



Lost in Translation – the challenges of multilingualism

Martin Flynn
Victoria and Albert Museum
London, United Kingdom
E-mail: m.flynn@vam.ac.uk

Meeting:

79 — Tackling the challenges of multilingualism in the arts: catalogues, databases, digital collections and other material in the global context — Art Libraries Section

Abstract:

Art information increasingly operates in a global environment but without the benefit of a lingua franca which some other disciplines enjoy. Information continues to be generated in a variety of vernacular languages which limits cross-cultural understanding and exchange. Easy access to cheap, instantaneous and accurate translation would be the ideal but this is unlikely in the foreseeable future. A selection of online initiatives and tools are assessed in an attempt to identify potential trends and opportunities.

Keywords : art information, machine translation, non-verbal search, crowdsourcing

Introduction

Art production, collecting and documentary information have always had a significant international dimension and this continues to be the case in our increasingly globalised world. However, unlike other subject areas such as science, aviation and information technology, it has not evolved its own lingua franca. Montgomery demonstrates the dominance of the English language in science¹. Today 80-90% of papers in scientific journals are now written in or translated into English and this has increased from 65% in the 1980s. This phenomenon is an enormous bonus for international co-operation and information exchange.

¹ Scott L Montgomery, 'English and Science: realities and issues for translation in an age of expanding *lingua franca*' *Journal of Specialised Translation*, January 2009: http://www.jostrans.org/issue11/art_montgomery.php

In the absence of anything resembling a lingua franca in the arts we need to turn to other opportunities for information transfer and understanding in a global art environment. The following is a selective survey of some potentially relevant initiatives to provoke further thought and discussion on this matter and it neither aims nor claims to be exhaustive or definitive.

Translation

One of the obvious solutions for library problems with language issues is to have access to easy translation of relevant documents into the required language. However, the creation of high quality translation is extremely time-consuming and consequently expensive. It is often assumed that to be bilingual is enough to be a good translator. However, a fully competent translator needs to be not only bilingual but also bicultural. They need to have a profound understanding of the etymological and idiomatic correlations between source and target languages. They require a familiarity with the subject matter of the text being translated as well. It is also preferable for translation to be produced from the translator's second language (the source language) into their first language (the target language) to obtain the optimal results. Translators also need to have a highly developed sense of when it is appropriate to paraphrase (translate literally) and when to paraphrase (rewrite) so as to ensure accurate, rather than chance, equivalents between source and target language texts.

Claude Piron, an experienced translator with both the United Nations and the World Health Organisation explains the biggest obstacle faced by translators². On average it will take a translator a whole workday to translate five pages of text. Approximately 90% of this work will normally be achieved within an hour but the remaining 10% will require six more hours of work. The reason for this is that the translator will normally have to do extensive research to resolve the ambiguities in the source text through the grammatical and lexical challenges of the target language. Coincidentally these very ambiguities form one of the biggest challenges to the development of the semantic web³. If language was more scientific and predictable both translation and the development of a semantic web would be infinitely easier. However, organic growth, exceptions to rules and unpredictability are linguistic norms.

Machine Translation

The challenges of translation and the free transfer of information in all disciplines across the language divide might be considerably mitigated by current developments in computer translation. There are two main approaches to machine translation: the dictionary, grammar and rules approach and statistical matching. Popular online translation services such as Google Translation and Yahoo! Babel Fish use text corpora from organisations which routinely deal in multiple languages such as the Canadian government, the European Union and the United Nations. They identify statistically significant patterns in existing translations in order to produce their translated texts. Google recently improved their translation ability

² Claude Piron, *Le défi des langues* (The Language Challenge), Paris, L'Harmattan, 1994.

³ Clay Shirky, The Semantic Web, Syllogism and Worldview:
http://www.shirky.com/writings/semantic_syllogism.html

by inputting 200 billion words from United Nations translated material in order to train their system⁴. The amount of available translated documents in any particular language will inevitably affect this ability to detect patterns. As a result French to English translation tends to work relatively well by comparison to other language pairings⁵. Statistical machine translation also tends to be more effective when operating in subject areas with restricted vocabularies such as weather forecasting and sport.

However, the overall user experience of automated translation services tends to be very poor, often including apparently nonsensical and obvious errors. Swapping common terms for non-equivalent terms in other languages as well as inverting sentence meaning are also common. Piron estimates that, at most, such machine translation will cater for about 25% of a translator's needs with the remaining 75% requiring human intervention to produce publishable quality translation. However, machine translation is continually improving and its value in giving an approximation of the meaning of a text can be invaluable for functional needs, particularly for those who have some familiarity with the language of the source text.

