

Greetings from the PI

Welcome to the 18th edition of the CorCenCC project newsletter. Schools are out and the sunshine is still, miraculously, blazing!



We are in to the final 15 months of funded research and progress on the project is heating up too...data collection levels are ever-growing; final edits and test runs of our

new crowdsourcing tools are being carried out (look out for details on our website and social media accounts); and our expansive team of transcribers are helping us to process the spoken data that will eventually be integrated into the corpus. In this edition of the newsletter, we provide an update on recent data collection exploits, a report on our trip to IVACS 2018 in Malta, and introduce you to new and existing members of the CorCenCC team.

Happy Reading – Dr Dawn Knight

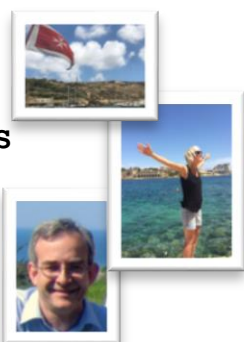
Contents

P1: News and events

P4: WP1 update

P5: Meet the team

P6: Contact us



+ News and events

IVACS Conference (13-15/06/18)

Following Jenny Needs and Mair Rees's presentation at the IVACS Symposium in Maynooth, Ireland earlier this year, the connection with IVACS [*Inter-Varietal Applied Corpus Studies*] continues! Dawn Knight and Steve Morris went to the 9th IVACS Conference held at the University of Malta (L'Università ta' Malta), Valletta Campus in June to share more information about CorCenCC. Dawn presented a paper on '*Corpus design and construction: the challenges faced by minoritised languages*' to a room overflowing with delegates on the second day of the conference and on the final day, Steve gave a presentation entitled '*CorCenCC: applying the sociolinguistics of new speakers within a contemporary corpus of Welsh*'. The striking thing about being at the IVACS conference is the emphasis on *application* in the context of work

on corpora. In this respect, the conference didn't disappoint at all and there was an obvious interest in CorCenCC – how we were going to collect data in a country where the demographics of the speakers is different to some of the larger linguistic communities, the response of some of our contributors (e.g. '*my Welsh isn't good enough to give to you*') and problems in ensuring that every variety of the language is represented in the corpus. It became obvious when networking with our fellow conference attendees working with the Irish, Flemish and – even though it is one of the country's official languages – Maltese languages that we face many of the same challenges and that there is a great potential for us to learn from each other as well as work with each other to facilitate corpus work in the context of the world's minoritised languages. This is an element of the project which we intend to develop – read the Newsletter for further details as they happen.



Welcome to the team: Alex Lovell, Swansea University

In the January 2018 edition of the newsletter, we said a sad farewell to Mark Stonelake who retired from his position at Swansea University and left the CorCenCC team. This month we are pleased to announce that we have recruited a new member of the team, stepping into Mark's shoes. We would like everyone to say a big welcome to Alex Lovell, a lecturer in the Welsh Department at Swansea University. Alex is currently finishing his PhD in which he looks at how best Welsh as a second language can be successfully presented in comparatively non-Welsh speaking areas of Wales. We asked Alex to introduce himself to you all...

"Shwmae! My name is Alex Lovell and I am a recently appointed lecturer in the Welsh Department at Swansea University. I am originally from Caldicot, Monmouthshire, but I am hoping to move to the Bridgend area in the coming months! After enjoying studying Welsh in school, I decided to study a degree in Welsh at Swansea University in 2010. As someone who learned Welsh as a second language, I was determined to gain the skills necessary in order to be able to speak and write fluently in Welsh, before moving on to a career as a Welsh language teacher in a secondary school. However, by the end of my degree I began to take interest in research. Four years later and I'm still at the university!

In terms of research, I am currently finishing my PhD in which I look at how best Welsh as a second language can be



successfully presented in comparatively non-Welsh speaking areas of Wales. In addition to Welsh Second Language education, I have a strong interest in research fields such as language planning, second language acquisition, language and education policy, bilingualism and bilingual education. As a lecturer in the Welsh Department, I am primarily responsible for teaching a number of language modules for second language students, but I also convene and contribute to a number of other modules in the fields of language, applied linguistics and language planning. I am extremely pleased to be a part of such an important and innovative project and I very much look forward to working with the team!"

