

Towards a participative platform for cultural texts translators

Aurélien Béné

ICD/Tech-CICO Lab, UTT (Troyes, France)

Philippe Lacour

ENS (Paris); Marc Bloch Center (Berlin)

ABSTRACT

Although machine translation and translation memories are frequently used in business, they are inadequate to translate a text from a culture to another one. When faced with philosophy, literature or ancient texts, professional translators have to cope with the fact that the most important things to ‘translate’ are often in the style, in details, or even unwritten.

We advocate for changing the user interfaces and use patterns of a few *computer-assisted translation* techniques so that they could fit the interpretative tradition of *cultural sciences*. In particular, we will focus on what could foster intertextuality and enable the confrontation of different points of view on the same opus (several translators in several languages).

Provided as a participative Web platform, our software is designed as a collaboration and debate place for scholars around the world working on the same opus, author, time or genre. At the end of the chapter, this design is confronted with the observation of a face-to-face working session.

INTRODUCTION

To scholars in social sciences or humanities, the so-called ‘Web 2.0’ (O’Reilly, 2005) means both new possibilities of cooperation and very surprising incarnations of rather familiar theoretical ideas. Indeed, it embodies – though in a very ‘pop culture’ way – some key aspects of *hermeneutics*, the theory and methods of interpretation (Lacour, 2010a; Béné & Lejeune, 2009).

In these disciplines, translation has a singular status. Even if their discourses might include ‘formal moments’, the areas of human knowledge related to *culture* (as opposed to *nature*) are intrinsically bounded to the properties and possibilities of natural languages (Passeron, 2006). This assumption entails quite important consequences. In particular, these disciplines’ key concepts are formulated in idioms characterized by their irreducible diversity. Translation difficulties arise when semantic fields do not exactly match in different languages. For instance ‘citizenship’ can mean either ‘citoyenneté’ or ‘nationalité’. An accurate translation tries to preserve most of these existing ambiguities without adding any¹. However, no translation can be perfect, and it is usually necessary to compare different translations to reveal all the original possible meanings.

Translation takes such a part in the interpretative tradition of *cultural sciences* embodied by ‘Web 2.0’ that one should be surprised that cultural texts platforms provide *either* interpretative translation features

¹ For this reason, François Rastier (2007) notes that great translations often witness to a deeper comprehension of the original text than commentaries.

(e.g. *Perseus digital library*²) or participation features (e.g. *Wikisource*³). In our process to integrate both kind of features, we will first highlight several translation theory principles⁴ and how they match (or not) typical *computer-assisted translation* tools. Then we will present the mockups of our platform together with their design rationale. To finish with, we will confront this design (based on translation theory) with a practical example as observed during a face-to-face working session.

BACKGROUND

Although computer-aided translation tools are widely used in business, they are carefully avoided in cultural sciences. In fact both machine translation and translation memories implicitly embed very questionable hypotheses concerning language.

From machine translation to human translation

Machine translation (see Fig.1) embeds a language theory in which translating could be reduced to applying a set of rules from a source language to a target language. First, this would require that a form could be replaced by another while preserving the meaning. On the contrary, both language theory and practice show that form-to-form translation (e.g. 'London' to 'Londres') is rare at any level (term to sentence). For example, when a translator cannot express a connotation in a translated form, she can move it on a neighbour form. Second, this would require the existence of language rules. For current theories, rules exist for genre but not for a whole language. In other words, they are not universal rules, but practical norms in use (Rastier, 2007). This last objection to rule-based machine translation could explain the renewed interest in computer-aided translation based on human translation. Among them, statistical machine translation is still not accurate enough to be used by professional, but, on the contrary, translation memories are widely used in translation agencies.

