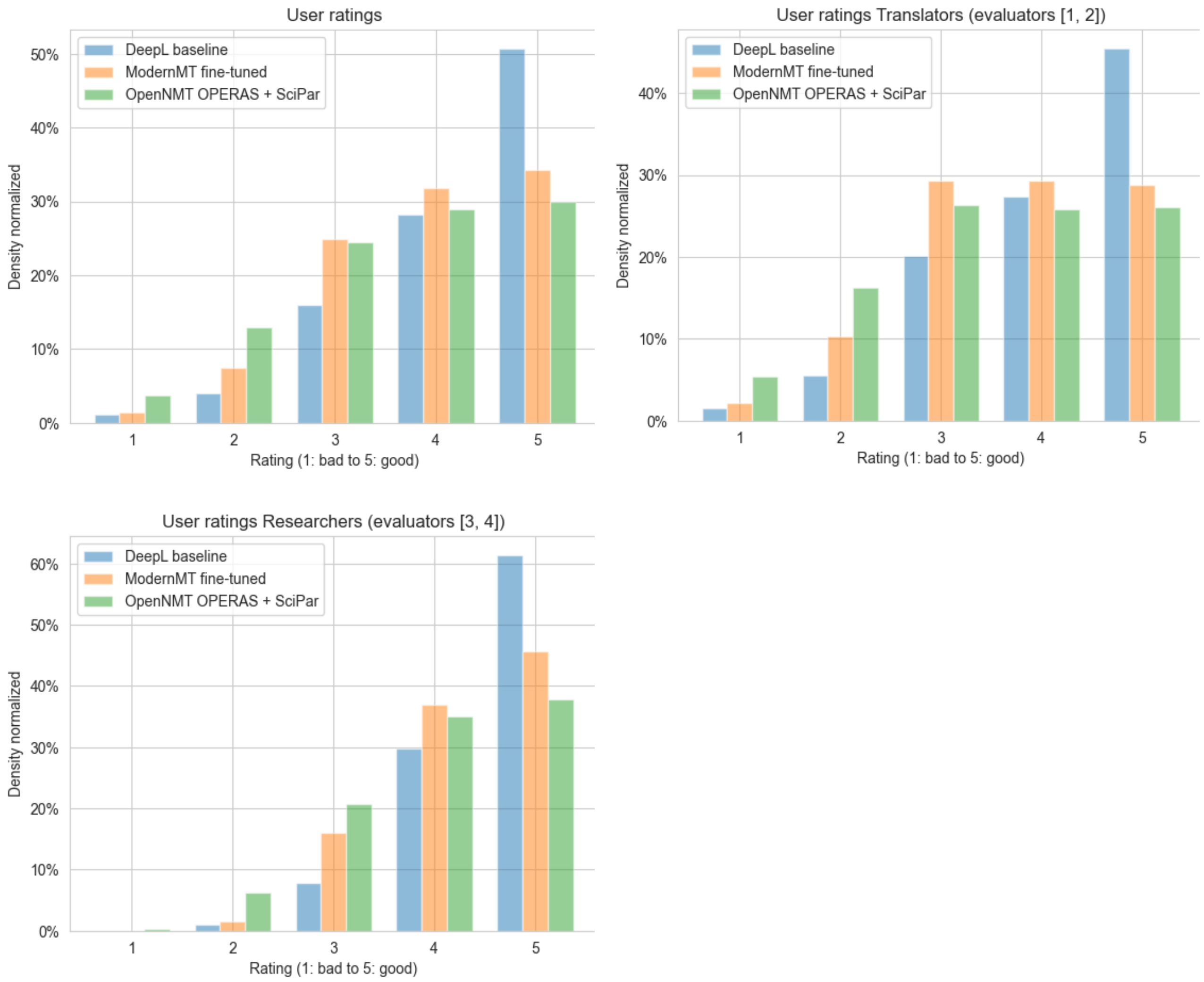


Figure 12: Adequacy user ratings for the LS5 domain according to user profile and all user profiles combined

