Delyth Prys
(Language Technologies Unit, Bangor University)

"Language and Technology for Less-resourced Languages."

David Crystal famously postulated that "an endangered language will progress if its speakers can make use of electronic technology". Today there is a general consensus that all disadvantaged languages, be they classed as 'endangered', 'minority', 'less-resourced', part of the 'long tail' or just 'small', will suffer and decline unless their users can use their native languages in the new digital media. Since 'disadvantaged languages' include all languages apart from English and maybe a half a dozen other economically significant languages, this is a significant threat to language diversity.

Activities to provide tools and resources for less-resourced and other disadvantaged languages in the digital sphere include the creation of basic language aids such as electronic dictionaries and proofing tools, more sophisticated software such as machine translation, synthetic speech and speech recognition, leading to cutting edge applications such as digital personal assistants, and the whole world of Artificial Intelligence and computer/human interaction as mediated through human language(s). Although this field is usually thought of as the realm of computer scientists, coders and geeks, it also needs linguists, and it overlaps with computational linguistics, corpus linguistics, phonetics and bilingualism studies.

This paper will give and overview of the relevant research undertaken in the Language Technologies Unit at Bangor University, in the context of revitalizing the Welsh language. These activities include the creation of core resources such as bilingual electronic dictionaries, spelling and grammar checkers, language models and acoustic models, and speech and text corpora. It will give examples of our pioneering use of crowdsourcing to collect linguistic data, and our tracking of the impact of our activities on language revitalization efforts.