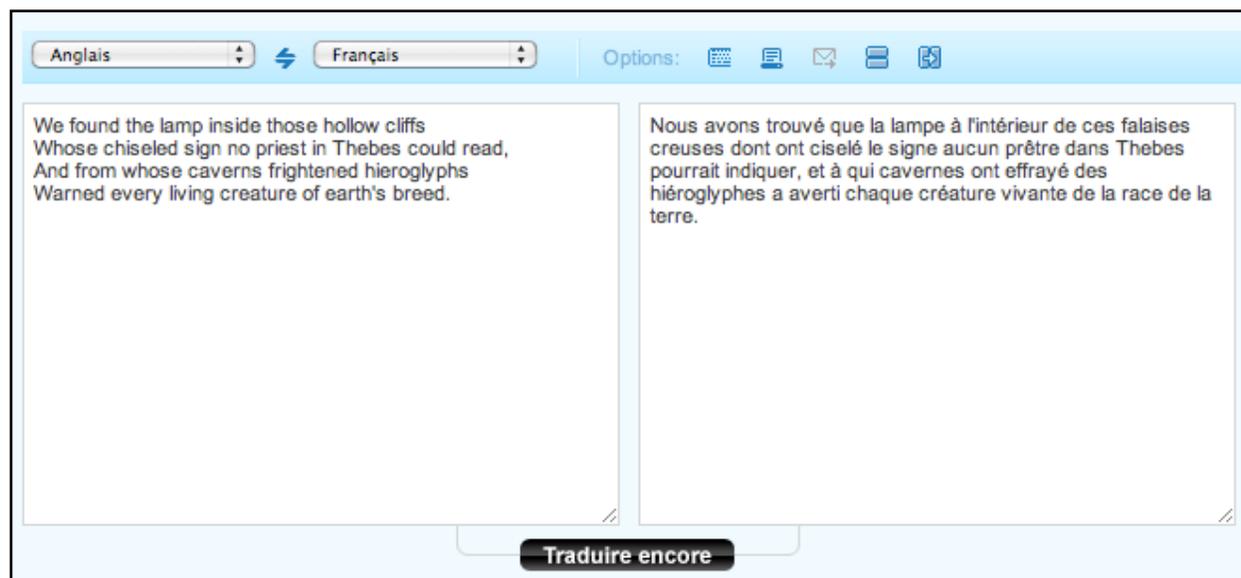


Figure 1. Machine translation of a poem (screenshot).

² <http://www.perseus.tufts.edu/>

³ <http://wikisource.org/>

⁴ Even in less erudite domains, we think that software designers should never consider an area of practice as if it were 'atheoretical'. Discovering users' conceptualizations is, in our opinion, one of the most important parts of the analysis.

From translation memory to concordance

A translation memory is a kind of database where translators store translated ‘segments’ so that they can be reused later. These segments are supposed to be large enough to be independent from their context, but small enough to be reusable as such or with a slight edit (see Fig.2).

The language theory embedded here may be less naive but is still questionable. First, getting context-independent segments is quite unrealistic. The context is indeed the whole text (Rastier, 1998). Second, large translation parts can only be reused on quotes or in very normative texts (laws, business letters, user manuals...). Third, translation memories sometimes mention a ‘domain’ on the segment but neither the author, nor the date and genre of the original text⁵. While their impact on translations can be low on mundane texts, they are known to be critical on cultural texts. Fourth, the name of the translator is usually neither store nor displayed, as if translations were ‘data’ rather than the result of a creative and interpretative activity.

These objections could explain why cultural texts translators do not use translation memories while they have used *concordances* and bilingual dictionaries for centuries. Contrary to translation memories, these traditional tools for classical studies mention precise references to the context, as well as the author, and sometimes his trend (e.g. ‘neoplatonicians’, ‘presocratics’...). Moreover, concordances are usually built from an homogeneous corpus (most of the time from texts by a single author). To finish with, it is worth noting that a comparative approach is embodied in some bilingual editions (especially for exegesis), in which concurrent translations are displayed on the same page.