HUMAN MOBILITY, ENVIRONMENT, AND SPACE (SH7) - ADEQUACY USER RATINGS

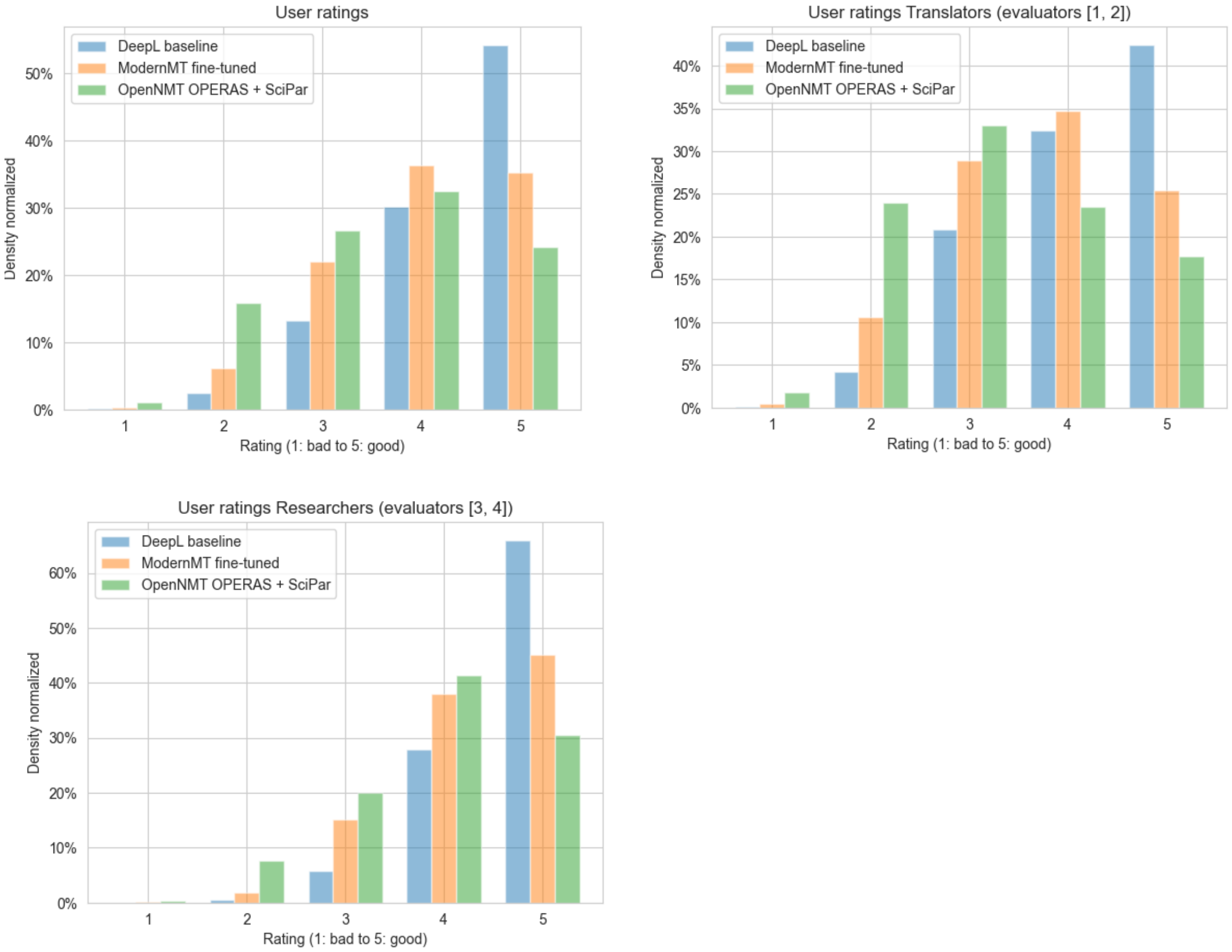


Figure 13: Adequacy user ratings for the SH7 domain according to user profile and all user profiles combined

3.3.2.2 Post-editing effort perceived by translators and researchers

The post-editing task consisted in asking the evaluators to produce a publishable translation (a terminologically valid, grammatically correct, fluent translation conveying the meaning of the source sentence), based on a source sentence, its context, and a machine translation output. The evaluators were also asked to provide a score from 1 to 5 to indicate the perceived post-editing effort for each sentence (5 being the worst - see the note on the post-editing effort measures below). This task was performed on a different test set than the one used for the adequacy task.

The perceived post-editing effort ratings obtained for the three domains show with significant confidence that post-editing DeepL outputs has a lower average perceived effort than post-editing ModernMT outputs, which in turn has a lower average effort than post-editing OpenNMT outputs.

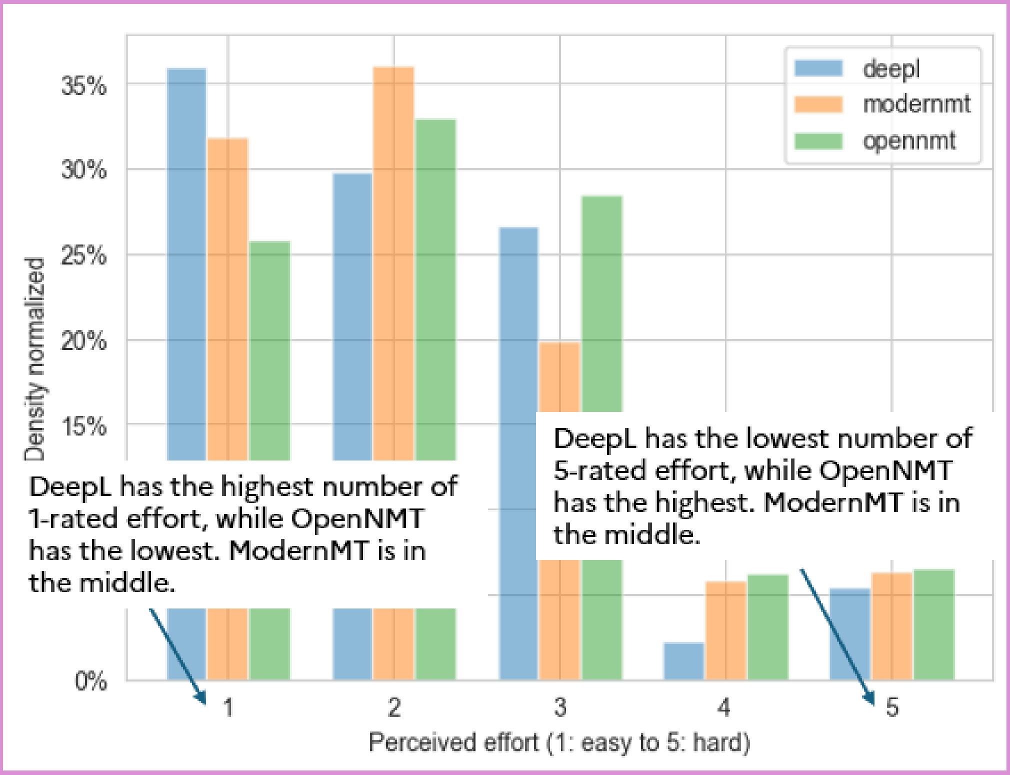
An overview of the post-editing effort perceived by the evaluators is presented in the charts below. For further details, please refer to the reports *Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space*, *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*, and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[30].

[30] Ibid. footnotes 18-19-20

Note on the post-editing effort measures

The post-editing effort measures are relevant indicators of machine translation output usability. In fact, even if no actual error is corrected in a given sentence during post-editing, the output may demand a great deal of effort to judge whether the translated sentence is acceptable or not.

As in the example below, the charts in the following pages (Figures 14, 15 and 16) show that **DeepL (light blue bar)** is the engine with the highest number of cases of low-perceived effort, while **OpenNMT (green bar)** has the lowest number of cases of low-perceived effort (see “Rating 1” column). Inversely, OpenNMT has the highest number of cases of high-perceived effort, while DeepL has the lowest number of cases of high-perceived effort (see “Rating 5” column). **ModernMT (orange bar)** is generally in the middle, except for the Human mobility, environment and space domain (SH7), in which ModernMT the highest number of cases of high-perceived effort.



DeepL has the highest number of 1-rated effort, while OpenNMT has the lowest. ModernMT is in the middle.

