



Translation and the New Digital Commons

Philippe Lacour, Any Freitas, Aurélien Bénel, Franck Eyraud, Diana Zambon

► To cite this version:

Philippe Lacour, Any Freitas, Aurélien Bénel, Franck Eyraud, Diana Zambon. Translation and the New Digital Commons. Tralogy I. Métiers et technologies de la traduction : quelles convergences pour l'avenir ?, Mar 2011, Paris, France. 14p. hal-02496086

HAL Id: hal-02496086

<https://hal.science/hal-02496086>

Submitted on 2 Mar 2020

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

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

TRALOGY

Translation and the New Digital Commons

Philippe Lacour

Dr., Ecole Normale Supérieure Paris (CIEPFC) / Centre Marc Bloch Berlin

Any Freitas

Dr., European University Institute

Aurélien Béné

Pr., University of Technology of Troyes (Tech-Cico Lab)

Franck Eyraud

Independent Developer

Diana Zambon

Independent Developer

TRALOGY I - Session 7

Date d'intervention : 04/03/2011

In this presentation, we have three key claims. First, we wish to make a case for an *alternative* conception of linguistic diversity, which sees language pluralism not as an “obstacle” or “barrier” that needs to be (dis)solved, but rather as a value that needs to be cherished and promoted, at different levels of politics and society. Second, we also hold that the “Digital Humanities” create new possibilities and thus open a “new age” for literary translation, which will also deeply impact research and education [Lacour, P. *et alii* (2010a)]. Third, as we should argue, the application of interpretive and corpus-driven linguistics to Computer Assisted Translation should foster collaboration on the realm of precise translation of cultural texts [Béné, A. and Lacour, P. (2011)] and therefore help reinforcing the sustainability of culture(s) and identity(ies), “on” and “off” line. Behind these claims, lies a conception of “language as a common good” which can (or should) be freely disposed by all its users. This paper aims consequently at proposing a more appropriate definition of the copyright for digital literary translation, especially for multilingual corpora.

lien video : https://webcast.in2p3.fr/video/translation_and_the_new_digital_commons



Introduction

Translation should probably become one of the most central institutions of the 21st century. As the pace of globalization processes seemed to increase during the new century's first decade, information and communication technologies (ICT) have definitely altered the patterns of human behaviour and social interaction. The Internet, the most powerful and ubiquitous of all, has made communication not only cheaper, simpler and impressively faster, but also virtually ubiquitous. After an initial period during which English reigned as the Internet "lingua franca", we have been witnessing an ever-growing process of language diversification of the web [Pimienta, D. (2001)]. As a clear illustration of that, the proportion of English-speakers using the Internet declined from more than 80% in the year the Web was born to 35% in 2005 [Unesco (2005), and specialists keep pointing to the great plurality of languages actually circulating on the net.

If an important part of this plurilingual content is in fact the translated version of web interfaces to local languages ("localisation"), users have been increasingly claiming the "right" not only to communicate, but also have access to other content in *their own mother-tongues*. Indeed, the promotion of cultural diversity, notably by the support of language pluralism, has become one important (political) claim – and *norm* – in recent decades. Translation appears hence a key (and particularly needed) concept in the plurilingual cyberspace, enabling people and cultures to communicate. The relevance of translation can be spotted as a number of international documents, institutions and declarations are established to promote it; and a new field of knowledge, 'Translation Studies', has been developing exponentially [Oustinoff, M. (2010)].

As other realms of human activity and knowledge, translation has gained a powerful ally with the development of ICT tools [Lacour P. *et alii* (2010a)]. In fact, one only begins to imagine the possible uses of ICT not only to facilitate and promote translation, but also as supports for the preservation and the concrete development of cultural pluralism. E-translation devices seem all the more important in the context of growing digitalization of cultural material (digitization of classical texts published in the public domain, due to the work of many libraries and other actors). Science and academia should particularly benefit from both the digitalization of knowledge and the spread of translation tools on virtual working spaces.

But how ready is the technological (and academic) world to make room for an *alternative* conception of linguistic diversity, which sees language pluralism not as an "obstacle" or "barrier" that needs to be (dis)solved, but rather as a value that needs to be valued and promoted, at different levels of politics and society? To what extent can "Digital Humanities" create new possibilities and thus open a "new age" for literary translation, and how deeply would it impact research and education [Lacour P. *et alii* (2010b)]? Finally, how could Computer Assisted Translation foster collaboration and help reinforcing the sustainability of culture(s) and identity(ies), "on" and "off" line? Behind each of these question lies a conception of "language as a common good" which can (or should) be freely at the disposal of all its users.

In this paper, we shall first attempt to show how translation in the Human and Social Sciences resists the alternative between mass automatization for pragmatic translation and human craftsmanship for difficult texts. How could one imagine *mass-customization* technologies in the digital era? How could we imply more traditional human skills in the man-machine process concerning translations? What kind of philosophy of translation, or linguistics, does it imply?

We shall then briefly examine the TraduXio project, as a concrete example of a technological attempt to illustrate the idea of recycling translations. We shall notably delve on TraduXio's *philosophy* of language, which has direct implications on the design of its interface.

Furthermore, by drawing from this concrete experimentation, we will try to imagine what a digital right to translate could be. By understanding both how this right to translate was historically

built upon the copyright and how contemporary needs, in a globalized world, potentially require translations, we shall try to make a plea for more legal openness.

As an illustration, we will then examine the possibility of setting-up a “translation license”, inspired by the Creative Commons Plus license, and being currently experimented by the ‘Fonds Ricoeur’, a French (non-profit) Foundation dedicated to the promotion of the works of Paul Ricoeur, including through online access to some of his disseminated texts.