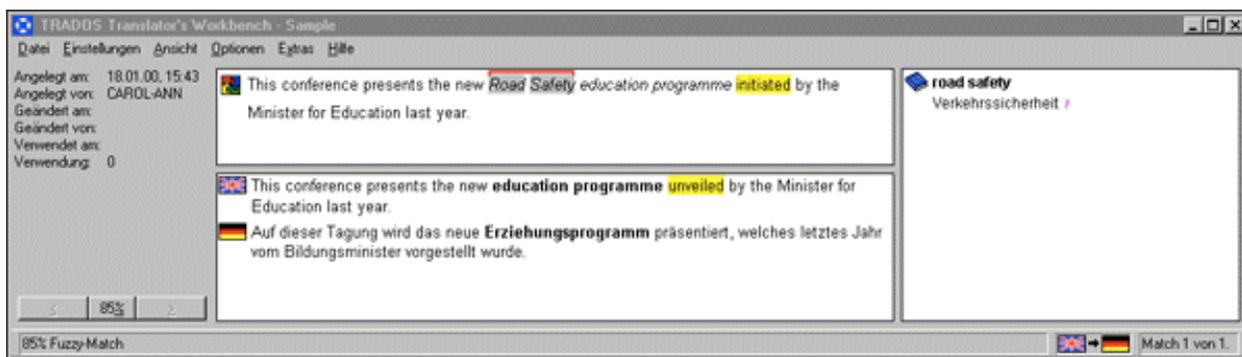


Figure 2. Translation memory used on similar sentences (screenshot).

TOWARDS INTERSUBJECTIVE TRANSLATION

These theoretical objections led us to design a participative platform for human translation featuring a concordance.

As the main page reminds the guests (see Fig.3), the platform we designed is not a translation service but a service for translators, and not a ‘database’ but a ‘texts-base’. The texts list at the middle of the page is primarily intended for translators. It aims at fostering the building of communities around a language and, if possible, around an author. The other list at the right side is intended for guests so that they can discover and read foreign texts recently translated into his/her language. One should note that, the only texts that

⁵ An in-house software has been reported to refute this as a general statement. However, it is significant that such clues about the original text are ignored by the format designed to “exchange translation memory data between tools and/or translation vendors with little or no loss of critical data during the process” (Savourel, 2005).

are listed in both lists must be in the public domain or with a free license. Other texts will remain private to the translator who uploaded them.

The screenshot shows the TraduXio website interface. At the top, there are navigation tabs: "Œuvres", "Mes traductions", and "Concordance". On the left, there is a featured image of a woman in an orange sari and a man in a white shirt sitting on a stone ledge by a body of water. Below the image is a quote in French: "Du point de vue de la communauté, le langage n'est pas simplement un fait social, [...] il est par l'altérité [...] le fondement de toute association humaine." attributed to Eugenio Coseriu. Below the quote is the text "Le site des traducteurs de textes culturels". In the center, there is a vertical dotted line separating the main content from a list of authors. The authors are grouped by language: ALLEMAND (Goethe, Kant, Meyrink, Schopenhauer), ANGLAIS (Carroll, Lovecraft, Machen), FRANÇAIS (Baudelaire, Lautréamont), GREC ANCIEN (Athénée), and RUSSE (Dostoïevski, Tchekhov). On the right side, there is a sidebar titled "NOUVELLES TRADUCTIONS FRANÇAISES" with a list of items including "Преступление и наказание", "Anthropologie statt Metaphysik", "The lamp", "Казаки I", "Жёны артистов", "Eristische Dialektik", "Vorspiel auf dem Theater", "Inni Mér Syngur Vitleysingur", "Die transzendente Ästhetik", "Новейший письмовник", and "Δελτοσοφιστών". At the bottom left, it says "À propos de TraduXio" and at the bottom right, "Droits d'auteur".

Figure 3. A 'texts-base' rather than a 'data-base' (mock-up).

