

# Technology and Welsh Language Conference Report 2019

*Friday, 25th January, 2019, at the Management Centre, Bangor University (9.30am to 4pm).*

The third conference that the Language Technologies Unit, Canolfan Bedwyr, has hosted for academics and practitioners interested in the use and dissemination of language technologies for minority languages, and in particular, Welsh.

The conference has been a success and has led to further talks about collaborative plans. Many of the smaller companies at the conference have asked for collaboration with the Language Technologies Unit to develop language technology.

1. Speakers
2. Software Exhibitions
3. Publicity
4. Attendees
5. Suggestions in discussions from the floor
6. Outcomes

Opened by **Eluned Morgan, Welsh Language Minister and International Relations** where she presented the Welsh Government's **Welsh Language Technology Action Plan**.

## **1. Speakers** (speakers' background can be seen in the briefing pack)

### ***Speech Technology***

- Dewi Bryn Jones: Language Technologies Unit, Canolfan Bedwyr: First Skills for Macsen, the Welsh Digital Personal Assistant
- Kelly Davis, Mozilla: Mozilla's Deep Speech Project for Speech Recognition
- Stefano Ghazzali, Language Technology Unit, Canolfan Bedwyr: The Lleisiwr project

### ***Digital strategies and language revitalization***

- Claudia Soria, Institute Computational Linguistics "A.Zampolli", Pisa, Italy: The DLDP's Digital Language Survival Kit

- Jason Evans, National Wikimedian, National Library of Wales: Welsh Wikipedia and Welsh Wikidata data

### ***Machine Translation***

- Delyth Prys in place of Teresa Lynn: Machine Translation: the experience of Europe
- Myfyr Prys, KTP Associate Cymen / Bangor University: Machine Translation for Limited Domains

2. There was also an opportunity to **exhibit Software** at the conference. This was opportunity for researchers and small companies to display their software and language technology equipment that includes Welsh. They can be finished product, prototypes, or an early proof of concept.

### **Software Exhibitions:**

- Interceptor Solutions Ltd (LinguaSkin, <http://www.linguaskin.com>)  
LinguaSkin is an innovative solution that enables online services to be multi-lingual as well as a quick creation of easy-to-use and inexpensive free language packages. LinguaSkin language packages make it possible to add language to an increasing number of web resources, including cloud / SaaS. It creates an agile interface easily without having to change a program. Based in Newport, they have worked with Newport, Swansea and Anglesey Councils, as well as the Welsh Language Commissioner, Arts Council of Wales, National Library of Wales and Coleg Cymraeg Cenedlaethol.
- Welsh Common Voice with Meddal.com  
Meddal.com has been working to encourage participants to donate recordings of their voices to Mozilla's "Common Voice" project and through this expand the speech corpus used to create Welsh speech recognition with Mozilla's Deep Speech engine. The Welsh Language Technologies Unit donated content/ sentences to Common Voice so that they could begin collecting Welsh Data immediately. The aim is to collect 100 hours of Welsh data as the technology requires a large amount of data to work effectively. There are 1,200 Welsh sentences to be read aloud by participants, so there is a constant appeal for Welsh speakers to contribute! @MeddalCom's twitter

account has often asked their followers to contribute their voices to Common Voice in Welsh, as well as an article in Golwg ('Will you contribute to Common Voice Cymraeg?', 10/10/18), item on Golwg 360 ('Appealing for voices for Welsh language technology', 08/08/18), and a host of community newspapers, magazines and organizations. Meddal.com and Language Technology Unit Members attended Mozfest (Mozilla Conference, London, 26-28 / 10/18) where they have been advertising the Common Voice in encouraging speakers of other less resource-makers to create their own version. An article on Meddal.com (01/11/18), states its hopes for the future of the project: to continue to raise awareness of the project on the website and social media pages of Meddal.com and encourage everyone to contribute.

- Astral Dynamics

Create creative solutions to technical problems. They have worked with clients in the field of health, education, data analysis, trade and travel, to create resources to solve problems that interfere with communication and information management. Some of these resources are designed to record staff working times, computer games, a visual resource for conferencing data, and the development of new applications.

- Geiriadur Prifysgol Cymru (University of Wales Dictionary of the Welsh Language) (GPC app ac GPC+)

GPC is a 'convenient way to use all the wealth of knowledge and scholarship of the Dictionary of the Welsh Dictionary' (GPC Website, Android App and iOS). The app is available on Apple, Android and Amazon Fire devices. Developed with Welsh Government financial support. The app includes all the University of Wales Dictionary's data and downloading the information to your device can enable you to work offline. GPC + offers access to new resources that were not on the original app, for example a wider context that was collected in the process of preparing the Dictionary itself. As a user of the app, you are part of the process that creates these new resources. When using the paper slips that show the words in context, they can be transcribed into the app and the user can contribute as a volunteer.

- LineageOS (Aled Powell)

LineageOS is a free open operating system. It works on a number of devices although it has been developed on an Android basis. They have created a community for

anyone with an interest, whether they are experienced Android users or not. It is possible to contribute through Gerrit (code editing system, using Git system), Wiki (as editor) or Translate (by creating versions of LineageOS in other languages using Crowdin).