Finally, we will discuss whether there is a specific legal problem for corpora, considered as the result of an original collection of texts, and therefore as a special legal entity (distinct from singular texts).

1. Resisting an Alternative: Translation for Human and Social Sciences

One of the main difficulties one has to face while entering the contemporary reflection in both *Translation Studies* and ICT is the *rigid* and *exclusive* alternative between two conceptions of translation. It seems as if mass automatization were required for pragmatic translation, sometimes with a slight human touch (most contemporary CAT device now include human correction to their automatic translators). Conversely, difficult texts, such as poetry, religious, philosophical texts, as well as all the productions in Humanities, could only be granted proper attention through careful, minute and patient human craftsmanship. The very idea of using technologies for literary translation, for instance, would therefore sound ludicrous.

When framing the problem in such a way, the bias lies therefore therein that little (if any) room is given to a reflection on more inspired uses of the technologies. But, in fact, the machines are only “intelligent” to the extent humans ask them to answer intelligent questions. Now, the idea of a (thorough) automatic translation is precisely grounded on very dubious philosophical assumptions, and therefore misguided from the start.

Consequently, (sheer) automatic translation cannot succeed when it comes to precise semantics, whatever the power of the algorithms used (be it based on statistical parsing or grammatical analysis). All too often, however, the conceptual failure of these endeavours is somehow hidden by the appeal to a touch of “human participation”. Although such “participative” approach might look like a progress, it is in fact only meant to improve the results of automatic production. Therefore, instead of tackling the issue of the semantic issues, which is at the core of the complexity of human sciences, it tends to delay the logical solution to the difficulty, and consider “literary difficulties” as (only) residual. Indeed, if translation problems and semantic nuances are notorious in poetry and philosophy, it is important to stress that human sciences face very similar issues, because of the semantic nuances. Their key concepts (care, nationality, etc.) do concentrate indeed a lot of intertextuality, thus making the translation in a different cultural context very complex. How could one address this specificity without rejecting technologies? A philosophical detour might prove necessary.

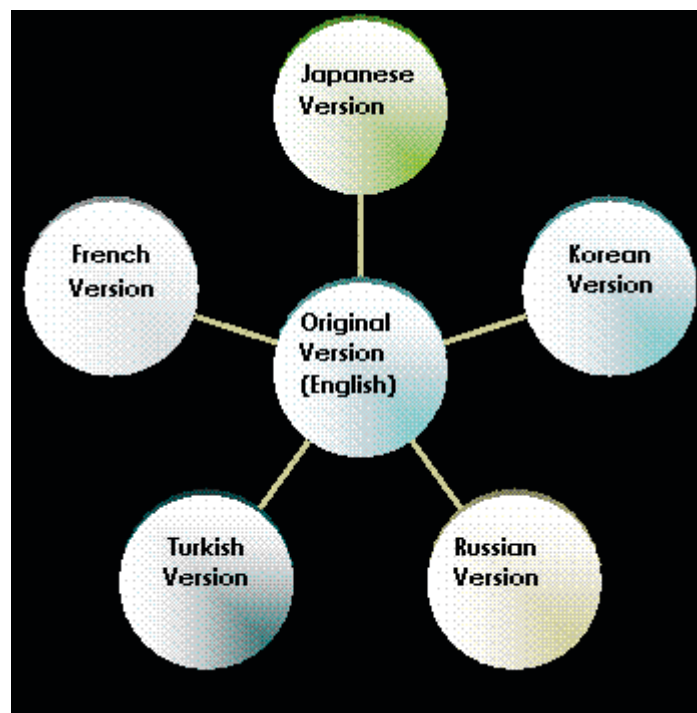
There are two ways of analysing the language, which are both important and not exclusive one from another. The first one insists on rules, which can be the grammar rules (as in the early Systran) or the statistical norms (the most frequent uses, popularized by Google Translate). The second one insists on singularities, along with the German romantic tradition, and the creativity of language [Lacour, P. *et alii* (2010a)]. Contemporary technologies have explored extensively the first option, with great success. But what if one *starts* with taking into account semantic nuances ? In this case, automatization might be used, but not for translation itself; rather it is meant, in a more modest way, to ask for suggestions, browse previous translations for relevant advice, compare one’s intuitions with translations already existing in related languages (like two roman languages, for instance), etc.

In other words, if one is to compare TraduXio with the Google Translator's Toolkit, for instance, one could say that the latter goes from machine to human to machine again (human participation is only meant to make machine translation more efficient), whereas the former starts from human translation and goes through automatization (of suggestions) in order to enrich other human translations. TraduXio therefore illustrates the switch of paradigm in the conception of Artificial Intelligence, from machines that think to machines that make people think [Bachimont, B. (1996)].

2. The TraduXio Case: a Recycling Endeavour

Employing one of the most original solutions available today in the area of web-based translation, *TraduXio* presents a number of advantages when compared to existing devices. It is developed by the Zanchin NGO, in collaboration with the University of Technology of Troyes, and with the support of the UNESCO, the International Organization of "Francophonie" and the Délégation à la Langue Française et aux Langues de France (among other partners). TraduXio is a free, open source, web based, collaborative, and computer assisted translation tool, developed with innovative technology. Inspired by the strong collaborative spirit of the Web 2.0, and available to different audiences, the software is aimed at becoming a mechanism of general interest. Though it was first tested in the field of Research, *TraduXio* also has a tremendous potential for online Education, artistic dissemination and social integration. The originality of the software resides in certain of its functionalities.