In prior works (Bénel & Lejeune, 2009), we offered a glimpse into 'Humanities 2.0' and how it could be brought by a digital workbench where documents would be interpreted by researchers from all over the world, and where their *viewpoints* could be compared. Contrary to our first experimentations where viewpoints were annotation structures, the viewpoints in this project are translations. More precisely, they are composed of translations of text fragments (see Fig.4). Even if our software could be used as *Wikisource*, our focus is not on getting an authoritative translation from different contributors, but to let the meaning of the text 'pop' from its confrontation with translations by different people in different languages.

Œuvres Mes traductions Concordance

Aurélien Philippe

Lovecraft, *Fungi from Yuggoth*

The Lamp

We found the lamp inside those hollow cliffs
Whose chiseled sign no priest in Thebes could read,
And from whose caverns frightened hieroglyphs
Warned every living creature of earth's breed.

No more was there - just that one brazen bowl
With traces of a curious oil within;
Fretted with some obscurely patterned scroll,
And symbols hinting vaguely of strange sin.

Little the fears of forty centuries meant
To us as we bore off our slender spoil,
And when we scanned it in our darkened tent
We struck a match to test the ancient oil.

It blazed - great God!... But the vast shapes we saw
In that mad flash have seared our lives with awe.

Lovecraft, *Fonges de Yuggoth*

La lampe

Nous trouvâmes la lampe à l'intérieur de ces falaises creuses
Aux signes ciselés qu'aucun prêtre de Thèbes n'avait su lire,
Et des cavernes desquelles, les hiéroglyphes effrayés
Mettaient en garde toute c

A propos de TraduXio Droits d'auteur

Figure 4. A different translation for every translator (mock-up).

Œuvres Mes traductions Concordance

living creature of earth's breed

Genre : narratif poétique Époque : antique contemporaine Auteur : anonyme Lovecraft

Domaine public	We found the lamp inside those hollow cliffs Whose chiseled sign no priest in Thebes could read, And from whose caverns frightened hieroglyphs Warned every living creature of earth's breed .	Nous trouvâmes la lampe à l'intérieur de ces falaises creuses Aux signes sculptés qu'aucun prêtre ne déchiffra jamais Et les effrayants hiéroglyphes de ces cavernes étaient Un avertissement pour toute créature vivante de l'espèce humaine.	© François Truchaud, Éd. Néo
Domaine public	Bring forth with thee every living thing that is with thee, of all flesh, both of fowl, and of cattle, and of every creeping thing that creepeth upon the earth ; that they may breed abundantly in the earth , and be fruitful, and multiply upon the earth .	Fais sortir avec toi chaque animal toute chair tout ce qui vole les bêtes tout ce qui se faufile ras du sol pour faire une foule sur la terre féconde et multiple sur la terre.	© Frédéric Boyer & Jean L'Hour, Éd. Bayard

Figure 5. In-context translations search (mock-up).

Provided that translations are stored fragment by fragment, it is possible to compute a multilingual concordance for every given expression (see Fig.5). For example, a French translator of *The Lamp* (see Fig.4) could search for other translated fragments containing “living creature of earth's breed”. The concordance would be made from every translation done on the software (see Fig.5). Even copyrighted texts could be displayed provided that the extract respects ‘fair use’: in particular, the short quotations would link to online stores whereto buy the complete work. In contrast, public domain texts (or texts with a free license) would be opened by such a link.

The link to the original context is not the only difference with a translation memory. First, because a random set of texts is not a real corpus, the concordance could be filtered on genre, time or author. Second, our software does not store short phrases but long fragments. They are not intended to be reused but only to give the translator insights. In our example, the sentence from *the Genesis* does not contain a translation of “living creature of earth's breed” that could be copied, but it suggests that the phrase connotes a creation story, and therefore could be translated by “toute créature vivante engendrée par la terre”.

