

BRIEFING DIGITAL IRISH



Lost for words

The Irish language is one of dozens of European tongues in danger of electronic extinction due to lack of digitised content. How can Gaeilge jump the digital divide, asks **John Burns**

BACKGROUND

Some European languages lagging



Research has been carried out into how 30 of Europe's estimated 80 languages are coping in the digital age, and the conclusion is most are unlikely to survive it. Language technology (LT) researchers estimate that digital support for 21 of the tongues is "non-existent" or "weak". The study by Meta-Net, a European network of 60 research centres in 34 countries, concluded that most of the languages (such as Basque and Catalan) are under-resourced, while some are completely neglected. Irish is among those at risk of being left behind, but "with a strong pedigree in translation, localisation and LT research, Ireland is in a strong position to develop such technologies", according to the authors, including Dr John Judge of Dublin City University.

INTERACTION

Irish-language users do have digital tools



LT covers everything from translating web pages to operating a mobile phone with spoken commands. There is software support for the Irish language, for example Google, Microsoft Windows and Mozilla Firefox all have localised interfaces. MS Office has a spell-checker, although it hasn't been updated recently, and there are online tools for spelling and grammar such as GaelSpell. "There are localised versions of word processors, internet browsers and other desktop tools as well as operating systems," said Judge. "These technologies are being used in the Irish-speaking community." However there is no automatic speech recognition in Irish, and "there is a long way to go" before such systems can be put in place, says the study.

TONGUE TWISTER

Irish has unique quirks



The Irish language has distinctive features which make speech- and text-based technologies tricky. Unlike most European languages, it has a verb-subject-object structure, and adjectives follow nouns. Irish has no words for "yes" or "no". To give a yes or no response, you must repeat the verb. There are two forms of "to be". Instead of a standard spoken form, there are three distinct dialects — Donegal, Connemara and Munster. As the standard spoken form has no native speakers — being a compromise of the three dialects — it "does not suit the building of a speech-synthesis system, which has to be based on a specific speaker". There is also an absence of resources which text-based technologies need, such as pronunciation dictionaries.

SPEAK UP

Urgent work needed to help Irish



Irish is largely out of the funding loop because it does not represent a market that investors are interested in, the report points out. Microsoft, Google and IBM all have centres for linguistics, and are based in Ireland, "but do very little work on Irish language data". Being an official EU language affords Irish some protection, according to Judge. "But we desperately need to start producing more core component tools like parsers, chunkers, speech recognisers and synthesisers," he said. "In order to do this we need the 'raw materials', ie linguistic data in a machine-usable form. There is a lack of continuity in the maintenance of what little we have, which means existing resources run the risk of becoming obsolete."