DeepL has the lowest number of 5-rated effort, while OpenNMT has the highest. ModernMT is in the middle.

A positive correlation is observed between the data collected from the different evaluators. However, given the very subjective nature of these judgements, their reliability could be further improved by extending the panel of evaluators.

CLIMATOLOGY AND CLIMATE CHANGE (PE3_10) - PERCEIVED POST-EDITING EFFORT

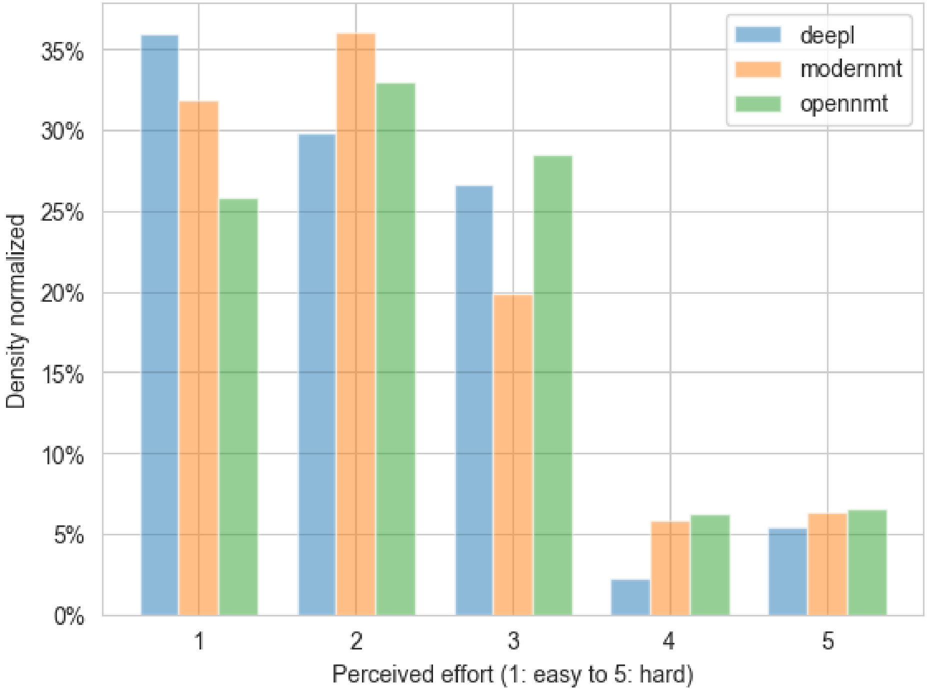


Figure 14: Perceived post-editing effort for the PE3_10 domain

NEUROSCIENCES (LS5) - PERCEIVED POST-EDITING EFFORT

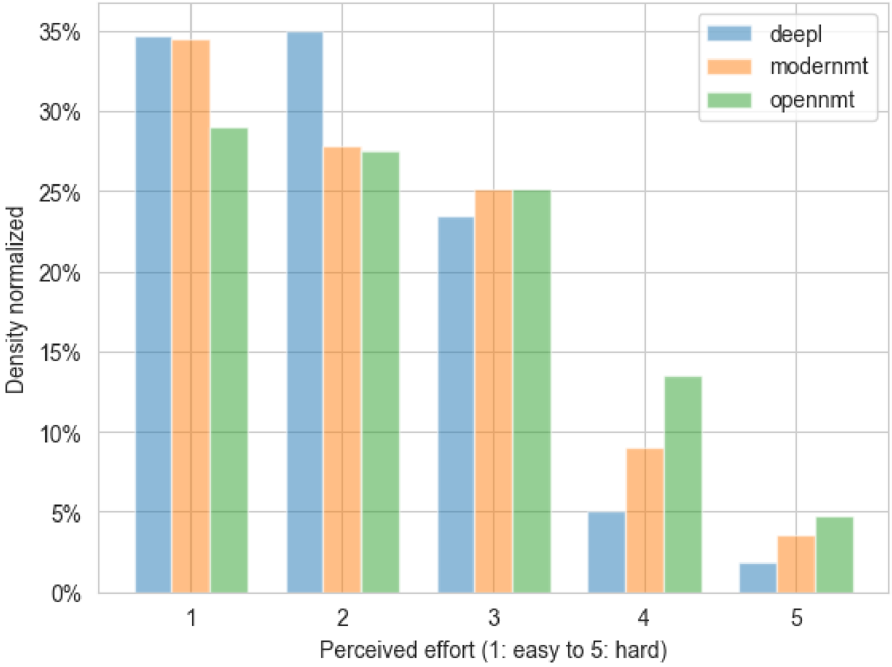


Figure 15: Perceived post-editing effort for the LS5 domain

HUMAN MOBILITY, ENVIRONMENT, AND SPACE (SH7) - PERCEIVED POST-EDITING EFFORT

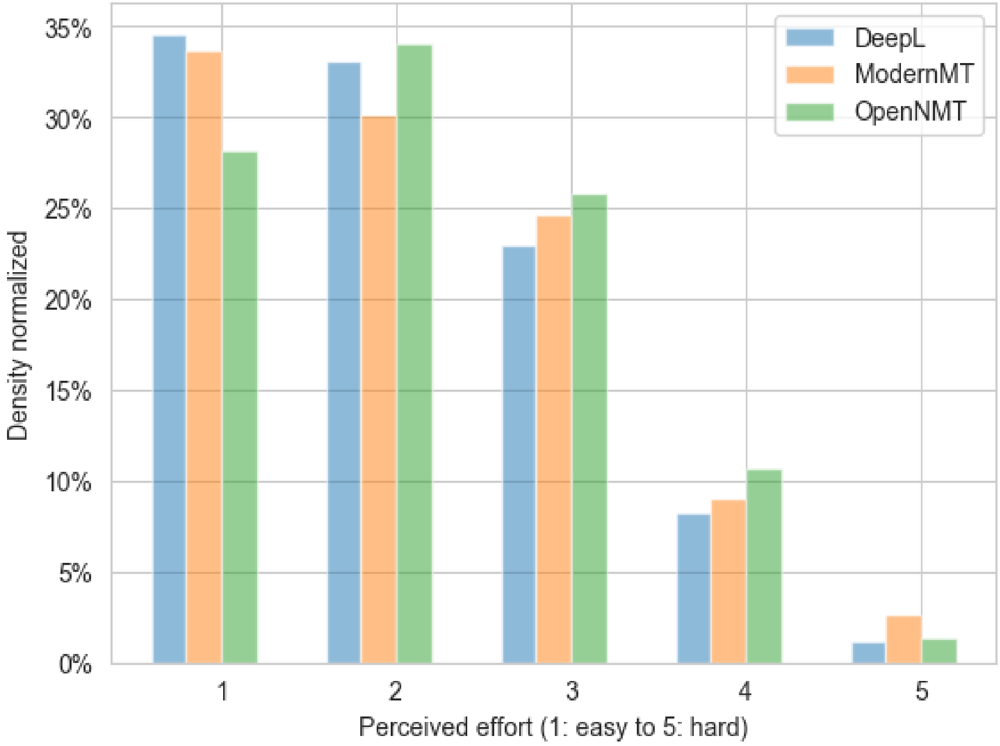


Figure 16: Perceived post-editing effort for the SH7 domain

3.3.3.3 Machine translation usability for gisting purposes

During the self-paced reading evaluation task, layperson readers were asked whether the machine translation output allowed them to get an idea of the content of the scientific text they read. For all the domains, more than half of all evaluators judged the machine translation output as sufficient. The detail of the assessments shows that in most of the cases the output judged as insufficient was produced by OpenNMT (36 times), followed by ModernMT (28 times), while DeepL output was judged as insufficient only 18 times.