There are a variety of other language aids available on the internet. Google Translate provides an aural pronunciation of words and phrases in both the source and the target languages. There are also a variety of free pronunciation sites on the web such as Howjsay⁶.

Crowdsourcing

There are some interesting developments in the area of leveraging the skills and interest of online users in relation to language translation. Microsoft Research India has developed a semi-automated translation facility, WikiBhasha, in order to encourage more articles in Indian languages on Wikipedia⁷. Wikipedia articles are developed from scratch in each individual language. However there is relatively limited content currently online in any of the Indian languages. Users of WikiBhasha can populate Wikipedia with translations of articles from other source languages into target Indian languages. The user is prompted not only by machine translations in the target language but also by alternatives created by human intervention. This is effectively a form of hybrid human and machine translation. As with other Wikipedia input users can contribute anything from a few words to a whole article.

Google Translate has a 'contribute a better translation' link on their results page for users to suggest improvements to any particular Google translation and to upload these. The aim is to use this human intervention to counteract the flaws inherent in machine translation and to further improve results.

Language learning websites such as Livemocha leverage the skills of existing learners to comment on and correct other student's work and pronunciation⁸. Teachers in one language

⁴ Google Translate Video: http://translate.google.co.uk/about/intl/en_ALL/

⁵ Google Translate, Wikipedia: http://en.wikipedia.org/wiki/Google_Translate

⁶ Howjsay: <http://www.howjsay.com/>

⁷ WikiBhasha Video: <http://www.wikibhasha.org/wikibhasha/tour.htm>

⁸ Livemocha: <http://www.livemocha.com/pages/about>

can become learners in another. It is a learning network where anyone in its gift economy can critique and help another student.

When Facebook launched its Spanish and German versions it didn't translate the site itself into those languages. Instead it created a platform for translation and handed the work over to potential users who did this work for free⁹. Librarians have had a long tradition of co-operation in areas such as inter-lending and standards development. Could they now move on to participate in the new networked gift economy by supporting one another, in the spirit of crowdsourcing, by undertaking small translating tasks online on a reciprocal basis to improve global discovery and understanding?

Visual searching

Methods of searching for information without the intervention of language are very much in their infancy. The Quaero project was initially announced five years ago¹⁰. It is a research and development programme funded by the European Union. This public and private partnership aims to facilitate multimedia and multilingual searching. However, all that has appeared to date are some online demonstrators. Given this lengthy timelag in a rapidly developing digital environment and the lack of any tangible progress its future looks distinctly unpromising.

The Victoria and Albert Museum in London is working on a joint development, the FABRIC project, with Dundee University, the retail outlet Liberty Art Fabrics and the IT company System Simulation. Dundee has developed the computer software and funding is provided by the UK Technology Strategy Board. The system automatically organises sets of images by elements such as colour, visual texture and shape. Items such as photographs, furniture, wallpaper and ceramics have been selected and presented on a beta site. The user is encouraged and enabled to make a variety of visual matchings through the functionality of the software. This test site facilitates and encourages user feedback¹¹. The underlying concept is based on the premise that designers are more likely to use images rather than textual descriptions to spark and fuel their creativity.

Conclusion

The initiatives outlined above begin to offer hope for overcoming the barriers of language to enable more effective international sharing of art information. However limited their current potential, it is essential to remember that we are only in a position to assess their present state of development and it is therefore impossible to predict what lies ahead for them given the exponential rate of development of digital technology. As there do not appear to be any immediately apparent parallel initiatives in the libraries this may be an appropriate challenge for the sector to rise to in order to secure a more central role in the networked world.

⁹ Jeff Jarvis, *What Would Google Do?*, Harper Business, p34

¹⁰ Quaero: http://www.quaero.org/modules/movie/scenes/home/index.php?FUSEBOX_LANG=2

¹¹ FABRIC: http://collections.vam.ac.uk/information/information_fabricvisualiser

Biographical Details:

Currently Head of Information Services at the V&A, London. Previously worked in public libraries, theatre management and publishing. Secretary of the Art Libraries Section of the International Federation of Library Associations (IFLA) and Manager of the Web and Communications Group of the UK Museum Librarians and Archivists Group (MLAG).