This White Paper has been prepared by the OPERAS Multilingualism Working Group

Coordinator:  
*University of Coimbra — UC Digitalis*  
Delfim Leão

Participants:  
*Georg-August-University Göttingen (UGOE)*  
Andrea Bertino

*Huma-Num*  
Suzanne Dumouchel

*National Documentation Centre (EKT/NHRF)*  
Marina Angelaki

*University Institute of Lisbon (ISCTE-IUL)*  
Frédéric Vidal

All OPERAS White Papers: [operas-eu.org/aboutoperas/working-groups](http://operas-eu.org/aboutoperas/working-groups)

OPERAS Multilingualism Working Group contact: [leo@fl.uc.pt](mailto:leo@fl.uc.pt)

OPERAS website: [operas-eu.org](http://operas-eu.org)

OPERAS contact: [pierre.mounier@openedition.org](mailto:pierre.mounier@openedition.org) and [suzanne.dumouchel@huma-num.fr](mailto:suzanne.dumouchel@huma-num.fr)

CC BY 4.0
EXECUTIVE SUMMARY

Scholarly publication is indisputably boosted by the use of the English language. However, the need to publish in English in order to get visibility and recognition represents an impoverishment of certain research fields, particularly in Social Sciences and Humanities. Taking this backdrop as reference, the challenges for OPERAS are to support researchers that want to continue publishing in their own language and to develop transnational scientific cooperation at the same time. Thereof, the proposed intervention areas are: translation, multilanguage discovery tool and the endowment of national languages. An overview of the situation respecting the OPERAS partners is provided, and suggestions are made in order to make their joint work more convergent with emerging trends.
1. SCOPE

**Premise:** Research dissemination is undeniably boosted by the use of the English language. However, language is not a neutral medium, but something that sets limits and possibilities for the scientific thinking and for scholars’ communication. In addition, the choice of a language system often implies the choice of a frame of references, of a methodology, of a school. Therefore, the usage of a *lingua franca* implies for a non-native speaker not only an impoverishment of its expressive means, but it can also be misunderstood as closeness to a certain field of research.

If a simply and poorly articulated English language spoken by a non-native speaker can be suitable for conveying research results in mathematics and the natural sciences, the humanities and the social sciences require specific means of expression because language in those disciplines is not only a communication tool but is entangled into the object of study itself. Therefore, an important part of the work in those disciplines is to reflect upon words and their meaning, both in the way they are naturally used and in the way they might be used as conceptual tools to elaborate a theory.

On the other hand, some specific traits of a language or culture (as happens quite often in scholarly work dealing with linguistics and literature) are very difficult to transpose in their full entirety to a different language, thus inhibiting the clarity or the scientific depth of a work (Cassin, 2012).

It has hence been recognized that “the overwhelming dominance of English as a *lingua franca* in the academic domain is having an insidious effect upon other languages, leading to the curtailment or erosion of their traditional scholarly discourses” (Bennett, 2013: 169). In addition to this, there is as well a risk of “self-perpetuation” of English as the scientific language *par excellence*, because, as is pertinently stated by Flowerdew (2013: 302) “the greater the number of scholars using English, the more research can be disseminated (to researchers who know English), and the more that research is disseminated in English, the more scholars will be encouraged to publish in English”.

Therefore humanities are at odds with the search for the highest possible dissemination of research results by a *lingua franca* on the one hand, and the usage of different national languages on the other hand, which would allow a richer, more rigorous and refined articulation of one’s own thinking.
2. STATE-OF-ART

As was highlighted in the scope section, a *lingua franca* risks to be perceived more as a necessary evil rather than a real resource. One might even suspect that the emancipation from the ideal of a *lingua franca* could unleash the creative potential of disciplines that are otherwise often accused of following the consolidated mainstreams too much. Something like what happened in the early modern era, when the progressive emancipation from the Latin language has led to an explosion of many original views and ideas formulated in different national languages.

To take a striking example, the recent [INTERCO-SSH project](#), which studies internationalization of SSH, found that, despite the growing importance of English as a communication language in social sciences and even humanities, the need for academic publications in native languages remains central in many cases. According to the same project findings, most of the time the internationalization of SSH does not mean going from local to global, rather it goes through what is described as transregional integration reflecting the structure of scientific networks that connect researchers across national boundaries and not always globally.

