

MOST and Initiative B@bel funded research:

Language on the Internet

Sue Wright

Aston University

s.m.wright@aston.ac.uk

The origins of the research

This research was undertaken within the framework of UNESCO's multidisciplinary project *Initiative B@bel*. This initiative was conceived to promote cultural diversity in general and to seek ways to protect the lesser used languages of the world in particular. The research team¹ was asked to seek ways to amass data on actual language practice on the Internet, in order that policy making in this area could be better informed.

There were several key research questions, including:

- Can endangered and lesser used languages be promoted among their populations by establishing areas of use for them on the Internet?
- Are there new language forms resulting from using a written medium for interaction which retains many of the features of spoken discourse?
- Is the Internet contributing to any kind of language convergence?

However, the first and perhaps most important task was to assemble some hard evidence about the role of English as the lingua franca of the medium:

- Is the Internet encouraging language shift to English?

There is a widespread assumption that English is the 'dominant' language on the Internet and that its presence blocks the use of other languages. This research set out to interrogate some of the premises on which this is based and to find evidence to back or refute that impression. The research looked at two areas

- the extent to which English is the vehicle for the unprecedented flows of

¹ The core research team was David Block (Institute of Education, London), Helen Kelly Holmes (University of Limerick) and Sue Wright (Aston University).

information of the World Wide Web.

- The extent to which English is the language chosen for Internet interaction exchanges among groups and networks.

This research is presented in more detail the MOST Electronic journal and this report is a summary of some of the findings.

The modalities of the research

Current research on who uses what on the WWW is based largely on counting hits on sites. We felt that this methodology should be challenged, largely because the record of a website visit does not necessarily mean that the information on it was read. A visit can be aborted as soon as it becomes clear that it is in a language that is not known to the searcher. An alternative and more accurate research strategy would be to interview users on the languages that they use on the Internet. The validity of such research would depend, of course, on finding a sample from which it would be defensible to generalise.

The group chosen was students in higher education. We felt that university students would be among the most computer literate in their societies, would have access to the computers at their institutions and would be at a level of study where they would be asked to research independently. They would thus have the skills, the opportunity and the need to use the Internet.

The subjects were drawn from a wide range of disciplines, but all were also taking courses in English, and it was through their English language teachers that we contacted them. The rationale for this decision was to have a sample in which every respondent was at the very least bi-literate, with full literacy in the national language of education and with varying degrees of literacy in English. As they were all studying English, theoretically, they should have been in a position to use that language as a medium on the Internet if they chose to. We thus had a sample with a repertoire of languages in which to go online as well as the likely need to do so.

Moreover, since university students can be seen as the group that, in meritocracies, will constitute the future elite in their societies, data on their language behaviour would

reveal developing practices among future national and global actors.

After invitations to a very large number of universities on all continents, data were eventually collected in universities and high schools in Tanzania, Indonesia, the United Arab Emirates, Oman, France, Italy, Poland, Macedonia, Japan and Ukraine. These ten countries were felt to be representative of a number of different linguistic situations. In some, the long established national language is widely viewed as prestigious and is often learnt as a foreign language. In others, the official language is also a regional lingua franca, allowing local networks to function. Among some groups, recent independence has meant that the national language is still in the process of being elaborated for all the functions and domains of the public space. The ten countries included some with very little language diversity within their borders and some with immense variety. Thus, although there are several areas of the world where we were unable to secure partners, we felt that the spread of partners that we did have represented many of the linguistic situations that exist.