Welcome to the team: Cardiff University CUROP students

In the March edition of the newsletter we brought you news of our success in obtaining funding from the Cardiff Undergraduate Research Opportunities Programme (CUROP), for two student placements over the summer months. We are pleased to announce that two great students have been recruited to these posts. Allow us to introduce Alys and Sali...

Alys Greene, CUROP student

Hello! My name is Alys, and I have just begun my work on my CorCenCC Summer placement. Originally, I come from a small village in the Gwendraeth Valley, but I am currently living in Cardiff studying Law and Welsh, about to enter into my third and final year. As someone who has been raised in a Welsh speaking family and received my entire education through the medium of Welsh, I can really appreciate the significance of the CorCenCC project's work. The position appealed to me because of the focus on the collection of data of the Welsh language as it's used naturally from day to day, and I strongly believe that the used, spoken language is as important as the way it's used in any formal context. My role in particular within the CorCenCC project will be concentrated on WP3, that is the semantic tagging of the data collected but I look forward to helping in any way that I can with building this corpus of contemporary Welsh.



Alys Greene, Cardiff University

Sali Nichols, CUROP student

It is ironic that living in Cardiff, the capital city, the Welsh word I am asked to recite most often is *Llanfairpwllgwyngyllgogerychwyrndrobwlilllantysiliogogoch*, with the odd *popty ping* and maybe a *chwyrlligwgan* here and there. In Swansea, my home city, it is normal for even non-welsh speakers to use *bola* and *cwtch*, therefore the amount of shock people express when they first hear I am a native Welsh speaker still baffles me; how do people live in the capital city of Wales and not realise we have our own language? “Welsh is a real language?” and “I didn’t realise people still spoke that, I thought it had died!” are sentences I’m sure every Welsh speaker has heard at some point in their life, however they seem more prominent in Cardiff than anywhere else; even in the proudly patriotic ‘Clwb Ifor Bach’ you get the odd look of bewilderment as you converse with the bartenders in Welsh.

My entire academic life has been through the medium of Welsh, and as a result, I was brought up bilingual by hilariously patriotic parents; it is not unusual to see the national anthem sprawled on a piece of wood or to see the flag draped proudly on the wall of my house – we are even partial to an inflatable daffodil every now and again!

Growing up in a small post-mining village, I was teased mercilessly for having the same name as Sali Mali (the correct way of spelling Sali might I add) and was constantly asked how Jac-y-Jwc was doing in my little ‘caffi’. However looking back as an adult, I have come to realise that the unique way both Sali Mali and I spell our names is what makes us us; our little land of poets and singers has a mind-blowing history and a language beautiful enough to bring a tear to the eye.

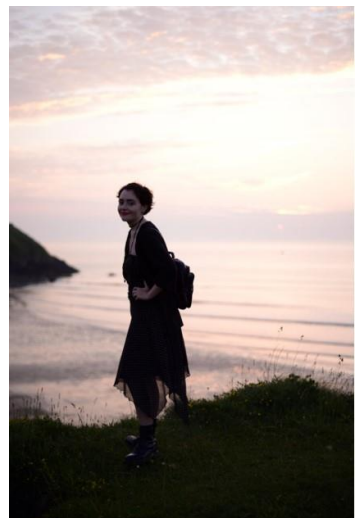
Languages have been my passion for a while and during sixth form (wherein I studied Welsh, English Literature and French) I was surprisingly struck with the realisation that I love these subjects, these crazy grammatical rules and history that we take for granted every day. One of my favourite things that we studied was the play ‘Siwan’ by Saunders Lewis. Lewis implicitly amalgamated the intricate history of Llywelyn Fawr and his relations with his unparalleled style of writing that I will never tire of reading. It is without doubt that my bilingual upbringing has deeply rooted my love for languages and has paved my path in the world of linguistics; without the Welsh language I would not be who I am now and most probably would NOT be reading French and Italian at University.