In order to support disciplines that want to continue publishing in their own languages and to develop transnational scientific cooperation at the same time, at least three scenarios are conceivable in which OPERAS can play an important role:

2.1 TRANSLATIONS

Open access indirectly supports the translations into different languages, as there are no longer any rights and licenses to pay for the translation of the work. OPERAS could accompany activities of advocacy for open access with targeted support for translation activities. One can imagine a kind of mediation aiming at stimulating the interaction between authors of open access publications and translators.

In this scenario, OPERAS could set up appropriate boards to provide scientific, technical and financial advice for the realization of translations. These scientific committees could, for example, select and evaluate translation proposals for open access publications: this would be something like a kind of peer review of translation projects. They could then assist authors and/or translators in finding funds for translations.

If this path proves to be promising, OPERAS could provide support in order to create a digital platform for submitting translation requests, bringing into contact the authors of OA works and their potential translators. Particular attention should be paid to the translation of contributions from European humanities into Chinese, since China manifests a growing interest in European humanities, which is also signaled by the fact that European scholars from humanities receive more and more professorships there.

On the other hand, a pendular movement could also be stimulated by promoting the translation into European languages of works written in less accessible languages, taking as reference the research networks and collaborative projects that are being developed. Europe has unique conditions to stimulate a global movement for the fostering of multilingualism, both to disseminate its own scientific and cultural matrix, as well as to give visibility to experiences developed in other geographic and cultural environments. In fact, multilingualism is a powerful agent of moderation and of inclusion.
2.2 MULTI-LANGUAGE SEARCHING TOOL

A second area of intervention would be the creation of a multilingual search tool, so that when someone looks for a concept in one language, that person may be referred to the same concept in other languages within a corpus of works. Example: when looking for ‘Freiheit’ one may be stimulated to find as well ‘freedom’, ‘liberté’, ‘libertad’, ‘liberdade’, etc.

Such a process presupposes the development of collaboration procedures between author and publisher, since the creation of a multilingual ontology is hardly conceivable as being fully automatable. It is about developing a workflow in which the publisher asks the author to provide a list of the most important terms in his/her work - just like the index of concepts at the end of a traditional monograph. Then, publishers and/or authors should indicate references in other languages for those concepts. In some cases, this could perhaps even be automated, but in many situations the operation will require a hermeneutic effort and ultimately a conscious decision on the part of the author. Example: ‘Geist’ --> ‘mind’ or ‘spirit’; ‘saudade’ --> ‘homesick’ or ‘nostalgia’?

This semi-automatic indexing and disambiguation service would allow a more precise cross-linking between works in different languages, thus constituting a very powerful resource that would add real value to Open Access digital monographs and stimulate original and comparative thinking.

Furthermore, the creation of open linked data and, more in general, the transformation of the internet into a semantic web, open up new perspectives for multilingualism both in research and publishing process: open linked data allow a “better data integration than existing models of linguistic data, due to the ecosystem of tools provided by the Semantic Web, such as query and federation.” (McCrae, Moran, Hellmann & Brümmer, 2015: 315). There is a wide range of multilingual and monolingual data that can be linked together in the context of multilingual research and publication environments: corpora of texts, lexical resources, language descriptions, inter-language comparisons and — of particular importance for OPERAS activity in support of scholarly publishing — “resources, which capture metadata about language resources.” (ibid.) OPERAS should therefore explore how the development of adequate architectures of open linked data could improve harvesting and collection of metadata descriptions of records in multilingual archives.

2.3 ENHANCE THE PROMOTION OF NATIONAL LANGUAGE SCHOLARLY LITERATURE

With very few exceptions — like the case of the “Loi Toubon” (France, 1994, version consolidée au 16 juillet 2018) and “The Declaration on a Nordic Language Policy” (Nordic Council of Ministers, 2007) —, governments and institutions are for the most part not explicitly involved in promoting national languages as alternative valuable modes of scientific expression. This absence of high-level programmatic involvement, combined with global driving forces that affect everyday options of the scholarly community (e.g. the quest for international prestige, academic recognition, competitive funding), lead to a general expression of resignation towards the usage of English as predominant scholar language (and increasingly as the only language fully recognized as plainly scientific).