An overview of the assessments for each domain is presented in the charts below. For further details, please refer to the reports *Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space*, *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*, and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[31].

[31] Ibid. footnotes 18-19-20

CLIMATOLOGY AND CLIMATE CHANGE (PE3_10) - EVALUATION OF MACHINE TRANSLATION USABILITY FOR GISTING PURPOSES

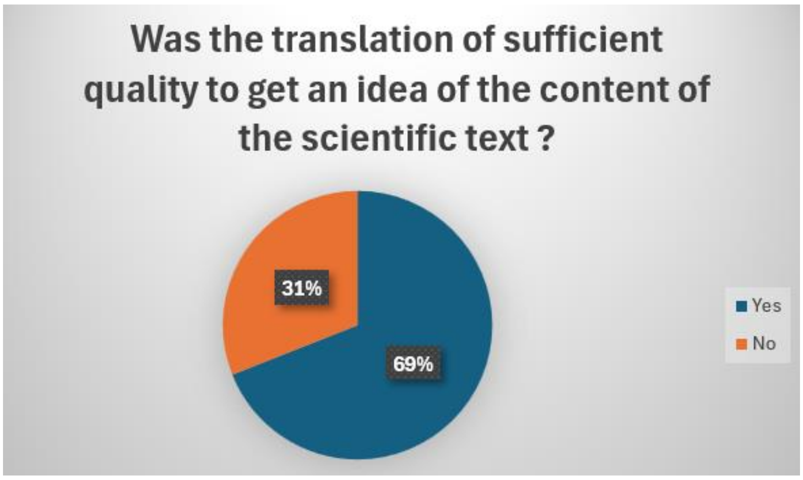


Figure 17: Evaluation of machine translation usability for gisting purposes for the PE3_10 domain (all machine translation engines combined)

NEUROSCIENCES (LS5) - EVALUATION OF MACHINE TRANSLATION USABILITY FOR GISTING PURPOSES

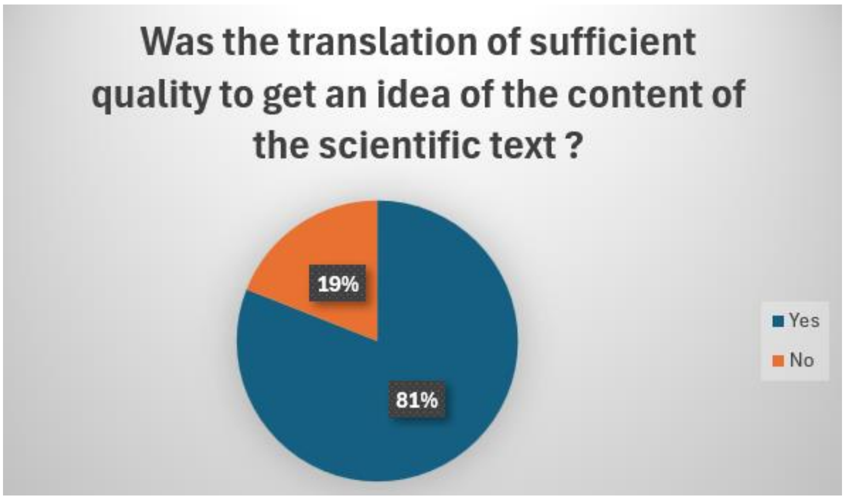


Figure 18: Evaluation of machine translation usability for gisting purposes for the LS5 domain (all machine translation engines combined)

HUMAN MOBILITY, ENVIRONMENT, AND SPACE (SH7) - EVALUATION OF MACHINE TRANSLATION USABILITY FOR GISTING PURPOSES

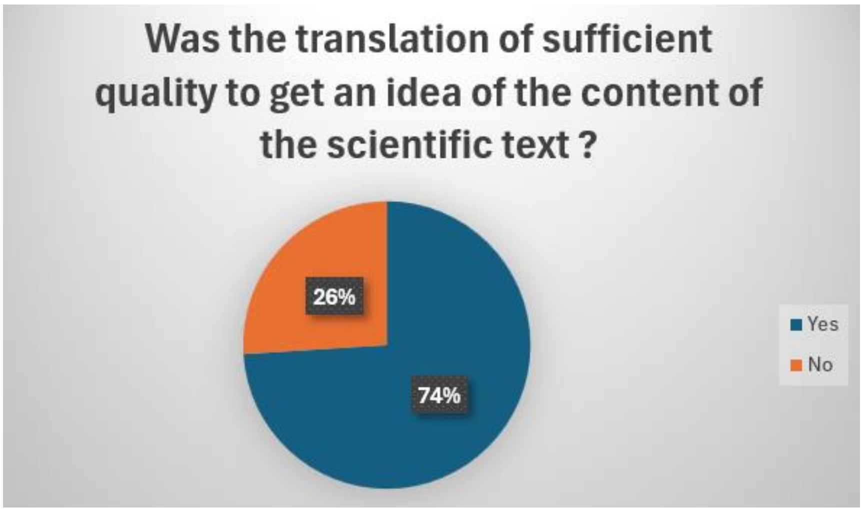


Figure 19: Evaluation of machine translation usability for gisting purposes for the SH7 domain (all machine translation engines combined)

3.3.3.4 Critical and terminology errors resulting from MQM annotation

The dataset used for the self-paced reading experiments was manually annotated to classify machine translation errors. MQM scores as well as counts of total errors, ratio of sentences with errors, terminology errors and critical errors were produced. The general ranking obtained from the scorecards and analyses confirms that DeepL scores better than ModernMT and OpenNMT.