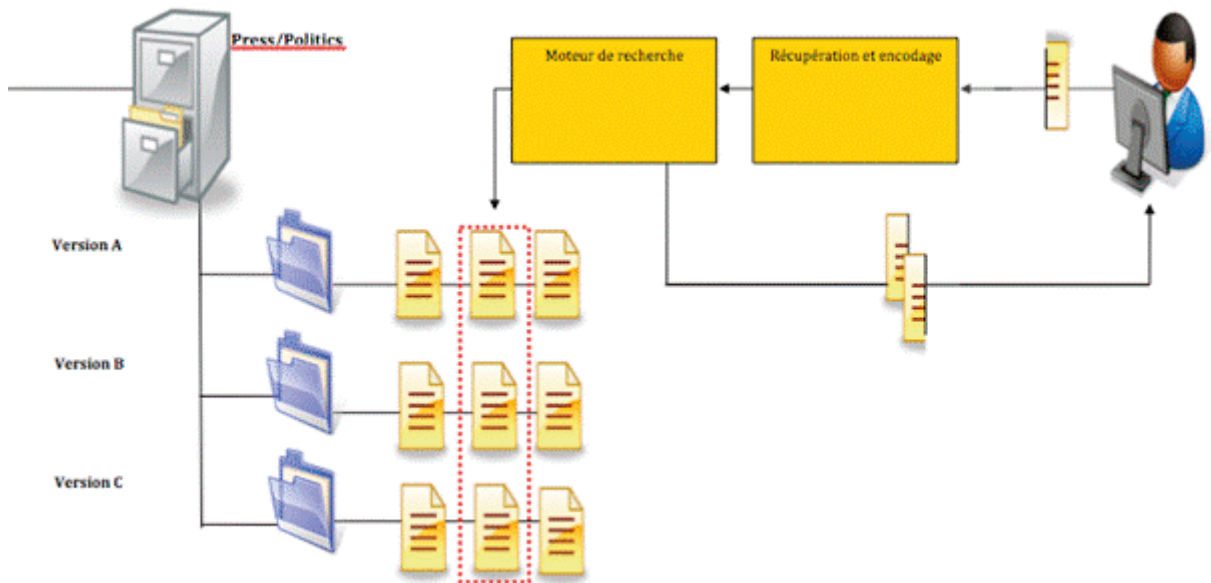
(i) Traditional Computer Assisted Translation tools, especially Translation Memories, are limited to two languages (the source / the target), thus enforcing a "star" system, in which a privileged language is set at the centre – one only, and always the same. In this device, like the Google Translator's toolkit¹, for instance, one can only go from language B to Language C, through language A (in the centre: English).



Star system (e.g. Google Translator's toolkit)

(1) See : <http://blogoscoped.com/archive/2008-08-04-n48.html>

On the contrary, TraduXio enables multilingual translation, through the comparison of different versions of the same text. In this case, a translated text is not considered as an independent segment, but rather as a version of the initial text in another language. TraduXio's inner structure is therefore a serial system, and not a star one. Also, the original version can be in any language, not only one (as in the "star" system).



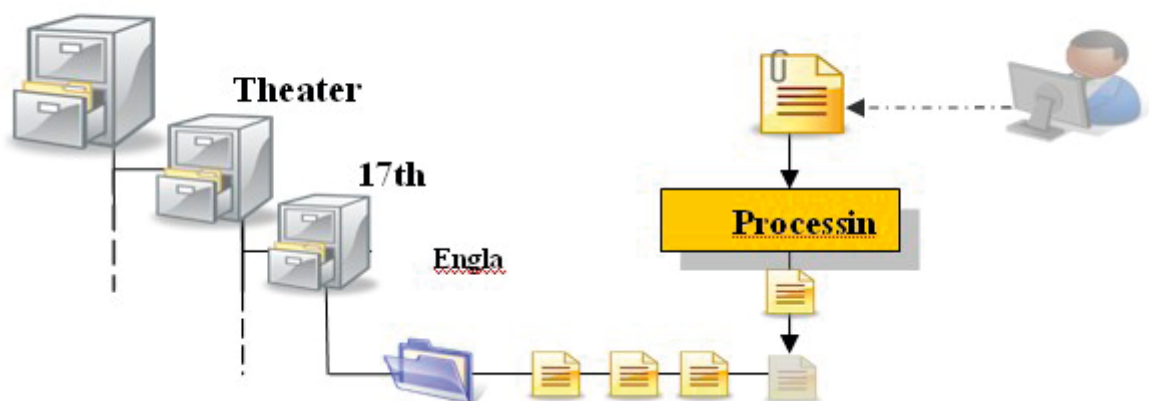
Serie system: TraduXio

The serie system allows *multilingual* translation, which are visible at the same time

The screenshot shows the TraduXio interface with a sidebar on the left containing options: 'Options', 'Modifier cette traduction', 'Gestion des privilèges du texte', 'Éliminer cette traduction', 'Historique des modifications', and 'Cacher les blocs'. The main content area displays a comparison of two versions of a text. The left column is labeled 'Espanol' and the right column is labeled 'Italian'. Both columns show the text of 'Miguel de Cervantes, Capítulo I'. The text is displayed in a table-like structure with two columns, allowing for side-by-side comparison of the original and translated versions.

