

CorCenCC Newsletter

Issue 4: July 2016



Corpws Cenedlaethol Cymraeg Cyfoes
National Corpus of Contemporary Welsh

Greetings from the PI



Welcome to the 4th edition of the CorCenCC newsletter! Time on the project seems to be flying and as the summer months are driving ahead, so too are developments on the planned work. The current issue documents some of the activities from June (and the very start of July), which have mainly included conferences and a range of public events – where we have been spreading the word about the vision for the project, and collecting the names of potential contributors to the corpus. We will be heading to the National Eisteddfod in August too (details below) – so do pop in and say hello if you are there! There are also conferences/invited talks in Tokyo and Slovenia for the team to look forward to – but more about that in a later edition.

The rest of the summer includes some exciting work on the final developments of the CorCenCC crowdsourcing app that Research Associate Dr Steven Neale is busy developing. We are hoping to launch this in the next few months, then it will be full steam ahead with data collection (we will be calling on all Welsh speakers to get involved with that ☺). We are also planning to launch the official CorCenCC project site soon (www.corcenc.org and www.corcenc.cymru). Keep an eye open for both of these!

Back to this month's newsletter – this edition brings you recent project news (from June), write-ups from some of the events that members of the project team have attended, and we also take the opportunity to introduce you to another member of the CorCenCC team, Dr Gareth Watkins.

Happy reading, Dr Dawn Knight (Cardiff University)

News

We want to take this opportunity to say a big welcome to the newest member of the CorCenCC team, Professor Colin Williams. Professor Colin Williams is a Fellow at St Edmund's College, University of Cambridge. In addition, he is also an Honorary Professor at the School of Welsh, Cardiff University. He was a member of the Welsh Language Board during the period 2000-2011. He is an international specialist in language policy and language planning. Colin joins the team as a member of the Project Advisory Group. Welcome, Colin!



Conferences and events

1. IVACS 2016

The biennial IVACS (Inter-varietal Applied Corpus Studies) conference was held on 16th and 17th June at the beautiful campus of Bath Spa University. This was an opportunity for members of the CorCenCC team to spread the word about our progress and our plans for the corpus, to familiarise ourselves with current research in the field of Corpus Linguistics, and to get to know other researchers ... under the influence of good food at the conference dinner at the Bath Pump Rooms!



all glad to be working on European languages when we heard, in another presentation, about the additional challenges involved in developing corpora of oriental languages.

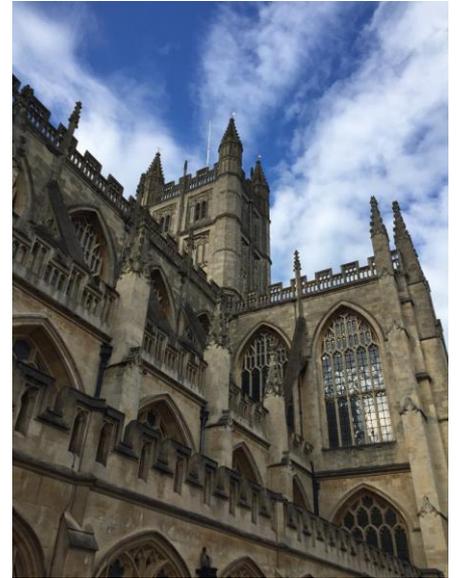
Steve Neale and Gareth Watkins (above) gave a paper about one of the most innovative aspects of the CorCenCC project: namely, the use of crowdsourcing for corpus construction. Through a bespoke app, members



of the public will be able to record their Welsh conversations and upload them to the corpus together with relevant information such as the participants' ages and where they live. The app will enable Welsh speakers

from all corners of Wales to contribute to CorCenCC easily.

Another interesting paper, entitled 'Corpus-informed discourse analysis: a methodology for exploring context in spoken corpora', was presented by CorCenCC principle investigator, Dawn Knight, together with a colleague, Michael Handford. We also benefited from thought-provoking presentations about the use of corpora for language learning, which will be really useful as we design CorCenCC's pedagogical toolkit (WP4). Thanks for the inspiration!



Mair Rees and Jennifer Needs

Contributed by Jennifer Needs

2. Tafwyl

In July, The CorCenCC team went to the Tafwyl festival – Cardiff city's annual Welsh language festival. Although a number of the visitors there were still suffering following the national team's victory against Belgium in Euro 2016 the previous night, a good crowd came to the Literature Wales tent to hear Steve Morris and Mair Rees talk about the project, the work that has been achieved so far and how they could contribute to CorCenCC themselves. This was one of the main aims in attending Tafwyl, namely trying to recruit people to contribute to the corpus.



Steve Morris and Mair Rees



Dawn Knight and Gareth Watkins joined in the work – and we were fortunate that one of our ambassadors Nia Parry was there to spread the word too. At the end of the day, we had succeeded in getting a good list of names of possible contributors.