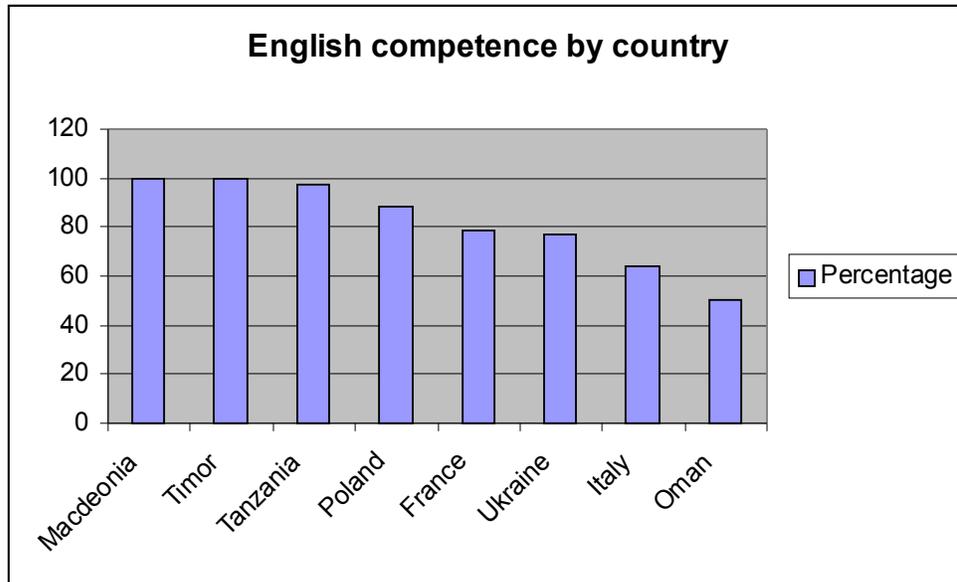
The tool that we developed to monitor language choice and behaviour was a questionnaire that asked students to give details of their language competence and to record the language and purpose of ten consecutive Internet sessions. We felt that, although the questionnaire depended on self-reporting, the detail required in the answers meant that it was unlikely there would be major misrepresentation of use. Respondents would have needed to go to great lengths to fabricate.

Findings

With more than 3000 responses to this questionnaire, we felt that we acquired some knowledge of developing practices among intellectual elites world wide.

The first finding was that largely comparable groups in the different universities reported widely differing estimations of their English language skills. The 92 Macedonian respondents all claimed levels of competence in English that would allow them to use the language with confidence for Internet purposes. 77% of the 555 Ukrainian sample felt that they had a level of English competence, adequate for using the language on the Internet. This fell to 50% among the 48 Omani respondents. So, although we had chosen

them by the fact that they were all studying English, the self-reporting showed that not all felt confidence in the language. However, this finding may have as much to do with confidence in the language and attitude towards it, as actual competence in English.



When the language of Internet use was calculated, it was discovered that, for the sample as a whole, English is used less than the national language in Internet usage. There was, however, wide divergence within the sample.

COUNTRY	PERCENTAGE OF REPORTED SESSIONS IN NATIONAL LANGUAGE	PERCENTAGE OF REPORTED SESSIONS IN ENGLISH
France	86% (French)	29%
Italy	91% (Italian)	17%
Japan	89% (Japanese)	15%
Macedonia	13% (Macedonia)	84%
Oman	100% (Arabic)	25%
Poland	84% (Polish)	49%
Tanzania	82% (Kiswahili)	90%
Indonesia	19% (Indonesian)	84%

These figures may add up to more than 100% because more than one language was used

in some sessions.

Reporting showed that the Internet is mainly used by these respondents for academic work. This was what had been expected when targeting this sample. The data further showed that recourse to English language sites on the WWW was less than had been anticipated. In the event English did not even dominate as the language of web research.

Further findings related to the language complexity of the English language sites used, the strategies the respondents adopted for overcoming technical difficulties, the composition of their networks and the presence of lesser used languages:

- The respondents, who were selected for follow up interviews, provided evidence that many of the English language websites visited were not particularly 'language rich'. For example, it became clear that some of the reported English language use related to pop music or to greeting cards sites, particularly among those who were reporting less use any way. The possibility that the language of English medium sites was formulaic or in the form of song lyrics that were not necessarily understood underscores the fact that the sample divides into two: those who use English for academic work on the Internet and those for whom English is only an aspect of their recreational use.
- Anecdotal evidence prior to the study had indicated that respondents whose languages do not use the Roman alphabet often write their language 'in English' when mailing and chatting if they do not have access to appropriate keyboards and software. For example it was noted that the Arabic users wrote Arabic employing Roman characters and attempting to replicate the sounds phonetically 'in English'. This was born out by the survey and in the interviews. The practice is not only restricted to Internet use. Interviewees claimed that text messaging on mobile phones is also widely done in this way.
- A further finding relates to the parochialism of networks. In theory, the Internet may allow us to contact the 20% of humanity that is said to have access to email, but, of course, people do not. The respondents reported that they had contacted their close networks of family and friends, and although they could move outside them, they did not report doing so regularly or often. To delve deeper into this, one set of

interviewees, were asked if they consulted the home pages of researchers in their own specialism. They mostly did. They were then asked whether they had ever contacted the researchers. None had. This may of course change as they gain seniority and recognition themselves and become part of trans-national networks.

