

**SCHOOL OF LANGUAGES, CULTURES AND LINGUISTICS
&
THE MONASH e-RESEARCH CENTRE**

Seminar

25 March, 1:00-2:00pm

Venue: CL_13A/M2

**Potential development of a Monash
Ranking system for accurate estimation of
Internet language diversity**

*Dr Peter Gerrand, Adjunct Senior Research Fellow,
School of Languages, Cultures and Linguistics, Monash University*

SUMMARY:

In recent years both UNESCO and the OECD have recognized the public policy benefit of obtaining and publishing credible information on human language usage on the Internet. Unfortunately the techniques for estimating or measuring the relative usage of different online languages are fraught with errors in source data or the limitations of inadequate measurement tools, and as a result, they often provide unreliable results.

Apparently contradictory trend results have been published, e.g. that the use of English on the Net has remained fairly steady between 70% and 80% since 1996, versus other results indicating that the use of English has declined rapidly from 80% to about 35% over the same period. The paradox is resolved by creating a functional taxonomy to distinguish between three fundamentally different approaches to estimating human language usage on the Internet: by user activity, user profile and web presence (Gerrand 2007). This paper analyses the published methodologies in each category, and reveals the essential weaknesses in each approach.

An overview is given of UNESCO's relevant initiatives via its World Summit for the Information Society, including its Tunis Commitment to foster greater cultural diversity including the use of local and indigenous languages on the Internet.

There is an opportunity for Monash University to develop a 'Monash Internet Language Rankings' system for estimating relative language usage on the Internet as a whole, and to publish the results every six months. It will be explained why no other body has provided any credible estimates of Internet linguistic diversity since 2002. It will be suggested how such estimates can be implemented by collaboration between Monash researchers in language and linguistics, statistics, computational linguistics and IT systems development.

Reference:

Gerrand, P. (2007). "Estimating linguistic diversity on the Internet: a taxonomy to avoid pitfalls and paradoxes", *J. Computer-Mediated Communication*, Vol. 12 (4), July 2007, article 8, pp. 1298–1321, accessible at <http://jcmc.indiana.edu/vol12/issue4/gerrand.html>.

ALL WELCOME

Enquiries to: Helen Marriott (Helen.Marriott@arts.monash.edu.au; 9905 2272)