With a view to assess machine translation usability for discoverability purposes, the charts below present the terminology error and critical error counts for each domain. For further details, please refer to the reports *Machine translation evaluation - Outcome for discipline Human mobility, Environment, and Space*, *Machine translation evaluation - Outcome for discipline Neuroscience and Disorders of the Nervous System*, and *Machine translation evaluation - Outcome for discipline Climatology and climate change*[32].

[32] Ibid. footnotes 18-19-20

Note on the MQM annotations

Multidimensional Quality Metrics (MQM) is a framework for analytic Translation Quality Evaluation (TQE). It provides a comprehensive list of error types, with standardised names and definitions, as well as a mechanism for applying them to translation quality analysis. Error annotation was performed by a single annotator based on the following error types: terminology, accuracy, linguistic conventions, style, locale conventions, audience, appropriateness, and design and markup.

The composition of the dataset used for the MQM annotation is as follows: PE10_3 domain: 64 segments, 1510 words; LS5 domain: 64 segments, 1840 words; SH7 domain: 80 segments, 1994 words.

CLIMATOLOGY AND CLIMATE CHANGE (PE3_10) - CRITICAL AND TERMINOLOGY ERRORS

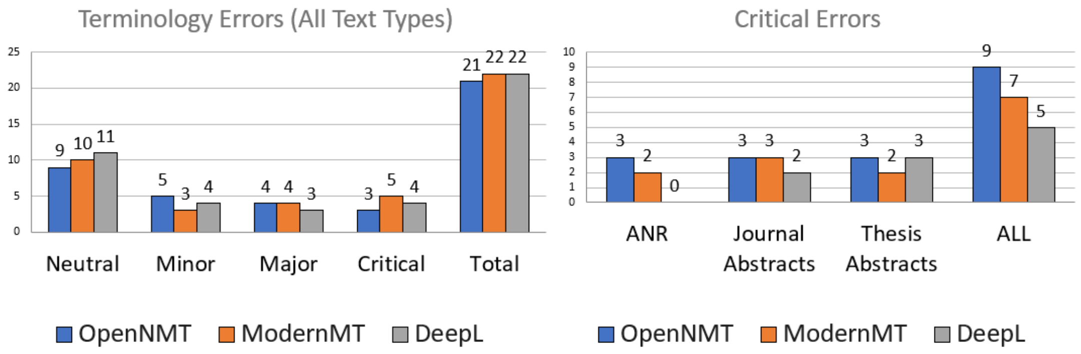


Figure 20: Critical and terminology errors for the PE3_10 domain

NEUROSCIENCES (LS5) - CRITICAL AND TERMINOLOGY ERRORS

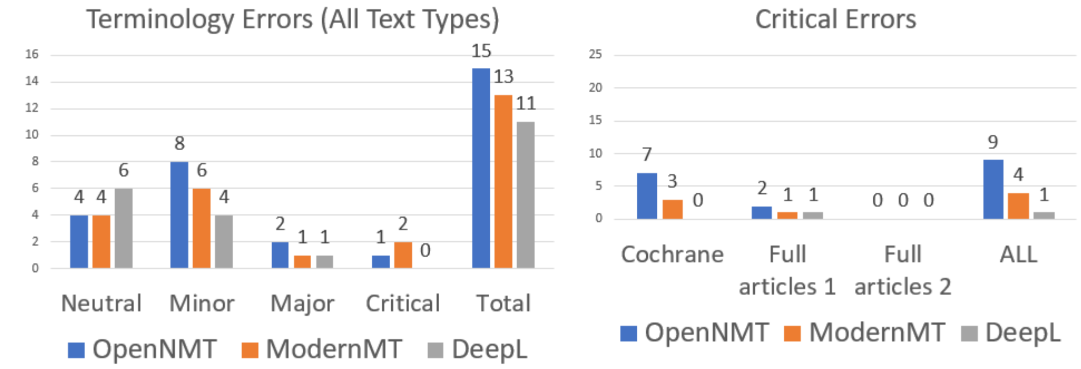


Figure 21: Critical and terminology errors for the LS5 domain

HUMAN MOBILITY, ENVIRONMENT, AND SPACE (SH7) - CRITICAL AND TERMINOLOGY ERRORS

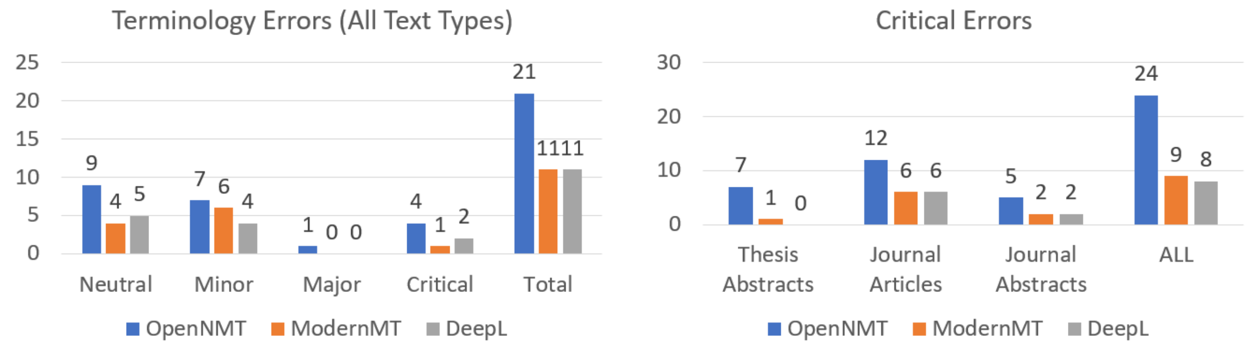


Figure 22: Critical and terminology errors for the SH7 domain

3.4 Conclusions from the study

At a first glance, DeepL appears to globally outperform ModernMT, which in turn outperforms in average OpenNMT for both the automatic evaluation and the human evaluation tasks. Besides the BLEU scores and human ratings detailed above, this ranking is confirmed by the total number of errors annotated for each of the engines: depending on the domain, the number of errors observed in OpenNMT raw output almost doubles or even triples compared to DeepL (+77% in PE3_10, +200% in LS5, +136% in SH7), against a more moderate increase in ModernMT raw output (+30% in PE3_10, +107% in LS5, +29% in SH7 compared to DeepL).

In conclusion, all the analysed metrics outline a ranking that leaves little room for doubts. However, these evaluation results should be interpreted by taking into account at least four key aspects that go beyond the mere question of machine translation output quality.

Quality of datasets: the collection of bilingual scientific datasets for machine translation fine-tuning raised various challenges. Among these, the difficulty to determine the origin of the translations collected, which means that it was not always possible to easily identify and exclude low quality-translations or *translationese* from the datasets. With this in mind, it is plausible to assume that, in the present evaluation, OpenNMT might have been more affected by the variable quality of fine-tuning data compared to the two other engines. Inversely, DeepL may have benefited from the extensive use over time of its free version by researchers, as also suggested by the use case study conducted as part of the project (see Section 1 of the present report).

Transparency and confidentiality issues: no information is available on which data has been used to train DeepL, nor on the way data have been preprocessed, the specific architecture of the engine or the hyperparameters applied during training and inference. Moreover, when texts are submitted for translation in the free version of DeepL, they are stored and reused to train the engine with no opt-out possibility for the user - which can also cause confidentiality issues when translating sensitive or protected documents. With all these elements in mind, it is difficult to submit the engine to a transparent and fair evaluation: its high scores may at least in part be attributable to overlap between test and training data. Moreover, the exact version of the translation engine is unknown. When performing a continuous evaluation of an engine, it is important to have a precise view of what has changed in order to know what variables are being measured. As for ModernMT, while its software is open-source, its users typically make use of an existing engine (model) provided online rather than training it from scratch. While such an online engine can be fine-tuned to a higher extent than DeepL, the same remarks apply as above: the details on baseline training data are not disclosed, there may be overlap between training and test data, and the exact version is unknown. On the other hand, with an open solution like OpenNMT, it is possible to ensure data transparency, and therefore the absence of overlap between training and test datasets. Moreover, the output can be continuously improved based on user feedback and thanks to an advanced management of parameters and variables.

