



City Research Online

City, University of London Institutional Repository

Citation: Bawden, D. (2010). Documenting Babel. *Journal of Documentation*, 66(2), pp. 177-178.

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <http://openaccess.city.ac.uk/3127/>

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Editorial

Documenting Babel

Languages, in one guise or another, have been a constant feature of the landscape of the information sciences for many years.

There are, for example, the various artificial 'languages' – more usually thought of as notations, nomenclatures or ontologies - which have been devised to describe such things as chemical structures and reactions, medical diagnoses and treatments, and the burgeoning data-rich fields of modern biology. There is the presence of linguistics as a subject of seemingly perpetual potential relevance to the information sciences. As Sparck Jones and Kay (1973, p. 1) put it in their seminal textbook: "linguistics and information science are natural bedfellows ... but there has been relatively little contact between the two fields": the situation has not changed much in the intervening decades. There is the now ubiquitous searching of 'full text' databases, requiring a greater or lesser amount of 'intelligent' processing of the natural languages in which the content of such databases are couched.

But primarily, there is the continued need for handling communication of information in all of the world's languages. Neither the earnest advocacy of 'universal' or 'auxiliary' languages, from Leibnitz' logic-based *characteristica universalis* to Esperanto, nor the long-anticipated advent of English as a *de facto* global language (Crystal 2003), has reduced the demand for support for national and local languages, as the provision for 23 official languages in the European Union testifies.

This naturally has consequences for research and practice in the information sciences. A facility with languages other than one's own has always been one of the requirements of the practising librarian and information officer, even in the traditionally language-averse United Kingdom. Sadly, the requirement for some facility with two languages other than English, a requirement when I studied information science at Masters level, has long gone from the UK, though an equivalent requirement is still largely present in Continental Europe. This manifested itself in a variety of detailed language tools for the information professions, Allen's 1975 *Manual of European Languages for Librarians*, being a typical example.

In research terms, language issues have stimulated work on a variety of topics. An early example was the study of the value of 'cover-to-cover' translations of scientific journals, particularly from the Russian language following the shock to the Western scientific complacency caused by the *Sputnik* satellite of 1957 (Tybulewicz 1970). Other long-standing concerns, in the English-speaking world at least, were focused on the 'language barrier', the belief that valuable information, particularly in scientific, technical and medical subjects, was being missed because it was not published in the English language (see, for instance, Hutchins, Pargeter and Saunders 1971, Chan 1977, Thorpe, Schur, Bawden and Joice 1988). More recently, attention has been focused on such topics as the information practices of translators, natural language processing and cross-language information retrieval; Table 1, below, shows some examples of recent *Journal of Documentation* articles reporting such research, as an indication of its variety.

These thoughts were stimulated by my attending the INFuture conference in Zagreb, Croatia, in November 2009 [<http://infoz.ffzg.hr/INFuture/Conference.aspx>]. A substantial proportion of this conference, which dealt with the future of information science, was devoted to language technologies – including machine-aided translation and natural language processing - and to languages issues in general. The topics covered included the European Union's CLARIN (Common Language Resources and Technology Infrastructure) project, which aims to compile a series of digital archives with data sources for language-based materials (text and speech corpuses, dictionaries, etc.) together with language and speech technology tools. Particularly aimed at academic users in the arts and social sciences, CLARIN adopts the philosophy that all languages – irrespective of the number of speakers or of their commercial importance - are of equal importance.

It seems clear that the predictions, or fears, of the adoption of artificial languages, and of the ubiquitous adoption of any single one, are very far from fulfilment. We may expect that these issues will be an important feature of the information research agenda for the foreseeable future.

David Bawden

References

- Allen, C.G. (1975), *A manual of European languages for librarians*, London: Bowker
- Chan, G.K.L. (1977), Mushroom poisoning, thioctic acid and the foreign language barrier, *Aslib Proceedings*, 29(6), 237-240
- Crystal, D. (2003), *English as a global language (2nd edn.)*, Cambridge: Cambridge University Press
- Hutchins, W.J., Pergeter, L.J. and Saunders, W.L. (1971), University research and the language barrier, *Journal of Librarianship*, 3(1), 1-25
- Sparck Jones, K. and Kay, M. (1973), *Linguistics and Information Science*, New York NY: Academic Press
- Thorpe, R.A., Schur, H., Bawden, D. and Joice, J.R. (1988), The foreign language barrier: a study among pharmaceutical research workers, *Journal of Information Science*, 14(1), 17-24
- Tybulewicz, A. (1970), Cover-to-cover translations of Soviet scientific journals, *Aslib Proceedings*, 22(2), 55-62

Table 1

Examples of recent language-related papers in Journal of Documentation

- White, M.D., Matteson, M. and Abels, E.G. (2008), Beyond dictionaries: understanding information behaviour of professional translators. *Journal of Documentation*, 64(4), 576-601
- Pinto, M. and Sales, D. (2008), INFOLITRANS: a model for the development of information competences for translators, *Journal of Documentation*, 64(3), 413-437
- Airio, E. (2008), Who benefits from CLIR (cross-language information retrieval) in web retrieval, *Journal of Documentation*, 64(5), 760-778
- Peng, F. and Huang, X. (2007), Machine learning for Asian text classification, *Journal of Documentation*, 63(3), 378-397
- Talvensaari, T., Lauriika, J., Järvelin, K., and Juhola, M. (2006), A study on automatic creation of a comparable document collection in cross-language information retrieval, *Journal of Documentation*, 62(3), 372-378