The multilingual and multinational nature of OPERAS places itself in a central position to promote multilingualism as a key-concept to enhance different ways of perceiving the world, of stimulating originality and ground-breaking ideas. In fact, it has been recognized that “multilingual publishing is thus seen as a way to protect both national languages and English and to sustain the diversity of academic rhetorical traditions” (Kuteeva & Mauranen, 2014: 3).
However, multilingualism must not be envisaged as implying a kind of programmatic opposition between English vs other languages. It is not English that must be blamed for being efficient as lingua franca, but other languages that have to find new ways of efficiently reaching new publics, wider audiences (or at times of giving voice to local audiences as well) in a complementary way.

An example of that kind of collaborative process can be perceived in the usage of English for more impacting academic purposes and of the mother tongue as being more prone to public dissemination (Bocanegra-Valle, 2013: 20-21). The latter has been perceived as a less important indicator, but things may change drastically: “if for instance knowledge dissemination to the public becomes a prime factor in securing funding, writing for the international community of scholars may have to take the back seat” (Kuteeva & Mauranen, 2014: 3). This is a particularly fertile field of operation to OPERAS.
3. EMERGING TRENDS

Multilingualism and internationalization used in the making of web sites are not equivalent concepts, and they may operate independently, as is made clear by the World Wide Web Consortium (W3C). Despite this, they can be combined at different levels in order to reach more embracing results, and as an enlightening illustration of the way regional and cultural differences can be detected in the manner the same information is displayed in different languages, regardless of their degree of complexity.

One can find as well, relatively often, simple solutions developed in order to promote multilingualism, as is the case of a blog curated by the National Library of Wales, which publishes postings only in the original language, and although it displays an equal number of blog posts in Welsh and English, thus keeping a balance in the usage of both languages, it is made clear that “they are not the same postings”, and that “for a translation of the blog readers may wish to try facilities such as Google Translate”. This is an easily deployable way of promoting multilingualism and of stimulating linguistic intersection, even if there are limitations to the quality and reliability of the translation thus obtained.

In what pertains to multilingual tools, LINGEA is an interesting case, which has been, for over 20 years, developing language tools and collecting linguistic data, thus acquiring a very relevant experience, visible moreover in the fact that it is available in thirteen different national websites. Its multilingual search tool is particularly promising, as it “automatically translates information in database, picture captions, product titles, or key words, into preselected languages”. It is as well capable of expanding the query into other languages, although it faces the limitation of not being suitable for the translation of complete sentences or when a higher degree of comprehensibility is needed.

Neural machine translation seems to be a promising and indeed already powerful technology. For example, the translations made by Deepl are at times already very impressive (specifically German-English; Italian-English less good). Attention should be paid to the development of such translators, possibly as tool for "working translations" which can facilitate (international) peer reviews of manuscripts that are not written in English. In particular, this could be very helpful in order to have a referee on global content (using a working translation and not a publishing translation). In the case that the content proves to be valuable, this first referee could be combined with a second one directed not only to content but also to the quality of writing, thus keeping the possibility of publishing in the original language and not necessarily in English.

Although directed to methods and tools respecting digital humanities, and not primarily to the question of multilingualism, the OpenMethods project has been developing a very promising activity respecting this issue. It is a new DARIAH initiative created by the Humanities at Scale project in cooperation with OPERAS. Basing on the contribution given by the editorial team, whose members come from different countries and combine different expertises, the platform has been able to select and enhance the visibility of a very significant amount of material available online (blog posts, articles, expert reports, etc.). The OpenMethods editorial team deals with content written in 15 different languages, which undergoes a preliminary evaluation and, when it proves to be worthwhile, is displayed through the platform, preceded by a short introduction in English, prepared by the editors. This has shown to be very efficient when it comes “to highlight and promote especially valuable multilingual and multidisciplinary open access content in the field of Digital Humanities Methods and Tools”. Despite these promising achievements, it still depends a lot on the time availability of the editorial team, and therefore not only faces limitations in what respects regular feedback, but also the risk of not being able to grant long-term continuity.
4. OPERAS SITUATION