(ii) *TraduXio* also offers a better management of the translation context, because it provides for a contextualized classification of the source (i.e. classification of the text according to the history, genre, author, etc.), in a very liberal and flexible way (according to the users' or community's choice). Thanks to this relevant classification device, information can be more easily

assessed and treated, thereby helping users finding the appropriate translation for particular words, expressions, and so forth. Indeed, the documentary base can be browsed for relevant suggestions, and not for the “most frequent use” (as in statistical approaches to translation), since it pays considerable attention to the nuances of the language:



Relevant suggestions through the use of the concorder

release 2009		release 1851 1851 ca.1855		to sleep perchance to dream Search
Hamlet, III, 1 (monologue) Anglais Shakespeare undefined	Or to take arms against a sea of troubles, And by opposing end them? To die: to sleep ; No more; and by a sleep to say we end The heart-ache and the thousand natural shocks That flesh is heir to, 'tis a consummation Devoutly to be wish'd. To die, to sleep ; To sleep : perchance to dream : ay, there's the rub;	ou bien à s'armer contre une mer de douleurs et à l'arrêter par une révolte? Mourir... dormir, rien de plus... et dire que par ce sommeil nous mettons fin aux maux du cœur et aux mille tortures naturelles qui sont le legs de la chair: c'est là un dénouement qu'on doit souhaiter avec fervor. Mourir... dormir, dormir! peut-être rêver! Oui, là est l'embarras.	Hamlet, acte 3, scène 1 (monologue) Français François-Victor Hugo undefined	
Hamlet, III, 1 (monologue) Anglais Shakespeare undefined	Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles, And by opposing end them? To die: to sleep ; No more; and by a sleep to say we end The heart-ache and the thousand natural shocks That flesh is heir to, 'tis a consummation Devoutly to be wish'd. To die, to sleep ; To sleep : perchance to dream : ay, there's the rub;	¿Cuál es más digna acción del ánimo, sufrir los tiros penetrantes de la fortuna injusta, u oponer los brazos a este torrente de calamidades, y darlas fin con atrevida resistencia? Morir es dormir. ¿No más? ¿Y por un sueño, diremos, las aflicciones se acabaron y los dolores sin número, patrimonio de nuestra débil naturaleza?... Este es un término que deberíamos solicitar con ansia. Morir es dormir... y tal vez soñar. Sí, y ved aquí el grande obstáculo,	Hamlet, III, 1 Español Inacio Celenio undefined	
Hamlet, III, 1 (monologue) Anglais Shakespeare undefined	Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles, And by opposing end them? To die: to sleep ; No more; and by a sleep to say we end The heart-ache and the thousand natural shocks That flesh is heir to, 'tis a consummation Devoutly to be wish'd. To die, to sleep ; To sleep : perchance to dream : ay, there's the rub;	Ob's edler im Gemüt, die Peil' und Schleudem Des wütenden Geschicks erdulden, oder Sich waffend gegen eine See von Plagen, Durch Widerstand sie enden. Sterben – schlafen – Nichts weiter – und zu wissen, daß ein Schlaf Das Herzweh und die tausend Stöße endet, Die unsers Fleisches Erbteil – 's ist ein Ziel, Aufs innigste zu wünschen. Sterben – schlafen – Schlafen! Vielleicht auch träumen! – Ja, da liegt's:	Hamlet, III, 1 Allemand Schögel / Tieck undefined	

(iii) As a collaborative translation software, *TraduXio* is more than a common workbench for digital translators. It is also a network and a platform where translators can meet and create joint projects, exchange ideas, create corpora and glossaries. In this sense, it creates one of these collaborative devices which have become increasingly famous through the (so called) “Web 2.0” movement, such as versioning, management of privileges, social tagging (not to mention wikis, forums, social networking, etc.).

Privileges Management

Traduxio Project - Welcome, philippe.lacour (Logout)

Back to the text Switch to CUSTOM PRIVILEGES

Standard Privileges

Text visibility

- ☒ Private
- ☐ Public

Save

Note: A blue double-headed arrow connects 'Standard Privileges' and 'Switch to CUSTOM PRIVILEGES'. Another blue arrow points from 'Text visibility' to the 'Private' radio button.

Custom Privileges Management

Traduxio Project - Welcome, philippe.lacour (Logout) Works Concorde

Back to the text Switch to STANDARD PRIVILEGES

Custom Privileges

Remove an existing permission

- ☐ User admin is allowed to do everything for this text
- ☐ User philippe.lacour is allowed to edit this text
- ☐ User diana.zambon is allowed to read this text

Remove Privilege

Add a new permission

User:

Privilege:

Note: A blue double-headed arrow connects 'Custom Privileges' and 'Switch to STANDARD PRIVILEGES'.

3. Enforcing a Digital « Right to Translate »?

In the last decade, legal tools (open licenses) have been developed to allow for the (re) appropriation, sharing and mutual (cultural) recognition of cultural production (and identities' reproduction) worldwide. This sophisticated system of licensing represents not only an alternative to more "traditional" conceptions of (intellectual) property (the "Commons"), but also provides legal support to those who view translation as a fundamental *right* [Basalamah, S. (2009)].

Indeed, in his compelling book, Salah Basalamah analyses how the translation right has been build upon a very traditional and restrictive conception of the copyright (or "droit d'auteur"), thus depriving potential translators of original versions of any co-authorship status. As revealed from a historical perspective, which includes a thorough analysis of the Bern Convention on

Translation Rights and of its most recent revisions, the exclusive focus on the protection of the author's rights has resulted in impeding dissemination and appropriation of culture on a very vast scale.

On the contrary, the new era of globalization calls for an alternative, a general revision of the author's right, which can take the form of legal tools meant to escape the strictness of the copyright regulation. Such "alternative" licensing system is based on the recognition that intellectual property rights quite often work as barriers to the free circulation of culture and thus need to be adapted to the needs and aims of both authors and the community. The idea is not to deny the authors' legitimacy over product of their creation, but to facilitate sharing, (re) appropriation, "mix", and ultimately foster the circulation and (mutual) recognition of cultures worldwide.

With TraduXio, each translator will be given the possibility to tag his/her creation with a legal license of his choice. The range of rights' attribution goes from public domain to full copyright through "open" licenses, for both texts and corpora. Users will be however encouraged to choose at least an attribution license, in order to avoid the positivist conception of translation as an invisible operation [Venuti, L. (1995)], and the symmetrical idea of the collection of data as neutral (rather, they are documents for a certain use – [Zacklad, M. *et alii* (2004)]). Attribution of authority over translations is indeed important since the re-utilization of "memory matching" depends on the identification of the author for a given semantic creation. Frequent users could thus benefit from a form of public and non-financial recognition (a system of "points"), which eventually might turn into a sort of professional reputation. The same liberal approach would apply to databases (the concordancer constituted through a specific use of *TraduXio*).