Cost-efficiency and long term costs: it should be taken into account that training and maintaining over the long term an open solution as suggested above will require significant investments. As an example, funding will be needed in order to build or collect high-quality specialised corpora, as well as to deploy the technical efforts and the long-term vision required to fill the gap with the existing solutions. A few assets should however be highlighted. First of all, as shown by the use of the all-domain scientific corpus *SciPar*, the addition of scientifically oriented training data beyond the domains being studied can be beneficial for additional domains, providing potential for sustainability of an open translation engine. Secondly, the pricing model of commercial solutions may change over time, making the costs unpredictable over the long term. Finally, it should be considered that without investments into open environments, the existing solutions will continue to improve - including by relying on the opaque collection of data and user feedback - to a level that will be impossible to achieve with other solutions in the future.

Independence: commercial engines offer an increasing variety of features that meet a wide range of use cases and needs. While such features are ready to use and generally made available to users and organisations without requiring additional technical and financial efforts, they are developed, maintained and improved according to the vision promoted by the owner company. In the light of the specific translation practices and needs observed in scholarly communication and academic publishing, an open solution would make it easier to promote an independent, tailored development strategy, without relying on the business vision of a third party. An open engine may be in a timely fashion enriched with new languages, domains and features to take into account specific elements of the translation workflow (for instance, APIs to ensure interoperability, user interaction and feedback, language-specific components, etc.)

IV. Study No. 4: Operating model for a technology-aided collaborative translation service

Study overview

This section presents an overview of the study *Operating model for a technology-aided collaborative translation service*. The aim of the study was to identify technical and organisational requirements as well as suggest economic and ethical models for the deployment of the service.

In addition to the overview, more information from the study is available in the study report *Operating model for a technology-aided collaborative translation service dedicated to open scholarly communication*[33] (report in French).

4.1 Scope of the study

The picture emerging from the previous studies raised a number of challenges related to technical, organisational, legal, economic as well as ethical dimensions. These challenges are mainly related to the following aspects:

- **Building and maintaining specialised language resources of high quality and volume**: language resource collection, production and management must comply with legal, ethical and technical requirements, including respect of copyright, textual data diversification, and processing efficiency which can be compromised by format and translation quality issues. In order to meet these requirements and create the best possible conditions for building and maintaining specialised language resources, it seems crucial to collaborate with academic publishers and researchers so as to identify – and when necessary produce – multilingual data that complies with the expected criteria in terms of quality, volume, representation and transparency.
- **Implementing a sustainable machine translation solution:** the machine translation evaluation carried out in study No. 3 leaves little room for doubt regarding the gap between a hypothetical open engine and commercial solutions, be it in terms of output, training data and algorithms. There are also questions about cost effectiveness and environmental impact, since the creation of a new engine is expected to require more resources than relying on an existing one, at least in the short term. However, in the current commercial landscape it is difficult to find solutions complying with all the expected criteria of openness, independence, data protection and transparency.
- **Promoting a sustainable use of Artificial Intelligence in general:** sustainability encompasses much more than the environment. Promoting a sustainable use of AI means eliminating the linguistic and cultural biases that these tools tend to reproduce and even reinforce, valuing human skills rather than exploiting them,

[33] C. Talbot, R. Torres, 2023, Operating model for a technology-aided collaborative translation service dedicated to open scholarly communication. https://doi.org/10.5281/zenodo.10972976

protecting and remunerating expert intellectual work and user contributions, in particular by respecting intellectual property and guaranteeing adequate working conditions for the experts involved in the translation service development and the translation workflows in general.