- The last finding in this short review of the research is concerned with the almost total invisibility of lesser used languages in this domain among this sample. Many of the group reported their first language as a language other than their national language, but there was virtually no reported use of these languages on the Internet.

Possible trends

In this multilingual population with the capability and motivation to use the Internet in more than one language, national languages were used more than English for both email/chat interactions and research on the WWW.

These statistics result mainly from the large number of respondents whose national language is a “big” or “major” language with a high prestige factor attached to it (for example, in the case of this survey, French, Arabic, Italian) who used English only minimally or not at all on the Internet. These groups reported less need for bilingual skills, probably because there are Internet resources available to them in their own languages. The reverse applied to those who spoke a language where speakers were either fewer in number or less favoured economically. These groups reported high levels of English.

However, there was also what appeared at first to be a slightly contradictory finding in that in the samples in which English is least used (French, Italian, Polish, Japanese), it was actually being used primarily to gather information and for academic work. In contrast, in the samples in which it is most used (Macedonian, Tanzanian, Indonesian) English did not dominate in any particular domain and was used both for both accessing the WWW and writing emails and contributing to chat rooms. On reflection this contradiction can be reconciled, since the inference is that where English language use increases, the user tends to employ the language both for passive understanding and for active communication. This has, of course implications for literacy practices in such

settings, and is the phenomenon of 'English domination' that has been noted and accepted as a general truth. The set of practices here could indeed contribute to language shift to English.

Other trends that were noticed included the increasing domestication of the IT resource. For a long period, software only supported languages which used the Roman alphabet and so disadvantaged groups which did not, even those that were technologically advanced. Now technical solutions to this are developing swiftly. However, even where these are not available, those whose bilingualism has given them competence in the Roman alphabet adapt that knowledge to other languages in which they are literate. Practices are developing which allow language with diverse alphabets to be used with Roman keyboards and software that does not support them. It is competence in English language that appears to permit these strategies. In yet another way this dents the argument that English inevitably restricts the use of other languages on the Internet.

Limited diversity and alerts for policy making

The general assertion that English dominates on the Internet needs to be reviewed. Certainly the use of English is extensive. However, it is only in a minority of groups and in a minority of situations that non-native speakers of English decide to use English rather than their national language. Trends seem to be towards diversity on the Internet away from the monopoly of English. English was a side effect of the geographical origins of the technology and the location of early users. As access opens up and participation grows in all areas of the world, the trend to a diversity of languages can only increase. English was also a side effect of the profile of early users who belonged to elite trans-national groups. As access becomes more democratic and there is penetration vertically into societies, the linguistic effect will be the same. These university students who can be seen as proto-elites demonstrate the linguistic diversification inherent in much Internet growth.

In summary, English usage on the Internet appears to coexist fairly comfortably with usage of one of the major national languages. Those bilinguals whose repertoire includes English and a prestigious national language with large numbers of speakers and an economy, which can support and promote this language on the Internet, do not seem to be

shifting massively to English in this medium. Here, even though English is sometimes employed, it seems unlikely that this will contribute to language shift away from the national language. The likelihood will diminish even further as the large language groups continue to domesticate sources to be in their own language and within their own text traditions. However, using English on the Internet may be contributing to language shift for those whose languages do not have many speakers or where the language community does not have the economic resources to fund sites in the language. The flexibility and competence of bilinguals in these settings contribute to this shift. These findings are not unsurprising and point to where effort should be directed in organisations concerned to foster diversity on the Internet.

The consequences for lesser used and endangered languages need further investigation. Diversity on the Internet seems to mean more sites for national languages, rather than full diversity. A post-national development that would see lesser used and endangered languages coexisting more easily with International English than with national languages is certainly not apparent from these data.