To store fragments of texts rather than out-of-context sentences is not only important for translation quality but also for authors' rights. In the platform we designed, there are neither derivative data that would be so ‘objective’ that they would not have any author anymore, nor anthologies that would be the site owner's property. There are only texts that have authors. Once a text is removed, every fragment is removed too.

Because the license of the text affects the handling of the text on both the main page and the concordance, and because the attribution is displayed on both the bilingual pages and the concordance, all of these pages would allow a guest to report a licensing or attribution error (see Fig.6).

Œuvres Mes traductions Concordance

SIGNALER UNE MISE EN LIGNE ABUSIVE

à la page :

Référence erronée ou manquante (auteur, éditeur, licence...)
 Diffusion d'une oeuvre protégée non-assimilable à une citation courte

Détails (optionnel) :

Système anti-spam :

Adresse de courrier électronique (optionnel) :

A propos de TraduXio Droits d'auteur

Figure 6. Copyright compliance management (mock-up).

FUTURE RESEARCH DIRECTIONS

Before implementing the platform, we tried to test our hypotheses by observing a face-to-face translators' working session. The attendants were from a lab specialized in antique theater, sports and shows. They started to collectively translate an ancient Greek book about those topics.

The observed practices (illustrated here with verbatim quotations) corroborated the need for comparing translations of the same text, especially ancient and modern translations because criteria of what is a good translation changed. But they emphasized that the translations to study are not necessarily in the 'target' language. Concordances will have to be designed like that as well.

*The 18th [century] guy, he translates anyhow but he knew Greek well!*⁶

**I* look at the Italian translation.*⁷

The need for a concordance was corroborated, but, what was more surprising was the value of a concordance *on the translation being made* to ensure coherence. Therefore the concordance could even be used on a platform with a single text. Because our main challenge for the next year will be the adoption of the platform by translators, this could be a critical advantage for us as a 'bootstrap' use.

*'Poïeticos' is that there elsewhere?*⁸

*'Kinesis', we already met it and we already translated it by 'motion'.*⁹

What was quite unexpected in a discipline where expertise is considered as the attribute of single scholars, was the team work being performed. For each session, one attendant had to prepare a rough translation, which had to be refined and augmented by the group. To handle such a cooperation, it is still unclear if it should be modeled as a single translation by a group (maybe with revisions history) or as a "patchwork" of translations done by single authors.

*It is Valérie who translated [it]. I filled in what was missing.*¹⁰

The type of the text being translated highlighted a few specific needs. First, as a text from the antiquity, the text has not an original typographical structure. Instead, punctuations, standard numbering, and section headings (called 'rubrics' because they were written in red) were added by the philologists who 'established' the text. Section headings and numbering could be handled as 'tags' on fragments. However editing the inner structure of the original text while preserving references is still a research challenge (Nelson, 1999). Second, because the 'original' text is in fact the result of the manuscripts transcriptions, our platform would probably need to be integrated with existing digital philology tools for transcribing a digital facsimile into plain text and preserving the variants from different manuscripts (Bozzi, 1993; Calabretto & Bozzi, 1998). Here again, referencing parts of such a composite document will be quite challenging.

⁶ Verbatim quotation: "Le type du XVIII^e, il traduit n'importe comment mais il connaissait bien le grec !"

⁷ Verbatim quotation: "Moi, je regarde la traduction italienne."

⁸ Verbatim quotation: "'Poïeticos' est-ce que ça y est ailleurs ?"

⁹ Verbatim quotation: "'Kinesis', on l'a déjà rencontré et on l'avait déjà traduit par 'mouvement'."

¹⁰ Verbatim quotation: "C'est Valérie qui a traduit. J'ai complété ce qui manquait."

Last but not least, what we observed on a physical session of an existing group cannot be directly extrapolated to distant communities that the platform aims at creating. In particular, we will have to foster the progressive building of trust between members (e.g. through new members sponsorship, participative promotion of translations, etc.).