The partners in the OPERAS consortium are usually receptive to the question of multilingualism and globally aligned with the main emerging trends. As a whole, their clearest weakness is that initiatives are, for the most part, directed to internal demands or to specific needs deriving from national and institutional realities, thus lacking the means to reach the long-range impact that would derive from a programmatic and interconnected approach to this issue. The challenge and opportunity posed to OPERAS is therefore to take the key-areas of intervention identified in the “State-of-Art” section and turn them into a strong coalescent agenda. In what follows, an overview is provided respecting the situation of OPERAS partners.

COIMBRA UNIVERSITY PRESS – UC DIGITALIS

In what respects the scenario “2.1 Translations” on page 6, this is the kind of dynamic collaboration that is being put into practice very actively by Coimbra University Press and the authors it represents. The UC Digitalis platform allows them to have a closer idea of the works that are attracting more attention and which languages could be used in order to sustain a more efficient publishing and dissemination strategy. English naturally attracts attention, but there has been interesting experiences with other languages, like Spanish, Italian, French, and German, besides obviously Portuguese.

Until now, there is no project implemented respecting directly scenario “2.2 Multi-Language Searching Tool” on page 7. Even though, recent discussions seem to point somehow in that direction, as is the case with the planning to develop onomastics online tools pertaining to the Portuguese nouns taken from Classical works that are more commonly used in other languages. This field was selected because Coimbra University Press works as a publishing hub for Classical Studies in the Portuguese speaking countries. The first insights in that direction have proven to be an interesting path to follow, and thereby practical results are expected probably during 2019.

It is, however, with scenario “2.3 Enhance the Promotion of National Language Scholarly Literature” on page 7 that palpable results can be perceived in a very positive way. In fact, Coimbra University Press publishes quite often multilingual works and this has clearly given added value to the publisher. Besides Portuguese and Spanish, the more frequently used languages are Italian, French, German and English, quite often combined within the same volume. This option has attracted proposals of reputed scholars from different countries who cherish the opportunity of publishing in their native tongue. A complimentary phenomenon is also detected, involving especially works written in Portuguese (Leão, 2015), but also languages and literatures that share a common cultural heritage (as happens mostly with Spanish, Italian, and French): it is a trend with growing importance to take previously published works in one’s mother tongue and to combine them in a more comprehensive study, published this time in English and directed to a wider audience, frequently as well (co)published in a bigger international publisher. This has proven to be a balanced way of keeping multilingualism as an operative concept in the academic field, combined with the option to use as well a lingua franca, but not in exclusive terms.
EKTE PUBLISHING – SERVICE PROVIDER

EKTE ePublishing is supported by a Greek public institution, yet the 32 journals hosted in the platform provide access to content in four languages (Greek, English, French and German) with content in Greek and English accounting for the majority. Quite often articles in different languages are hosted in the same volume.

The majority of journals favor submissions in English in an attempt to increase their outreach to non-Greek audiences. Editors are requested to provide metadata in English in addition to a second language. Abstracts and keywords for each article are also provided in English. During the journal set up, English is one of the languages selected.

GÖTTINGEN UNIVERSITY PRESS

Göttingen University Press is the self-contained publishing facility of the Georg August University, Göttingen. It publishes documents from scholars who are associated with the Göttingen University. It gives assistance to the authors to publish their works electronically and as printed copies. Its aim is to offer free access to as many publications as possible and to publish reviewed high-quality books that due to their specialised nature will not easily be accepted by commercial publishing houses.

With regard to the topics covered in this white paper, there are some practices carried out by Göttingen University Press that may be of interest to the OPERAS consortium:

- Use of OA licenses (CC-by), so that the right to translation is granted.
- Publisher's website is made available in two versions, one in German and the other in English.
- Different publication languages are accepted as long as peer review is possible and scientific quality can be guaranteed.
- Where appropriate, multilingual abstracts are published and authors are encouraged to write summaries in a different language which is relevant for the specific research topic.
- Use of OAI-PMH and Dublin Core Qualified metadata with language specification to facilitate mapping and harvesting.
- With regard to the interrelationship between specialist language and non-expert language: a book collection, "Varia", has been made available for the publication of works directed to a wide audience.