- **Establishing a viable economic and operating model:** achieving all these goals comes at a cost, and these costs should be borne and shared as much as possible at the highest level (institutions, organisations, etc.) so that they are not imposed on users, according to open science principles. It seems therefore necessary to create a collective dynamic around translation tools, resources and activities in order to establish a sustainable and ethical model.

In order to address the above-mentioned points and to suggest a viable and sustainable operating model for the future translation service, the methodology described in Section 4.2 was followed.

4.2 Methodology of the study

The study consisted of two main phases: a series of design-thinking and collective intelligence workshops, followed by an analysis phase to refine the assumptions made during the workshops.

The aim of the workshops was to consolidate the usage scenarios and features identified in study No. 1, as well as to suggest the possible common interests and reasons for economic cooperation between stakeholders in order to establish a viable economic, ethical and legal model for the future translation service. The workshops were attended by a total of 13 participants. In particular the activities were organised as follows:

- 1 workshop on the “Researcher” and “Dissemination platform” usage scenarios, attended by 3 researchers, 3 technical experts from a dissemination platform, 1 university student, 1 legal expert, 2 facilitators;
- 1 workshop on the “Translator” usage scenario: attended by 3 translators, 1 legal expert, 2 facilitators;
- 1 workshop on economic models: attended by 3 translators, 1 publisher, 1 legal expert, 2 facilitators.

The assumptions developed during the workshops were then discussed and refined with 2 legal experts as well as with 2 AI and data management consultants.

4.3 Conclusions from the study

Given the scope of the challenges to be addressed, the hypothesis that emerged from the workshops was to establish a community-shared service, operating on the basis of common interests while serving at the same time individual practices and needs.

Thanks to the design-thinking and collective intelligence approaches leveraged in the workshops, the following picture was suggested.

4.3.1 General interests that the future translation service could serve

- Fostering knowledge and science dissemination in different languages and

contexts;
- Promoting sustainability in the use of Artificial Intelligence;
- Building and maintaining qualitative specialised language resources in order to promote a fair data approach and reduce AI bias;
- Diversifying scientific productions;
- Promoting collaboration and interactions between the stakeholders involved in scientific translation and multilingual scholarly communication.

4.3.2 Individual-user or profile-level interests that the future translation service could serve

The stakeholders involved in the workshops identified the following interests and benefits as possible drivers of user participation or contribution to the future translation service. The list is not exhaustive and further feedback could still be needed from the community.

4.3.2.1 Researchers' interests

***This category includes PhD and academic students.**

- Gaining visibility for their research, for example by taking part in exchanges and making contributions on the service platform;
- Saving time and benefiting from optimised content production thanks to specialised and contextualised translation tools and workflows;
- Saving time and benefiting from optimised content production thanks to the use of a specialised glossary, especially in fields that lack recognised multilingual terminology;
- Increasing their academic credit and impact thanks to the contributions to the language resources of the service.

4.3.2.2 Publishers' interests

***This category includes dissemination platforms.**

- Gaining audience, traffic and visibility;
- Increasing productivity and lowering operating costs through the use of shared reliable tools and resources;
- Encouraging authors to write multilingual content to reach an international audience.

4.3.2.3 Translators' interests

- Gaining visibility as users and contributors of the service, for example by helping improve language resources;
- Increasing productivity through to the use of reliable specialised tools and resources at a lower cost;
- Leveraging the service tools to take care of simple or repetitive tasks;
- Creating or consolidating a professional community;

- Promoting the value and ensuring a healthy future for the profession.

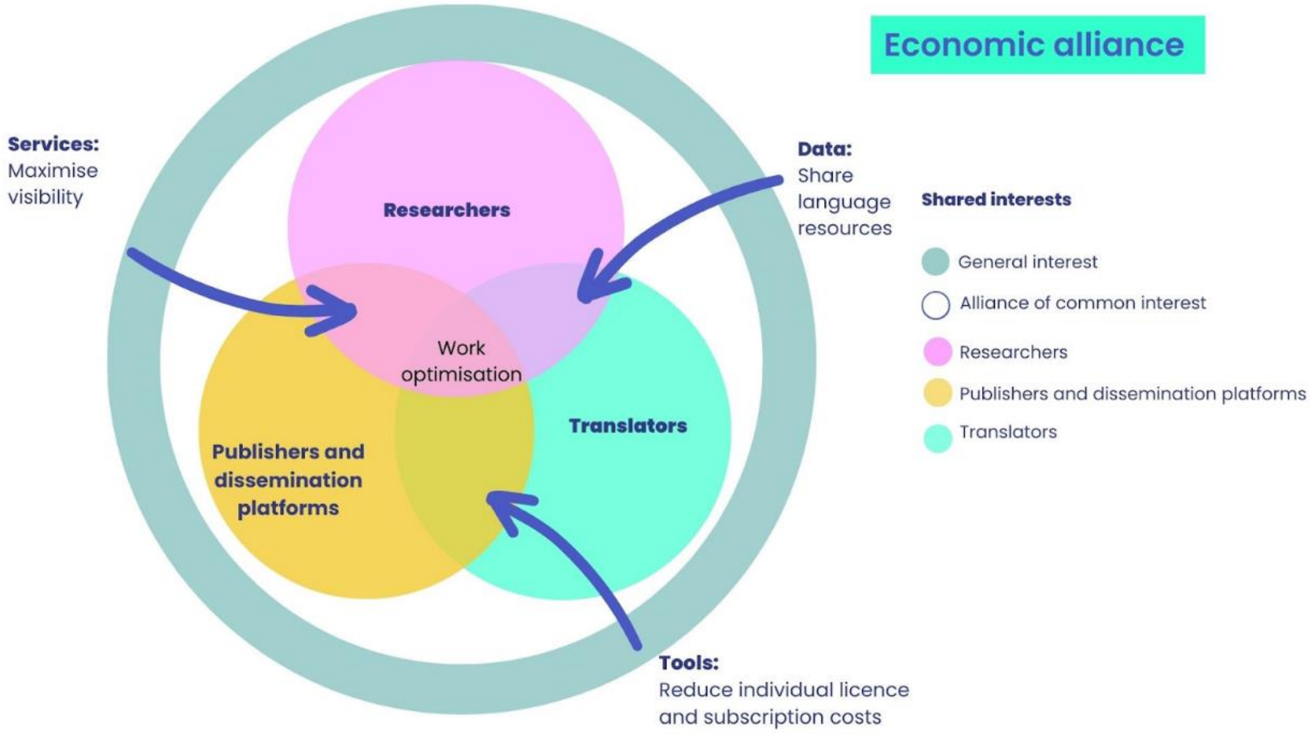


Figure 23: Symbolic representation of an economic alliance based on common interests between stakeholders