Now, if the "Commons" represent an opportunity to cultural diversity and language pluralism, the challenges that copyright law poses to E-translation have been thus far addressed in a rather superficial and irregular manner. The tendency is quite the same in different disciplines, even in the trend of Translation Studies that focuses on the rather cultural and legal aspects of translation. Despite the increasing development of literature, the recent and quite dramatic transformations engendered by the application of new Information and Communication Technologies (ICT) to translation – especially on dataset issues– remain virtually unexplored. The next section of this paper aims precisely at proposing a more appropriate definition of copyright for digital literary translation.

4. Case Study: Designing a Licensing System for Philosophical Translations

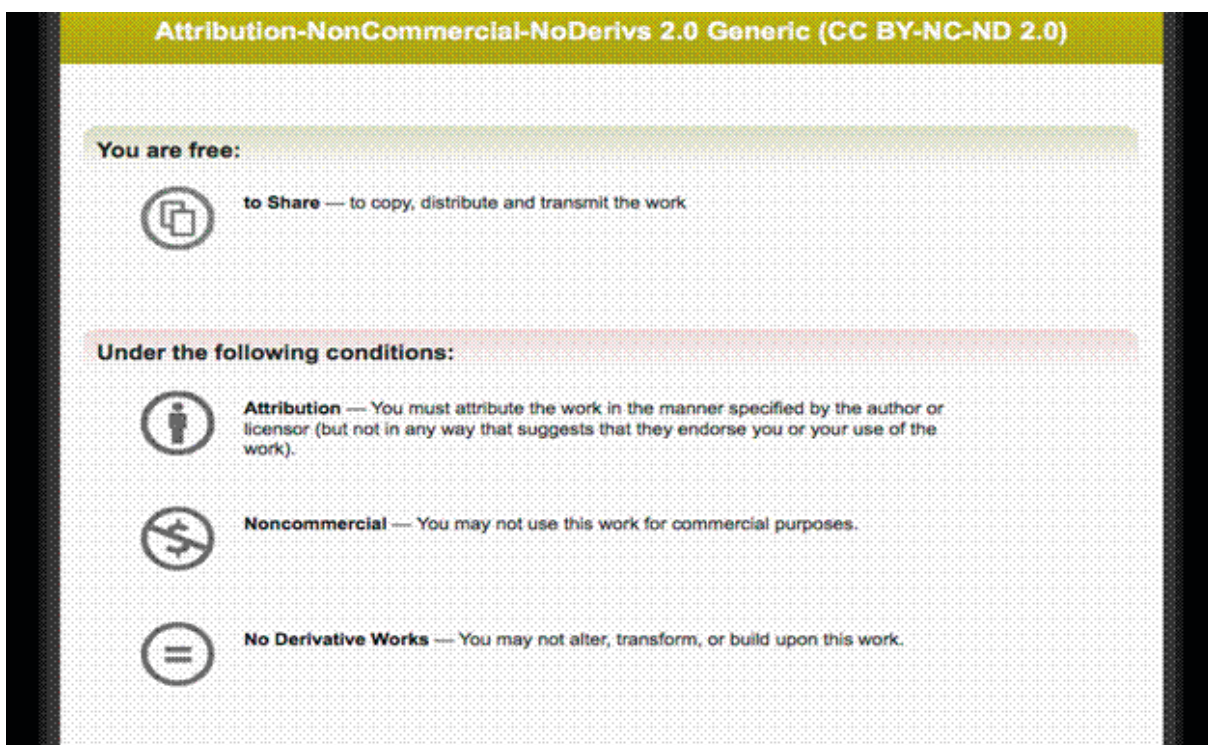
In the previous sections, we have discussed the more abstract/theoretical dimensions of an emerging, *sui generis* "right to translate". Considering "language as a common good" the right to translate takes seriously the claim – embraced by several international institutions such as UNESCO, the Council of Europe or the European Commission – that culture should freely circulate among the people, that language plurality need to be encouraged and promoted and that languages should be freely disposed by their users.

In this section, we will discuss some of the concrete challenges that copyright law poses to translation (in particular, to 'e-translation'), and how these challenges can be effectively overcome. We will proceed in this discussion by analysing the "Fonds Ricoeur" situation, a case that illustrates quite clearly the specific obstacles that the current copyright system represents to authors (or, in this case, their legal representatives), translators and the general public. We will then explore the possible alternatives to this limitation, notably by proposing the adoption of an "adapted" Creative Commons license in order to allow translations to be (always) authorized. We propose in fact that this should become the 'default' license for translations, in line with a 'right to translate' premise.

The Fonds Ricoeur (a Paris based non profit Foundation) wishes to disseminate online, through its own web site, certain texts of the famous French philosopher, which have become difficult to retrieve (they were originally published in foreign journals, or in journals that have disappeared since then)². The Editorial Board of the Foundation holds all the copyrights for these texts, as opposed to other (writtenly) published texts (belonging to regular book publishers). Along with a growing international reception of the work of Paul Ricoeur, the demand for the translation of his minor texts will probably increase.