CONCLUSION

We tried to understand why computer-assisted translation tools are avoided in cultural sciences. It appeared that machine translation embeds a very naive language theory, ignoring that there are usually no equivalent forms in two languages, mainly because of connotations and different available corpora. Translation memories would be more interesting since they are based on practical norms in use by humans rather than on formal and universal rules. However, because they are thought as a 'database', with the 'objective data' being the translated segments, they miss contexts, genres, authors, translators... everything a translator needs when he studies a concordance, prior translations, or a dictionary for classical studies.

By taking these objections into account, we designed a 'texts-base' rather than a 'database', a digital concordance rather than a translation memory. The software was designed in a 'Web 2.0' way, which means that all the content would be brought by the users themselves. We paid a lot of attention at the human interface so that it would allow the users to give access to useful published and unpublished content while not harming the potential value of it (as stated by the 'fair use' doctrine).

We tested our design by observing a face-to-face translators' working session. While it corroborated most of our hypotheses, it forced us to refine some of them and to list future research directions, such as computer-supported cooperative work models for cooperative translation and community trust building, as well as hypermedia models for digital philology.

REFERENCES

Bénel, A., & Lejeune, C. (2009). Humanities 2.0: documents, interpretation and intersubjectivity in the digital age. *International Journal of Web Based Communities*, Vol 5, No 4, 562–576. Retrieved November 15, 2010, from http://publi.tech-cico.fr/paper_details.php?paperid=106

Bozzi, A. (1993). Towards a Philological Workstation. *Revue Informatique et Statistique dans les Sciences humaines*, Vol XXIX, No 1-4. Retrieved November 15, 2010, from <http://promethee.philo.ulg.ac.be/RISHPdf/Annee1993/Articles/ABozzi.pdf>

Calabretto, S., & Bozzi, A. (1998) The Philological Workstation BAMBI (Better Access to Manuscripts and Browsing of Images). *Journal of Digital Information*, Vol 1, No 3. Retrieved November 15, 2010, from <https://journals.tdl.org/jodi/article/viewArticle/10/20>

Lacour, P. (2010). Portrait of the intellectual as a DJ: Wikipedia and the question of scientific expertise, *La Vie des Idées*, March 12, 2010. Retrieved November 15, 2010, from <http://www.laviedesidees.fr/Portrait-of-the-intellectual-as-a.html>

Nelson, Th. H. (1999). Xanalogical Structure Needed Now More Than Ever. *ACM Computing Surveys*, Volume 31, Issue 4, ACM Press. Retrieved November 15, 2010, from http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/60.html

O'Reilly, T. (2005). *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. September 30, 2005. Retrieved November 15, 2010, from <http://oreilly.com/web2/archive/what-is-web-20.html>

Passeron, J.-C. (2006). *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*. Albin Michel.

Rastier, F. (1998). Sens et signification. *Protée, printemps 1998*, 7–18.

Rastier, F. (2007). La traduction : Interprétation et genèse du sens. *Revue Texto !*. Retrieved November 15, 2010, from <http://www.revue-texto.net/index.php?id=2202>

Savourel, Y. (2005). *TMX 1.4b Specification*. OSCAR Recommendation. Localisation Industry Standards Association, April 26, 2005. Retrieved November 15, 2010, from <http://www.lisa.org/fileadmin/standards/tmx1.4/tmx.htm>

ADDITIONAL READING SECTION

Bénel, A., Zhou, C., & Cahier, J.-P. (2010). Beyond Web 2.0... And Beyond the Semantic Web. In D. Randall & P. Salembier (Eds.), *From CSCW to Web 2.0: European Developments in Collaborative Design*. London: Springer Verlag. 155–171. Retrieved November 15, 2010, from http://benel.tech-cico.fr/publi/benel_SPRINGER_2010.pdf

D'Iorio, P. (2007). Nietzsche on New Paths: The HyperNietzsche Project and Open Scholarship on the Web. In M. C. Fornari & S. Franzese (Eds.), *Friedrich Nietzsche. Edizioni e interpretazioni*. Pisa : ETS. Retrieved November 15, 2010, from <http://www.hypernietzsche.org/doc/files/new-paths.pdf>

Duteil-Mougel, C. (2004) Introduction à la sémantique interprétative. *Texto !*. Retrieved November 15, 2010, from http://www.revue-texto.net/1996-2007/Reperes/Themes/Duteil/Duteil_Intro.html

Granger, G.-G. (1994). *Les conditions protologiques des langues naturelles. Formes, opérations, objets*. Paris: Vrin.

