

# Collaborative computer-assisted translation applied to pedagogical documents and literary works

*Ruslan KALITVIANSKI, Christian BOITET, Valérie BELLYNCK*

GETALP-LIG, 41, rue des Mathématiques,  
38400 St Martin d'Hères, FRANCE

[ruslan.kalitvianski@imag.fr](mailto:ruslan.kalitvianski@imag.fr), [christian.boitet@imag.fr](mailto:christian.boitet@imag.fr),  
[valerie.bellynck@imag.fr](mailto:valerie.bellynck@imag.fr)

## ABSTRACT

This paper showcases three applications of GETALP's iMAG (Interactive Multilingual Access Gateway) technology. iMAGs allow internet users to navigate a selected website in the language of their choice (using machine translation), as well as to collaboratively and incrementally improve the translation through a web-based interface. One of GETALP's ongoing projects is MACAU (Multilingual Access and Contributive Appropriation for Universities), a platform that allows users to reuse existing pedagogical material to generate adaptive content. We demonstrate how student- and teacher-produced lecture notes can be translated into different languages in the context of MACAU, and show the same approach applied to textbooks and to literary works.

## Применение коллективного автоматизированного перевода к учебным материалам и литературным работам

## РЕЗЮМЕ

Эта статья демонстрирует три применения технологии iMAG (Interactive Multilingual Access Gateway) лаборатории GETALP-LIG. iMAG-и позволяют Интернет-пользователям посещать избранный веб-сайт на желаемом языке благодаря машинному переводу, а так же постепенно коллективно улучшать перевод через веб-интерфейс. Один из текущих проектов GETALP - MACAU (Multilingual Access and Contributive Appropriation for Universities), платформа, позволяющая пользователям использовать существующие педагогические материалы, чтобы генерировать персонализированные уроки. Мы показываем, как конспекты студентов и учителей могут быть переведены на различные языки в контексте MACAU, и демонстрируем применение данного подхода к учебникам и литературным работам.

---

KEYWORDS : MACHINE TRANSLATION, COMPUTER-ASSISTED TRANSLATION, PEDAGOGICAL DOCUMENTS, INTERACTIVE MULTILINGUAL ACCESS GATEWAY, iMAG, COLLABORATIVE TRANSLATION, POST-EDITING.

Ключевые слова: МАШИННЫЙ ПЕРЕВОД, АВТОМАТИЗИРОВАННЫЙ ПЕРЕВОД, УЧЕБНЫЕ МАТЕРИАЛЫ, iMAG, КОЛЛЕКТИВНЫЙ ПЕРЕВОД, ПОСТРЕДАКТИРОВАНИЕ..

---

# 1 Interactive Multilingual Access Gateways

An iMAG (interactive Multilingual Access Gateway) to a website is a web service that allows users to access the site in a language of their choice (by translating it with one or several machine translation engines) and to improve the translation by post-editing (Boitet et al, 2008).

When the user chooses to display the page in a new language, the textual segments of the page are substituted either by a new translation or, if the page has been translated before, by their best translation retrieved from the translation memory. The translation of a segment can be edited by hovering the cursor above the segment. This brings up a bubble containing the segment in the source language, an editing zone and a choice of a score to be assigned to the post-edited translation. Figures 1 and 2 show the interface of iMAG-COLING.



Figure 1: COLING 2012 website accessed in Japanese through iMAG-COLING

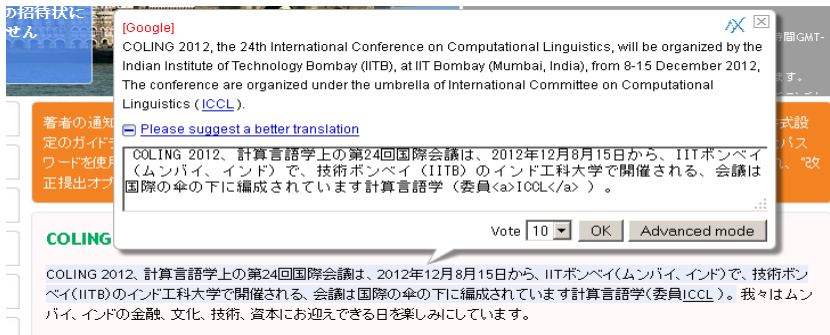


Figure 2: post-edition of the translation of a segment.