The Editorial Authority of this Foundation wishes to avoid two main problems, which are all the better identified since they actually occurred quite recently in the French intellectual history: the lack of circulation, and the monopoly over translation of the work (or of certain piece of the works), on the other hand. The baseline chosen to address the issue is the following: promote "openness", by allowing translation to be *always (legally) possible, for all the texts, disregarding case-based personal authorisations*. However, commercial rights are reserved. If commercial uses are considered for a certain translation, permission must be asked for to the Editorial Board of the Foundation (and will probably remain non exclusive). This device is meant first and foremost to promote scholarly translation, mainly through the interest of young PhD scholars, who are often keen on translating specific texts, and whose career might benefit from such an endeavour.

The difficulty consists in inventing a legal license enabling a certain "right to translate" these texts, from a both liberal though (somehow) restricted perspective: translating the whole text, nothing but the text and only from the original version. Such a license does not exist yet, mainly because translation is considered as a form of derivative work. It is therefore allowed or not, according to the license, such as the CC-BY-NC-ND case (see the simplified contract: <http://creativecommons.org/licenses/by-nc-nd/2.0/>)



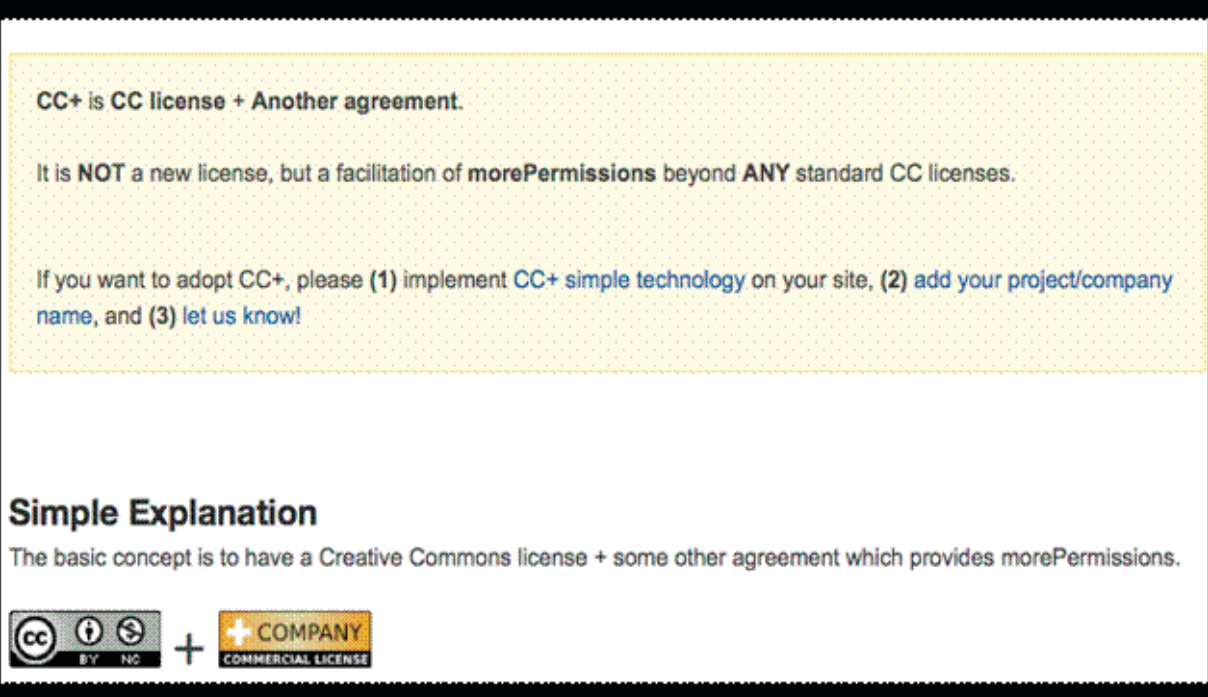
Details of the contract are available at :
<http://creativecommons.org/licenses/by-nc-nd/2.0/legalcode>)

(2) <http://www.fondsriceur.fr/intro.php>

Along with a reasonable definition of a translation as “an equivalent without an identity” [Ricoeur, P. (2004)], a possible solution would consist in using an already existing license while modifying it slightly, so as to authorize translations in foreign languages as some kind of specific derivative works. Restrictive conditions would however be given, and the translator would translate:

- the whole text
- nothing but the text
- only from the original version of the text

Certain modifications of existing « Creative Commons » licenses already exist (the so called “CC+” licenses):





CC+ is CC license + Another agreement.

It is **NOT** a new license, but a facilitation of **morePermissions** beyond **ANY** standard CC licenses.

If you want to adopt CC+, please (1) implement **CC+ simple technology** on your site, (2) add your project/company name, and (3) let us know!