Lacour, P. (2009). Diskursivität. Zur logischen Erklärung der Hermeneutik Ricoeurs. *Energeia – Online Zeitschrift für Sprachwissenschaft und Sprachphilosophie*. Retrieved November 15, 2010, from <http://www.uni-tuebingen.de/uni/nrk/coseriu/energeia/zeitschrift/2009/diskursivitaet.html>

Lacour, P. (2010). *La nostalgie de l'individuel. Essai sur le rationalisme pratique de Gilles-Gaston Granger*. Collection “Mathèsis”. Paris : Vrin.

Tasman P. (1957). Literary data processing. *IBM Journal of Research and Development*, vol. 1, No 3, 249-256.

Virbel, J. (1994). Annotation dynamique et lecture expérimentale : Vers une nouvelle glose ? *Littérature*, No 96. 91–105.

Zhou, C., & Bénel, A. (2008). *From the crowd to communities: New interfaces for social tagging*. Paper presented at the international conference on the design of cooperative systems, Carry-le-Rouet, France. Retrieved November 15, 2010, from http://benel.tech-cico.fr/publi/zhou_COOP_08.pdf

KEY TERMS & DEFINITIONS

Concordance: Alphabetical list of the principal words used in a corpus with their immediate contexts. Depending on their purpose, the extracts can be displayed in different ways (as a whole sentence, centered on the word, with a translation, etc.). The first known concordance was established by Dominicans on the 13th century. They are now computed automatically (Tasman, 1957).

Hermeneutics: The theory and practice of texts interpretation, historically created to decipher sacred texts, and generalized by philosophers (Schleiermacher, Dilthey, Gadamer...) as a foundation for the epistemology and methodology of humanities and social sciences. It constitutes the reflexive answer to the challenge of the polysemy expressed by a discourse under the complex (transphrastic) form of a text (Lacour, 2009). An interpretation stresses one – and only one – meaning. The polysemy of the text therefore logically unfolds into a variety of conflicting interpretations that constitute a polemic space of competing viewpoints. It also gave credit and legitimacy to an alternative approach of language, focusing not so much on rules and generalities as on creativity and singularity (Berner et Thouard, 2010).

Humanities 2.0: A provocative term to prompt thinking about the links that could exist between ‘Web 2.0’ and the social and human sciences. By focusing on documents (as opposed to data), on interpretation (as opposed to deduction), and on intersubjectivity (as opposed to consensus), Web 2.0 has surprising analogies with hermeneutics, and thus might better meet the needs of social and human sciences than traditional information technologies that are historically bound with logical positivism (Bénel et Lejeune, 2009).

Viewpoint: More than an opinion (anything goes) and less than a rigid and fixed semantics (Bénel et al., 2010; Zhou et Bénel, 2008).

Web 2.0: Contrary to a common misconception, the term does not refer to a technical innovation but to a change in Web use (O’Reilly, 2005). A few years after the bursting of the dot-com bubble, Tim O’Reilly noted that the most popular Web sites shared several characteristics. One in particular was to “harness collective intelligence”. Indeed, from books commentaries on *Amazon*, to ‘tagged’ pages on *Del.icio.us* or articles on *Wikipedia*, these contents are generated by the users rather than by the service provider. Internet participation, formerly reserved to a few specialized communities (open-source software, open directory project, open archives...) has become the major use model.