KNOWLEDGE UNLATCHED – SERVICE PROVIDER

While Knowledge Unlatched focuses primarily on obtaining institutional funding for English language collections, there are currently two open access collections, one in French and one in German:

- OpenEdition Select 2018: a collection of 30 French language books in SSH from a variety of publishers. The collection follows the KU Select model, where a group of librarians review the longlist of submissions and selects the most relevant titles. OpenEdition Books Select is the first crowdfunding program for open access development in French-language scientific publishing. The program is led by OpenEdition, the French non-profit initiative promoting innovative and fair open access models, supported by the main French research institutions.
Transcript OPEN Political Science: a collection of German language books in Political Science from transcript Verlag. The pilot invites academic libraries to help take collaborative measures in the move towards Open Access. The complete frontlist of 2019 titles (i.e. all new titles in the field of Political Science) will be bundled into one package, very similar to the established practice of “eBook collections”. Rather than individual libraries purchasing eBook licences, the crowdfunding model means that libraries collectively finance the full frontlist production being made available in Open Access.

LEXIS

Lexis (full name “Lexis Compagnia Editoriale in Torino”) is a commercial company based in Italy which makes business by (1) providing services to the publishing industry, and by (2) publishing on its own account through some brands, owned by or licensed to Lexis, in the non-fiction, mainly academic, sector.

With regard to the editorial production in the SSH academic field, multilingualism is widely accepted which means that in journals and collective works texts are published systematically in the original languages (apart from Italian, especially in English, French, German and Spanish) and are not translated. It is also quite common to publish single author monographs in the original language. However, we record a growing demand by authors to gain international visibility and diffusion of their scientific production, which sometimes results in the decision of direct publishing in English. Abstracts and keywords in English are the standard.

Considering economic cost and quality issues, we would look with great interest to instruments such as networks with other publishers and platforms to share translation services, proofreading and reviewing by native professionals.

OPENEDITION

OpenEdition, although based in France and supported by French public institutions, disseminates content in 14 languages on its different platforms. OpenEdition has been developing an internationalization program since 2012 which resulted in an increase of multilingualism. Currently, around 30% of content is in another language than French, including English. The internationalization strategy is based on the development of partnership with academic institutions and professional organizations in different countries (Portugal, Italy, Spain, Germany, Poland, so far). The management of a truly multilingual platform in humanities and social sciences is particularly costly and entails the ability to address issues at several levels:

– Community management to engage different types of academic stakeholders in the different countries (researchers, funders, libraries, publishers): community management relies on relations of proximity which means localization of staff and specific coordination tools to facilitate the circulation of information between the central technical team and the end users through national community managers

– Interface and metadata translation: OpenEdition policy regarding metadata is that journals and book publishers provide abstracts and keywords in at least one other language than the original language of the publication, usually English. To support usage of the platforms within the different linguistic communities, the public interface of the platforms has been translated into Portuguese,
Italian, Spanish, German, English. The most difficult part of this is not to add another language, but to be able to maintain different languages because it multiplies the work each time the platform is upgraded.

- Multilingual training ability and documentation: the hardest part of OpenEdition’s internationalization program is the difficulty to work with content producers, namely journals and book publishers, and to coordinate with them for the preparation of content accordingly to the platforms requirements. It supposes a sound understanding of the publication system that can be obtained only through intense training and solid documentation. In that domain, being able to deliver documentation and training in many different languages has a high cost.

In a few words, OpenEdition strategy fits well into scenario 3 mentioned earlier but could be rephrased differently as supporting the diversity of languages through infrastructure services. The cost and the difficulty of this endeavor should not be underestimated as it requires continuous effort.