While my friends and I sit in a beautifully rustic outside bar, we joke


that I am ‘spiritually French’ as I sip on my glass of Semillon Chardonnay attempting to keep my vibrant red lipstick in place. However, beneath the seemingly European façade, a choir singing and poetry reciting girl is itching to get out. As a child, every summer was spent frantically travelling to the next Eisteddfod, making sure not to get in trouble for having an out of place tie and messy hair.

Above all, languages, especially Welsh, are my life and the CorCenCC project is the perfect opportunity for me to broaden my knowledge of the Welsh language and how it differs around the country. Without doubt, in a digital and time obsessed world, one almost forgets how to communicate with finesse and skill, which is why the beauty of language and literature are becoming more precious than ever. Just like the dodo, if a language stays stagnant it will cease to function and become extinct. In order to survive they must evolve whilst maintaining their historical roots, and it is of the utmost importance that we remember the history of our native tongue.

Sali Nichols,
Cardiff
University
CUROP
student



WP1 update: an insight into data collection by Lowri Williams

The summer season is in full swing and so is our data collection! Members of the WP1 package have been travelling far and across Wales, collecting spoken data from different contexts and documenting it all by taking some lovely photos as seen here (provided by lowriawilliams )

There are several public events and festivals in Wales over the summer. We kicked off our data collection season with a bang, and in May, Jenny, Mair, and I attended the Urdd Eisteddfod which was held on the Royal Welsh showground in Builth Wells. With an estimated annual attendance of 90,000, this event attracts a diverse audience of ages, as well as Welsh language capabilities. The weather was great and it was nice to get out of the office for a couple of days. But it was especially nice to hear that many people wanted to contribute to the project and that several of those contributors were adult Welsh learners.

With just over a year remaining on the project, this was the last opportunity to collect data at the event. Our sampling frame, which guides us in collecting a balanced collection of language from a range of different contexts, types, and contributors, includes social conversations from the Urdd Eisteddfod. Following our attendance, we are pleased that we have collect almost double our target.

Following the Urdd, Jenny and I travelled up to North Wales where we spent a couple of days in Denbigh, Ruthin, Mold, and Wrexham. We were successful in recording in several local shops, cafés, pubs, libraries, and offices. I also had the opportunity to take one of my favourite pictures so far; a photo of the Welsh flag standing proud against the wind and grey sky from the inside of Denbigh Castle. We hope to return to these areas in the future to collect more data.



Our next event saw us return to the castle grounds of the capital city for Tafwyl. With an estimated annual attendance of 38,000, Tafwyl also attracts a diverse audience of Welsh speakers. The hot June sun was beaming, allowing visitors to enjoy the live music, foods, drinks, and stalls. Jenny and I met new contributors and introduced them to our crowdsourcing app. We also caught up with a few who have already contributed data and follow the project. Knowing that we have the continual support of the Welsh speaking community is a strong motivation for us during the last leg of the project.

The most recent stop was home, the Llŷn Peninsula. Historically, the peninsula was travelled by pilgrims who were en route to Bardsey Island located close to the very tip of the main land, where over 20,000 saints are claimed to have been buried. Its relative remoteness has not only helped to conserve the Welsh language, but also the unspoilt image that this area is, as J. Glyn Davies composed, a place for the soul to find peace (*"Lle i enaid gael llonydd"*).