## 2 Translation of pedagogical content

One of our ongoing projects is MACAU (Multilingual Access and Contributive Appropriation for Universities), a platform that would allow its users to reuse existing pedagogical materials to improve them and generate pedagogical content that best fits their needs, given their preferences and their level of domain knowledge. This includes multilingualism, as well as choices of presentation, including topics, types of material and levels of abstraction.

IMAGs can be used by foreign students to access course content in their language and to collectively improve the translations when needed. A simple scenario within the context of MACAU would the following.

First, a set of documents is received. We can, for instance, consider a teacher-produced LaTeX file containing a course on computational complexity theory, and two MS Word documents, containing student-produced lecture notes.

These documents are converted to an iMAG-compatible format, which for the moment is html. In this example, we use HeVeA for the LaTeX → html conversion, and MS Word's built-in export tool. The resulting html documents can now be annotated with semantic information, allowing for later selective access and extraction. Annotations can indicate the level of abstraction of a section, its difficulty, pre-requisites, etc. Here, we use simple “span” tags to indicate the type of a section of the document (e.g. definition, example, illustration).

```

<P STYLE="margin-bottom: 0in"> =  $n^2 - 2n(p+1) + n$ </P>
<P STYLE="margin-bottom: 0in"><BR>
</P>
<span class="macau" type="theorem">
<P STYLE="margin-bottom: 0in"> <U>Théorème</U> &nbsp;: Soit une MT
Z<SUB>k</SUB> à k bandes. Il existe une MT simple Z simulant <I>n</I>
pas de calcul de Z<SUB>k</SUB> par au plus <I>2n × (n + 2k + 1)</I>
pas de calcul de Z</P>
</span>
<P STYLE="margin-bottom: 0in"><BR>
</P>

```

Figure 3: semantic annotation of document sections.

The documents are then placed in a repository accessible to iMAG-MACAU that can be visited at <http://service.aximag.fr/xwiki/bin/view/imag/macau-fr>. Users can now navigate and post-edit the documents in different languages. Since iMAGs preserve the underlying html code of a page, the annotations are never lost and whatever content generated from these files using the semantic annotations, has its segments translated. We are currently developing tools that would allow reversible conversion in order to generate translations of documents in their initial formats.



Figure 4: a bilingual view of the document being translated.

### 3 Translation of books

Many universities now release their educational materials free of charge, however they are usually unavailable in more than one or two languages, or available for a fee. IMAGs provide a rapid, convenient and cost-effective alternative for obtaining multilingual versions of educational materials such as textbooks that are converted into an iMAG-compatible format. As an example, readers are invited to visit the demonstration iMAG for the book “Bioelectromagnetism” by Jaakko Malmivuo and Robert Plonsey at <http://service.aximag.fr/xwiki/bin/view/imag/BEMBOOK>.

The use of volunteer workforce for the translation of literary works is a novel approach. Fig. 5 demonstrates a chapter of Rohit Manchanda's “Monastery, Sanctuary, Laboratory” being translated from English to Hindi.



Figure 5: translation of Rohit Manchanda's "Monastery, Sanctuary, Laboratory: 50 years of IIT-Bombay"

Readers are invited to contribute at <http://service.aximag.fr/xwiki/bin/view/imag/xan-en?u=http://www-clips.imag.fr/geta/User/christian.boitet/iMAGs-tests/en/ManchandaArticles/IITB-Monastery-Sanctuary-Laboratory>

## References

- Boitet, C., Bellyneck, V., Mangeot, M., and Ramisch, C., (2008). *Towards Higher Quality Internal and Outside Multilingualization of Web Sites*. In Proceedings of ONII-08 (Summer Workshop on Ontology, NLP, Personalization and IE/IR), Mumbai, CFILT, IITB.
- Maranget. L., *H<sup>E</sup>V<sup>E</sup>A*, version 2.00. Sources and documentation available at <http://hevea.inria.fr/>
- Maranget. L., *H<sup>E</sup>V<sup>E</sup>A*, un traducteur de L<sup>A</sup>T<sub>E</sub>X vers HTML en Caml. Available at <http://pauillac.inria.fr/~maranget/papers/hevea/>
- Malmivuo, J., and Plonsey, R., (1995). *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*, Oxford University Press, New York <http://www.bem.fi/book/index.htm>
- Manchanda, R., (2008). *Monastery, Sanctuary, Laboratory: 50 Years of IIT-Bombay*, Macmillan India.