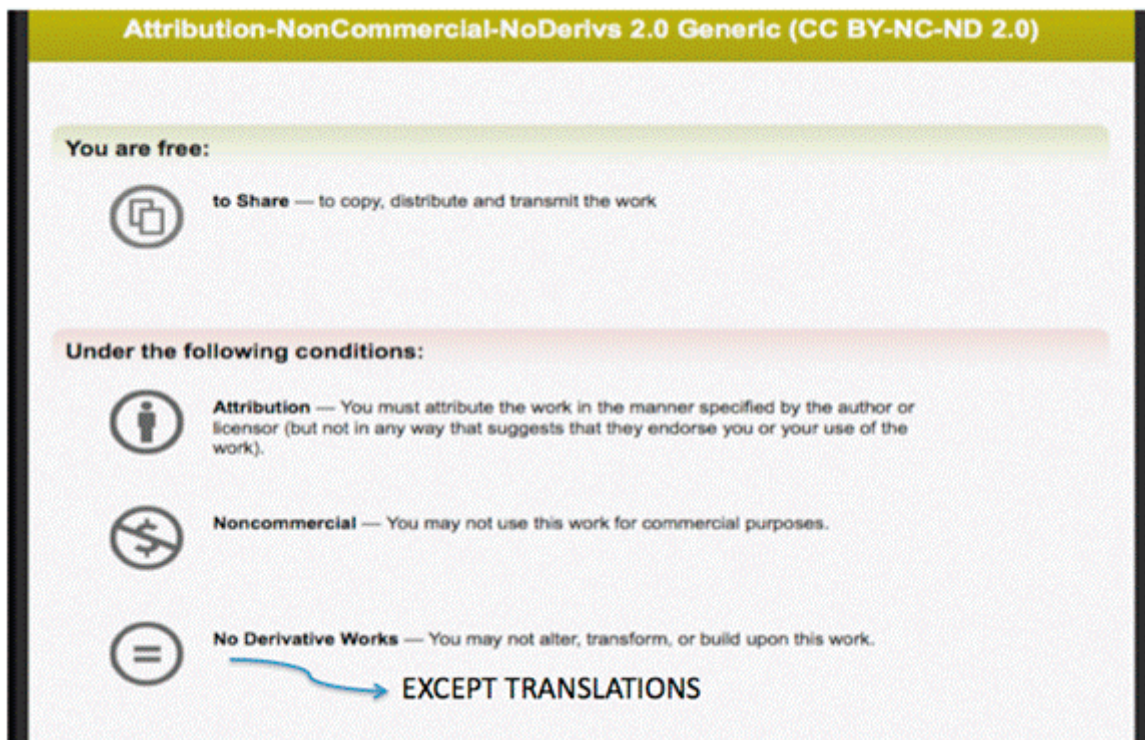
Simple Explanation

The basic concept is to have a Creative Commons license + some other agreement which provides morePermissions.

 + 

Although CC+ licenses have most of the time been used, until now, as an exception form to the “non commercial” clause, nothing prevents theoretically to imagine another mechanism, as suggests the bottom line: « The basic concept is to have a Creative Commons License + some other agreement which provides more Permissions ». By way of example, if someone wanted to apply BY-ND (attribution-no derivative) to a work, but wanted to allow for translations only, they could use CC Plus to allow that specialized additional permission, to anyone or only particular persons, with or without fee, and with or without additional conditions tied to the right to do translations.

Concerning the Fonds Ricoeur, the legal idea would therefore be schematically illustrated as follows:



in other words, and in using a CC+ form (to which one should add a ND clause):



CC+ is CC license + Another agreement.

It is **NOT** a new license, but a facilitation of **morePermissions** beyond **ANY** standard CC licenses.

If you want to adopt CC+, please (1) implement **CC+ simple technology** on your site, (2) **add your project/company name**, and (3) **let us know!**

Simple Explanation

The basic concept is to have a Creative Commons license + some other agreement which provides **morePermissions**.

 +  → **whole text, only the text, from the original**

5. Is there a Specific Corpora Problem?

There is a scientific bias in the positivist view pervading the very notion of “data”. In fact, there can be no such thing as “pure data”, since no data can exist without an underlying theory organizing data. Cultural Sciences are written in *natural* and not formal languages. Even if their discourses might include ‘formal moments’, these areas of human knowledge are intrinsically linked to the communicative properties and possibilities of natural languages. Such a close connection imposes particular constraints to the ways of reasoning. Indeed, these are sciences of *inquiry*, based on an *intermediary* rationality that is located in a logical space between the mere opinion and the robust formal thought. Cultural sciences are moreover essentially *interpretive* and work in a *casuistic* and *reflexive* manner.

This has many implications on the constitution of their sets of information. In particular, one should resist the idea of Web semantics as a superposition of standardized layers (“layer-cake”). Rather, one should promote a more socio-semantic or pragmatic approach, by insisting on the dynamics of uses and appropriations, rather than on fixed ‘datas’. This implies combining the notions of *document* (having a certain testimonial value, submitted to description, revision and signature, according to a particular investigation), *interpretation* (heuristic modelization) and *intersubjectivity* (rational comparison of view points, organization of the conflict of interpretations).

The particular case of multilingual corpora can now be addressed. We shall stress the fact that a corpus is not a bag of words but a set of texts gathered according to a certain question, and therefore a *documentary base*, as much as an *original* creation of the mind (not a neutral collection). Such orientation does not, however, hinder further distribution and reuse, neither from an intellectual nor from a legal point of view.

As a collaborative translation environment, *TraduXio* is a platform where translators can create corpora. However, the TraduXio project relies on *interpretive* semantics (hermeneutics) and therefore relies more on corpus-driven than on corpus-based linguistics. *Corpus-based* approaches focus on generalization and extraction of standardized information, and digitized corpora only reinforce such vision through the idea of data-mining and automatic parsing [Tognini-Bonelli, E. (2001)]. However, these approaches tend to overlook the text itself, which is given at best a subsidiary role [Rastier (2005)]. From this perspective, corpora only serve to illustrate a priori (linguistic) theories, of which they are considered “representative”.

Based on more reflexive perspectives [Mayaffre, D. (2002)], *corpus-driven* orientations consider, on the other hand, that the very construction of texts’ set is problem-centered, that is, connected to a particular “question”. In this conception, corpora are to some extent also singular – a “singularity” that is positively understood through the notion of *clinical* knowledge [Thouard, D. (2011)]. Stressing the problem-driven and hence auctorial character of corpora constitution processes does not contradict the possibility of further using or “recycling” these corpora, especially in the digital age. Corpus-driven analysis is therefore very open, and cannot be reduced to the identification of grammatical patterns [Hunston and Francis (2000)]. In fact, as we wish to stress here, corpus-driven linguistics is fundamentally hermeneutic and case-based. Such perception leads us to take into account the social dimension of text classifications, especially through the introduction of categories (such as “genres” and types of discourse), as observed Rastier [Rastier, F. (2008)], to follow the recommendation of the linguistics of norms. This switch in semantics [Rastier, F. (2004)] can be given a digital materialisation, for instance through an appropriate protocol: designed by the Tech-Cico Department of the University of Technology of Troyes, the *Hypertopic* protocol aims precisely at visualizing different view points on a subject [Bénel, A. *et alii* (2010), Bénel, A. and Lejeune, C. (2009)].

