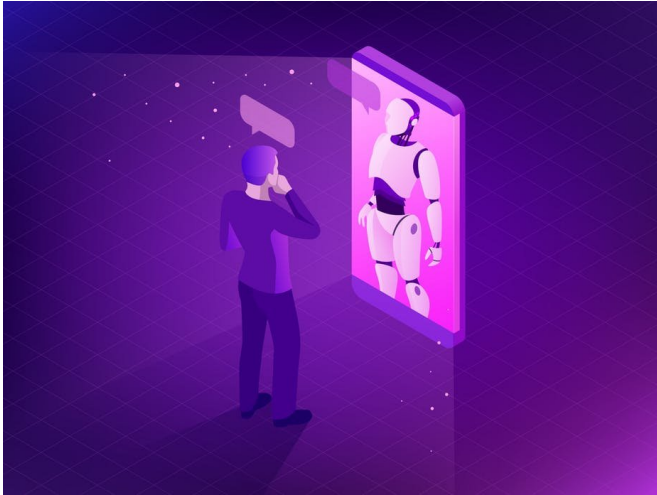


# How to teach AI to speak Welsh (and other minority languages)

29 October 2018, by Ian Mcloughlin



Ydych chi'n siarad Cymraeg? Credit: [Golden Sikorka/Shutterstock](#)

Pioneering smart home technologies and voice assistants don't, as a rule, speak Welsh – although the Welsh government now aims to change that through their [Welsh Language Technology Action Plan](#). But is their aim feasible, is it necessary, and how can it be done?

AI speech tools (like Google's Pixelbuds) are heavily reliant on the use of big data sets to learn a language, its pronunciation, grammar and semantics. The ability or quality of the resulting tools is mainly limited by how much data is available (and how "good" it is). This means that, in theory at least, tools for a minority language like Welsh cannot become as capable as those for a mainstream language.

Languages with limited amounts of good training data available are termed "low resource" languages. Compared to English, [Welsh](#) resources are sparse, but there are [several thousand languages](#) with [fewer speakers](#), and most likely

much poorer resources, than [Welsh](#).

Fortunately there is good research being done on a machine learning technique called "[transfer learning](#)". This allows systems to learn using one set of data and to then apply this knowledge in another. In China it is being used for [automatic speech recognition \(ASR\) of Tibetan](#), which has virtually no data available for training. The ASR system learned Chinese – which is linguistically very different to Tibetan – and was then retrained or finetuned to "understand" Tibetan. There is actually a lot of commonality between many languages – shared or borrowed words and pronunciation patterns – that helps this kind of technique.

## Retraining AI in Welsh

So there is no reason why AI systems cannot be produced to converse in Welsh or other minority languages. But is there any reason why it should? All of the speech [technology](#), smart homes and voice interaction systems used today are the products of commercial research. To put it bluntly, they exist to either make money from your data, to sell you more goods and services, or to influence your thinking. None of this AI [exists for the public good](#).

Making a system that works well with Welsh may not be as easy as engineering everything in English. With current technology, speech AI experts will be needed (and we are expensive). There will be a need for Welsh training and testing material, and Welsh speaking testers must be involved. The dangers of not having Welsh speakers involved in the translation process has been [amply demonstrated in the past](#), when an out of office email reply ended up on a road sign

Unless there is a strong enough economic argument, don't expect big companies to rush into producing Welsh, Gaelic or Cornish speech systems. Even tech giant Samsung hasn't yet

managed to produce a UK-English speaking version of their Bixby assistant (international English speakers need to [speak to it in fake American accents](#) to get it to work). Even the US-English version was [delayed due to a lack of resources](#).

And as long as Welsh speakers are happy to make use of English language AI systems, there may not be an economic argument – unless the Welsh government decides to pay to make it happen, which it has so far not done (the action plan is a "commitment" at this stage).

### AI to the rescue

Technology marches on and techniques such as transfer learning are becoming more capable every day. This has allowed previous research on [language adaptation](#) to be refreshed and extended into development of [multi-language deep learning techniques](#). Meanwhile, growing use of other kinds of digital technology by Welsh speakers has improved the collection of resources in the language, as has Welsh TV and radio. These advances mean that the cost of localising systems for Welsh (and other minority languages) is reducing.

Research on brain-like learning algorithms may just hold the key here. This is technology that can continually learn during use, just like humans learn to speak a new language. It is unlike most current AI systems that are trained in the lab, before being let loose in the wild – apart from a few exceptions some, like [Microsoft's Tay](#), notable for their [spectacular failures](#). Future systems will be able to gradually acquire skills in a second language just by having users gradually introduce more and more of that language in their daily interactions. Rather than funding research into Welsh speech AI, the Welsh government may well do better by backing research into [this new kind of adaptive learning technology](#).

Because all current speech AI systems handle the [speech](#) centrally (it's not done in the device, but [in a remote server farm](#)), these systems could gather data from hundreds of users worldwide (or all over Wales) to rapidly learn. So the message to Welsh speakers today may be to not buy that

English-language Google Home or Amazon Alexa if you want Google or Amazon to produce a system that works in Welsh. But if you do have one, as its software develops over the next few years, try speaking Welsh to it as much as possible. It may just surprise you and [Siaradwch â chi yn Gymraeg](#).

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

APA citation: How to teach AI to speak Welsh (and other minority languages) (2018, October 29)  
retrieved 23 October 2021 from <https://techxplore.com/news/2018-10-ai-welsh-minority-languages.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*