4.3.3 Proposed economic model

A freemium economic model could be imagined for the future translation service, which might therefore offer free access and features, paid access and features (mainly intended for organisations and institutions) and "contributor" access and features (free and/or subject to contributions from users, in particular intended for expert users such as researchers and translators).

4.3.3.1 Free usage

- Basic use of the machine translation engine;
- Access to the community directory.

4.3.3.2 Paid usage

Free usage features, plus

- Access to academically-validated translation memories and glossaries;
- Access to forum and collaborative features;
- Presence in the community directory.

4.3.3.3 "Contributor" usage (free and/or subject to contributions)

Free and paid usage features, plus

- "Contributor" features allowing users to contribute to translation memories and glossaries, as well as to peer reviewing processes;
- Traceability of personal contributions in the translation memories, glossaries and peer reviews, so as to ensure more visibility for the contributor and offer academically-validated resources to the other users;
- Statistics on the use of these translation memories and glossaries (ex: how many times they are consulted);
- Reward for contributions to translation memories and glossaries, according to the volume of the contributions;
- A "Contributor" status which could be confirmed every year based on the volume of annual contributions. Such volume could therefore be limited with minimum and maximum thresholds in order to ensure equally distributed and represented contributions.

4.3.4 Roadmap for the development of the service

The deployment of the service could be planned according to the 4 milestones below:

M0 – Demonstrator: The demonstrator could present the main features of the service, such as an existing machine translation engine, a preliminary structured data collection workflow, as well as a legal framework for data processing. This milestone could also include the study of the architecture required to support the overall system.

M1 – First version of the service: In the first version of the translation service, the database could be aggregated, the underlying system deployed and secured, and the first API made available. This first version could target the support of 1000 daily users and 2 additional languages.

M2 – Second version of the service: The second version of the translation service could anticipate the gradual increase in the number of daily users, segments in the translation memories, and daily contributions. The target could be to support of five major European languages.

M3 – Third version of the service: In this version, the translation service could reach 50,000 daily users, 100 million segments in the associated translation memories, 1 million daily contributions, and ten main languages to cover the 5 continents.

The collaborative platform mentioned in Section 1.3.2 (see footnote 7) should be developed in parallel with the translation service.

General conclusions

The four exploratory studies globally confirmed the interest of a collaborative, technology-aided scientific translation service to address a variety of needs in multilingual scholarly communication. In particular, the studies highlighted the importance of sharing and promoting best practices, resources, tools, and skills relating to scholarly translation and multilingual content production. In this context, translation technologies certainly have a role to play but given the complex nature of scholarly communication, such role must be carefully defined, while the community and their skills must remain at the heart of the translation service.

As previous research has shown[34], Artificial Intelligence has the potential to be an enabler of the Sustainable Development Goals identified by the United Nations[35]. The *Translations and Open Science* project is inspired by some of these goals, such as reduced inequalities and more inclusive access to information. However, AI fast development is becoming difficult to track, including for specialists, and therefore requires greater caution than ever when it comes to evaluating risks, in particular relating to quality and ethical standards. Regarding AI-based translation technologies, for example, a quality-risk matrix could be defined as in the following table.

[34] Vinuesa, R., Azizpour, H., Leite, I. *et al*. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 (2020). https://doi.org/10.1038/s41467-019-14108-y

[35] Sustainable Development Goals | Division for Sustainable Development Goals (DSDG) in the United Nations Department of Economic and Social Affairs (UNDESA)

Translation process	**Translation by language experts**	**Post-edited machine translation**	**Unsupervised machine translation**
Risk level	**Lowest risk of error**	**Medium risk of error**	**Highest risk of error**
Process details	Specialised language experts deliver a translation by leveraging a technology-aided process when relevant.	Machine translation is generated by an AI-technology and is fixed with basic intervention.	Machine translation is generated by an AI-technology and is used with no supervision.
Process result	Subject to the conditions under which a translation is made (translation skills and domain expertise of the language expert, time and resources available, quality of the source text, etc.), translation is expected to contain no or very few errors and inaccuracies. Thanks to their specialised skills, language experts are able to understand the meaning of the source text in all its nuances and to convey them into the target text, while also respecting domain terminology. Language experts can also provide advice and leverage relevant translation technologies according to specific translation contexts.	Translation may still contain errors and inaccuracies due to machine-translation subtleties, biases and priming effects. Subject to the conditions under which post-editing is performed (post-editor's domain and language expertise, time and resources available, post-editing cognitive effort, quality of the source text, etc.), this type of process should result in no or few **critical** errors in the target text. However, the result is not comparable to a qualitative translation made by a language expert.	Translation may contain errors, including critical ones, especially relating to meaning and specialised terminology. Users should be clearly warned about these risks. Failure to do so could seriously compromise the image of the author, the publisher or the dissemination platform of the publication.
Possible use cases	Translating highly visible or specialised publications, for which the highest quality is required.	Producing multilingual abstracts, metadata and other formats in order to make research publications internationally discoverable and readable.	Gisting a research publication, i.e. a reader wants to have an idea of the content of the publication.

Table 7: Proposal of quality-risk matrix to understand translation technologies, processes and outcomes

For the next steps of the project, it seems therefore recommended to focus on risk- and cost-benefit ratios in order to develop a service that is truly useful and sustainable for all the stakeholders in the research community and society at large.