From a legal perspective, the do-it-with-others corpus orientation of the TraduXio project implies to claim authorship over the set of texts one has designed in an original way. While remaining preferably “open” to reuse and recycling, multilingual corpora can therefore not pretend to be

"neutral". A vast movement, originating in the natural ("hard") sciences claims that databases should remain in the public domain and that author rights should be waived accordingly, in order to promote re-use on a very large scale. By doing so, the Creative Commons so called "zero" license³ conveys however a very positivist and questionable view of data collection. On the contrary, TraduXio, which insists on the originality of the act of interpretation underlying the gathering of documents for a certain use, would recommend to tag datasets with a least one attribution license.

Bibliography

Bachimont, Bruno (1996), *Herméneutique matérielle et artéfacture : des machines qui pensent aux machines qui donnent à penser*, thèse de doctorat de l'Ecole polytechnique en épistémologie, 24 mai ; [http://www. utc.fr/~bachimon/Livresettheses.html](http://www.utc.fr/~bachimon/Livresettheses.html)

Basalamah, Salah (2009), *Le droit de traduire. Une politique culturelle pour la mondialisation*. Ottawa : Presses de l'Université d'Ottawa.

Bénel, Aurélien, Zhou Chao and Cahier Jean-Pierre (2010), 'Beyond Web 2.0...And Beyond the Semantic Web', in Randall D. and Salembier P. (eds.) *From CSCW to Web 2.0: European Developments in Collaborative Design*. London: Springer Verlag. Available online at http://benel.tech-cico.fr/publi/benel_SPRINGER_2010.pdf (accessed 27 July 2010)

Bénel, Aurélien and Lejeune, Christophe (2009), 'Humanities 2.0: documents, interpretation and intersubjectivity in the digital age', *International Journal of Web Based Communities*, 5(4): 562-576.

Bénel, Aurélien and Lacour, Philippe (2011) 'Towards a Collaborative Platform for Cultural Texts Translators'. In Maret Pierre (ed.) *Virtual Community Building and the Information Society: Current and Future Directions*. Hershey (Pennsylvania): IGI Global (forthcoming).

Hunston and Francis (2000), *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*, Amsterdam and Philadelphia, Benjamins, 2000.

Lacour Philippe, Bénel Aurélien, Eyraud Franck, Freitas Any and Zambon Diana (2010a), « TIC, Collaboration et Traduction : vers de nouveaux laboratoires de translocalisation culturelle », *Meta*, 55(4), 2010, Presses de l'Université de Montréal.

Lacour Philippe, Bénel Aurélien, Eyraud Franck, Freitas Any and Zambon Diana (2010b) 'Managing Customization in Language Teaching through Collaborative Translation: *Traduxio*, an open source platform for Precise Translation', in Maerlein, Michael (ed.), *Proceedings of the First International Conference: "Mass Customization" for Language Teaching and Learning*. Dublin, 25-27 August 2010 (forthcoming).

Mayaffre, Damon (2002), « Les corpus *réflexifs*: entre architextualité et hypertextualité », *Corpus*, 1, nov. Available online at [http://corpus.revues.org/document11 .html](http://corpus.revues.org/document11.html) (accessed 27 July 2010)

Pimienta, Daniel *et al.* (2001) "The fifth study of languages on the Internet", available at <http://funredes.org/LC/english/L5/L5tendencies.html>

(3) <http://creativecommons.org/publicdomain/zero/1.0/>

Oustinoff, Michaël (2010), « Les *Translation Studies* et le tournant traductologique », *Revue Hermès. Cognition. Communication. Politique*, n° 56 : « Traduction et mondialisation, vol. 2 », Paris, CNRS, pp. 21-28.

Rastier, François (2004) « Ontologie(s) ». *Revue des sciences et technologies de l'information*. Série : *Revue d'Intelligence artificielle*. 18(1), pp. 15-40. Available online at http://www.revue-texto.net/Inedits/Rastier/Rastier_Ontologies.html (accessed 27 July 2010).

(2005) « Rôle et place des corpus en linguistique : réflexions introductives ». *Texte !*, décembre 2005, vol. X, n° 4. Available online at http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html (accessed 27 July 2010)

(2008) "Conditions d'une linguistique des normes", *Texte !*, available online at <http://www.revue-texto.net/index.php?id=1612> (accessed 27 July 2010)

Ricoeur, Paul (2004), *Sur la traduction*, Paris, Bayard.

Tognini-Bonelli, Elena (2001), *Corpus Linguistics at Work*, Amsterdam: John Benjamin's Publishing.

Thouard, Denis (éd.) (2011), *Herméneutique contemporaine. Comprendre, interpréter, connaître*, Paris, Vrin.

UNESCO (2005), *Measuring Linguistic Diversity on the Internet*, UNESCO Press: Paris.

Venuti, Lawrence (1995), *The Translator's Invisibility: A History of Translation*. London & New York : Routledge.

Zacklad, Manuel (2004), « Processus de documentarisation dans les Documents pour l'Action (DopA) ». In Savard R. (ed.) *Actes du colloque «Le numérique : impact sur le cycle de vie du document»*, Montréal, 13-15 octobre 2004, pp. 139-175. Lyon : École nationale supérieure des sciences de l'information et des bibliothèques, 2004. Available online at <http://www.enssib.fr/bibliotheque-numerique/notice-1223> (accessed 27 July 2010)