**OPEN LIBRARY OF HUMANITIES**

The Open Library of Humanities publishes several titles that are multilingual. It has had Italian submissions, for instance, to the OLHJ, it sponsors the *Journal of Portuguese Linguistics* and *Francosphères*, and it publishes the multilingual *Digital Studies / Le champ numérique, Le foucauldien,* and *ZFF Zeitschrift für Fantastikforschung.*

In addition, the Centre for Technology and Publishing at Birkbeck, University of London, where the OLH is based, created a piece of software for live dynamic translation of web pages, using the hypothesis annotation platform as its underlying engine. This software is called annotran and is freely available.

**UCL PRESS**

UCL Press is the university press for University College London. It is a fully open access press, which was launched in 2015. UCL Press publishes all its books and journals in English. Because UCL Press authors grant UCL Press a non-exclusive licence to publish, they are free to arrange translation agreements with publishers in other countries. UCL Press does not proactively seek translation rights agreements with other publishers, rather it responds to enquiries as they arise. While it is up to the author to make arrangements with other publishers, UCL Press provides extensive advice to authors in navigating often complex arrangements, and UCL Press usually checks the author contracts and the copyright pages for works in translation — more often than not, foreign language arrangements are with publishers who do not offer open access.

When the author is able to arrange translation themselves but has no other publishing arrangement in place, UCL Press is happy to publish the translations. This extends to making the work available on its own website and distributing it to other open access platforms, as well as making the PoD available for sale. It does not extend to local sales and marketing. It is only in the minority of cases that the author is able to make an arrangement with a foreign-language publisher, and this usually only applies to particularly popular books that have the potential for international appeal. Similarly, it is only very rarely that the author will have the funds available to arrange for a translation themselves.
LIST OF ACRONYMS AND ABBREVIATIONS

CC: Creative Commons licenses

DARIAH: Digital Research Infrastructure for the Arts and Humanities

INTERCO-SSH: The INTERCO-SSH project sets out to assess the state of the Social Sciences & Humanities (SSH) in Europe and to understand the factors that facilitate or hinder international exchanges.

OA: open access

OPERAS: European research infrastructure for the development of open scholarly communication, particularly in the social sciences and humanities.

SSH: Social Sciences & Humanities

W3C: The World Wide Web Consortium is an international community where Member organizations, a full-time staff, and the public work together to develop Web standards.
REFERENCES


ANNEXES


MULTILINGUALISM

Premise

Research dissemination is undeniably boosted by the use of the English language. However, language is not a neutral medium, but a means to set limits and possibilities for the scientific thinking and for scholars’ communication. In addition, the choice of a language system often implies the choice of a frame of references, of a methodology, of a school. Consequently, the usage of a lingua franca implies not only an impoverishment of expressive means, but also the narrowing of the reach of certain research fields. This is particularly sensitive in the area of Social Sciences and Humanities. A lingua franca can, therefore, be seen as a necessary evil rather than a resource. One might even suspect that the emancipation from the ideal of lingua franca could unleash the creative potential of disciplines that are otherwise often accused of simply following the consolidated mainstreams.

Challenges for OPERAS

to support researchers that want to continue publishing in their own language and develop transnational scientific cooperation at the same time.

Intervention areas

1. Translation

Open Access indirectly supports the translation into different languages, as there are no longer rights and licenses to pay for the translation of the work. OPERAS could accompany initiatives of advocacy for mediation between authors of Open Access publications and translators.

2. Multilanguage Discovery Tool

The creation of a multilingual search tool, so that when someone looks for a concept in one language, that person may be referred to the same concept in other languages within a corpus of works. Such a process presupposes the development of a collaboration procedure among authors and publishers, since, until now, the creation of a multilingual ontology is not fully automatable.

3. Protection of National Language Scholarly Literature

Most governments and organizations are not explicitly engaged in promoting national languages as alternative valuable modes of scientific expression. This absence of high-level programmatic involvement, combined with global driving forces that affect everyday options of the scholarly community, lead to a general expression of resignation towards the usage of English as predominant scholar language.

Conclusions

OPERAS is in the frontline of a broad, international strategy to leverage multilingualism in scholarly production and research dissemination – thereof, also recognizing other languages as plainly scientific. Multilingualism should not be envisaged as implying a kind of programmatic opposition between English and other languages, but rather as complementary. The cost and the difficulty of this endeavor should not be underestimated as it requires continuous effort.