The intention is to do something similar at the National Eisteddfod in Abergavenny in August thereby ensuring that we have a good list of potential participants by September.

Contributed by Steve Morris



Gareth Watkins, Dawn Knight, Steve Morris and Nia Parry

3. Other events: August 2016

➤ National Eisteddfod (29th July to 6th August, Abergavenny)

Members of the CorCenCC team will attend the National Eisteddfod to inform the public about the main aims of the project, and provide details of how they can get involved in the work. Informal presentations on the project will take place in the Cardiff University tent at 11am on 2nd August and in the Swansea University tent at 11am on 3rd August (but you will find us on site throughout both days). For more details, see: <https://eisteddfod.wales/>

Meet the team

Every month we will be featuring a different member of the extended CorCenCC team in our newsletter. This will give everyone a chance to tell you a little about their background; what they want to see from CorCenCC and how they think it might contribute to their own work, or more broadly, to the work of others in Wales. This month we spotlight Gareth Watkins, who is a Research Assistant (RA) at Cardiff University.

Profile: Dr Gareth Watkins



That's me on the left

As I was appealing to my muse for inspiration to write this article, as I let my mind wander, I was reminded of a language technology conference that I attended in Barcelona a while back. At the conference I presented research on how Welsh translators use translation memory systems, such systems being a kind of technology designed to help translators as they translate.

That's one of my research interests, my specialty, in fact. The subject in which I completed my doctorate a couple of years ago.

But back to the conference. While chatting over coffee, I happened to mention to one of my fellow delegates that I do not consider myself a translator, per se. My fellow delegate was both shocked and appalled (in a manner of speaking). How on earth is it possible to be an expert on translation technology without having a background as a translator? Well, I explained glibly, easily enough! After all, there are people in this old world who are shark experts, even though they have never attempted to bite a surfer and swallow them whole. This was a reference to a Welsh pop song, possibly

inappropriate at the time considering my fellow delegate had no experience of Welsh culture. It loses something in translation!

If not a translator, what then? An academic? Yes. A researcher? Certainly. A person interested in the Welsh language, and her relationship with technology? Two more big tick marks.

What has this to do with the CorCenCC project? For me, personally, the project is an opportunity to combine three of my main interests, that is research, the Welsh language, and technology. However, I am only one small Welsh speaker among over half a million Welsh speakers. What about the others? Well, in general, it can be stated with certainty that the CorCenCC project has the potential to touch all of them, whether young or old, first language or second language, a Northerner or a Southerner, and everyone else in between. CorCenCC is a project which Welsh speakers can take ownership of.



More specifically (and thus back to my area of expertise), such a resource will contribute to the field of language technology, and as a result, contributes to the livelihood and work of our translators. Perhaps it is not surprising, at least according to my research, that a very high percentage of Welsh translators do not use corpora as they translate. Not surprising by a long stretch. After all, comprehensive general corpora on the same scale as CorCenCC have yet to be created. Not surprising, but a shame nonetheless. Translators would find a corpus a very useful tool indeed.

After all, what better tool than a corpus to support a translator's natural talent as they strive to create a natural translation? What better than a corpus to support them as they strive to create a text that is not identifiable as a translation? What better than a corpus - a large reservoir of natural language, a bank of language as she is used by her speakers? Nothing that I'm aware of.

That's me on the right

CorCenCC online

You can keep up to date with developments on the project via Facebook www.facebook.com/CorCenCC/; Twitter <https://twitter.com/corcencc> (Tweet us @CorCenCC). You can also contact us on the project email address: corcencc@cardiff.ac.uk or visit our holding website at: <http://sites.cardiff.ac.uk/corcencc/>

CorCenCC is an ESRC/AHRC funded research project (Grant Number ES/M011348/1). The CorCenCC team includes **PI** - Dawn Knight; **CIs** - Tess Fitzpatrick, Steve Morris, Irena Spasic, Paul Rayson, Enlli Thomas, Mark Stonelake and Jeremy Evas; **RAs** - Steven Neale, Jennifer Needs, Mair Rees, Scott Piao and Gareth Watkins; **Consultants** - Kevin Donnelly, Kevin Scannell, Laurence Anthony, Tom Cobb, Michael McCarthy and Margaret Deuchar; **Project Advisory Group** - Karen Corrigan, Llion Jones, Maggie Tallerman, Mair Parry-Jones, Gwen Awbery, Emyr Davies (CBAC-WJEC), Gareth Morlais (Welsh Government), Owain Roberts (National Library of Wales), Aran Jones (Saysomethingin.com) and Andrew Hawke (University of Wales Dictionary of the Welsh Language).

If you have any comments or questions about the content of this newsletter please contact Dr Dawn Knight: KnightD5@cardiff.ac.uk