- Mobilise (a company from Swansea)

Mobilise helps organizations move their services to the Cloud. They help with the planning, implementation and organization process. Hyperscale Public Cloud is used to implement this, using computer systems that respond to voice without a need for a server. This enables the process to be less expensive and to be easier and safer to use. Some clients: Welsh Government, Football Association of Wales, DVLA and BT.

### 3. Publicity

- ‘Sharing Welsh language technologies with the world’, Bangor University, 28/01/19 <https://www.bangor.ac.uk/news/latest/sharing-welsh-language-technologies-with-the-world-39523>
- Twitter (#techiaith, <https://twitter.com/search?q=techiaith&src=typd>)
  - @canolfanbedwyr, ‘Eluned Morgan @walesintheworld emphasises the importance of technology to the vision of @WelshGovernment for increasing the use of Welsh. #cymraeg2050 #techiaith’, 25/01/19 (<https://twitter.com/canolfanbedwyr/status/1088745937298747392>)
  - @prifysgolbangor, ‘Rhannu technolegau iaith y Gymraeg â gweddill y byd yng Nghynhadledd Technoleg a'r Gymraeg 2019. Stori lawn yma: <https://bit.ly/2RizHU1> @canolfanbedwyr @techiaith @walesintheworld @mozilla’, 18/01/19 (<https://twitter.com/prifysgolbangor/status/1089844028189933568>)
  - @walesintheworld, ‘Pleser i agor cynhadledd dechnoleg @prifysgolbangor eleni. Wrth i dechnoleg barhau i drawsnewid ein bywydau, mae hwn yn gyfle gwych i rannu syniadau a dysgu oddi wrth ein gilydd er lles pob iaith. Mae technoleg yn symud yn gyflym ac mae angen i'r Gymraeg gadw lan’, 25/01/19 (<https://twitter.com/walesintheworld/status/1088816828485435392>)

- Other Twitter accounts that tweeted about event: @carwynedwards, @MeddalCom, @illtud, @melynmelyn, @digitalst, @stefanoghazzali, @WIKI\_NLW, @WICI\_LLGC, @melindrafod, @sarahcooper\_1, @DeianDCW, @IaithAmgueddfa, @dewibryn, @yrawrgymraeg, @cynog\_prys, @pjstead, @IndegElen

#### **4. Attendees according to sector**

- Local Authorities: 6
- Software Companies: 8
- Public Bodies: 30
- Private Companies: 25
- Universities: 19

Total = 88

In addition a number of other friends dropped in for one or more sessions without having registered formally, therefore we do not have firm statistics for their participation.

#### **5. Suggestions in discussions from the floor**

Looking to the future, some participants made the following points:

- There is an opportunity for Welsh to collaborate and plan language technologies so that they link in with solutions offered by Mozilla and others (which are open-source)
- It is important to take advantage of the DLDP / SIGUL networks to promote collaborations with European contacts
- Here in Wales, attention needs to be given to the ecosystem of language technology - especially to strengthen training in the field (e.g. Master's degree in language technology) and knowledge exchange ('transfer learning') to SMEs
- A new field of study is Welsh / English bilingual applications (rather than monolingual ones) should be a focus for the Unit, especially the ability to swap between the two languages easily
- The LTU could talk to the National Library of Wales to look again at Wikidata as a source of appropriate copyrighted data and explore potential collaboration opportunities
- The LTU could improve the packaging and help to industry of its open source resources and as a minimum develop:

- Effective speech recognition models
- Better artificial voices
- Audio base for 'chatbots' and 'intent parsing

## 6. Outcomes

- Dewi Bryn Jones (UTI) worked with Josh Meyer (Mozilla) to train DeepSpeech with the Common Voice corpus. They are improving the Common Voice corpus and 'tidying up' its language content in order to be more effective. (Debate discussion / results @dewibryn, <https://twitter.com/dewibryn/status/1088488287373799425>, <https://twitter.com/dewibryn/status/1088559247653261312>)
- Discussion about expanding engineering learning technology to games between on Twitter between @carwynedwards and @WeAreGameDev (<https://twitter.com/carwynedwards/status/1088812624723300353>).
- Talk about possible collaboration between Huw Marshall (@Marshallmedia) and Mozilla (in the question session after the Kelly Davis presentation).
- Discussions on copyright data from the BBC and S4C following the presentation of Dewi Bryn Jones. Also Supercomputing Wales was mentioned in discussing the computer power needed to train a speech recognition system (a day with the help of Supercomputing Wales, about a week on a PC). Dewi and Josh (Meyer) have been training DeepSpeech that took 5 hours
- Delyth Prys looked forward at preparing the field for the next 20 years and establishing a new generation of language technologists. Continue to train, to provide research and to create the technology (all of which are on an open license). Delyth also talked about the need to work with Ireland and learn from major international projects to create our own. It was noted that the use of these resources needs to be extended to other fields, not only machine translation. As well as Ireland, the other Celtic languages should be helped through 'transfer learning'.
- People were encouraged to contribute their Welsh voices to Common Voice in response to the software demonstration. @techiaith, 'More contributors needed! Contribute to your voice to Common Voice. People of all ages and all corners of Wales are needed. ' 25/01/19 (<https://twitter.com/techiaith/status/1088811966561435656>)