Having being born and raised in Llŷn, one may say that I went on my own data collection pilgrimage! It was nice to return to speak with friends, family, and members of the community who feel passionate about the Welsh language. The peninsula is a popular destination for both tourists and holiday home owners. Holiday homes remain a bone of contention among locals, who strive to keep the language alive by ensuring that their correspondence in several contexts is through the Welsh medium, contributing to Gwynedd's large Welsh speaking population. We have visited some areas in Gwynedd in the past and it is very interesting to hear the differences in the language when there is only a couple of miles which separate them. From Caernarfon's "ia" to the Llŷn and Meirionydd "de", we have collected data from several locations in the county. During this visit, I was given the opportunity to record at the local primary school that I myself attended. I could not help but feel sentimental returning to "where it all began". Not only because I graduated in July and all of my memories as a child came rushing back, but also to have the school being so helpful in their contribution to the project. It is safe to say that contributors of all ages are behind the project and the Welsh language. Our next stop will be the National Eisteddfod in August, which will be held in Cardiff Bay. Not only will we be collecting data, but we are also presenting the project and some of the current outputs at Cardiff University's tent. Come and learn more at 12 noon on Wednesday 8th, and please do come and speak to us if you see us around the 'maes' too!

Dr Lowri Williams, Cardiff University



+ Meet the team: Andrew Hawke, Project Advisory Group (PAG) member, Dictionary of the Welsh Language

My interest in language and computing began at school in the 1970s. I was aware of my father's Cornish ancestry and became interested in the Cornish language, whose revival was then beginning to gain momentum. I was also introduced to computing as part of a General Studies course in the sixth form, using manual card punches and the mainframe at the University of London Computing Centre at Imperial College. (Rwy'n dal i gofio mai'r cod am 'A' yw'r tyllau 0, 6, a 5, wedi eu pwnsio yn yr un golofn.) We would write our programs (in Fortran 66) and data (in EBCDIC format), laboriously punch the cards by hand, post them to Imperial, and await the return of the cards some days later together with some lineprinter output with the inevitable error messages! And so the cycle would begin again. This became strangely addictive, and I began searching for material to process. Eventually I acquired an 1859 edition of some of the Middle Cornish plays, and typed some 9,000 punched cards (this time using a modern card-punch



machine kindly donated to the school by NCR who had an office in the same town).

I decided to read Celtic Studies at university and applied to Aberystwyth, where my interest in computing continued. I had read about computer generated concordances, and some software was available in Fortran which went by the name of COCOA. I was soon producing wordlists and concordances using the remote facilities at the University of Manchester Regional Computing Centre (UMRCC), still using punched cards, magnetic tapes, and lineprinter output. Computing was advancing very rapidly at that time, and soon hard discs, teleprinters, and primitive VDUs (visual display units) were available, and the local mainframe hardware was upgraded so as to obviate using the UMRCC facilities, much to the relief of the long-suffering IT staff at Aberystwyth who occasionally returned from meetings in Manchester with boxes of lineprinter output for me. By the early 1980s, I was doing doctoral research towards a historical dictionary of Cornish, and persuaded the Welsh Department to provide a small room in the Old College for me, and the Computer Unit to run a cable down the hill from the computer room to a new VDU. This operated at a speed of 300 baud (i.e. 300 bits[*sic*] per

second) up to a maximum (in dry weather!) of 1200 baud, quite sufficient for typing and editing texts and code. At around this time the first 'proper' business microcomputers became available, and luckily the Aberystwyth Computer Unit was an early adopter of the technology and I started examining the possibilities of using it for text analysis.

I joined the staff of the University of Wales Dictionary of the Welsh Language in 1983, without finishing my research. The opportunity to get a job on the Dictionary was too tempting, and I soon began to think about computerizing the work. By 1987 we started to typeset the dictionary ourselves, so the text was available electronically. A 'retrospective conversion' scheme was started to turn the text of the first two volumes of the Dictionary into digital format – equivalent in length to six Bibles! The conversion was completed in 2000, and the Dictionary itself in 2002, and then we went back to the beginning of the alphabet to re-edit the earliest letters. It became clear that the Dictionary should be online, and thanks to a major investment from the University of Wales and the help of a software company called EMP from

Copenhagen, all the data was converted to XML and Unicode, and launched as GPC Online in 2014 (<http://gpc.wales>). Shortly after we were awarded a grant from the Welsh Government to create iOS and Android apps, and GPC is now available in full as an app, the first complete historical dictionary to appear as an app, I believe. (Search for 'GPC' in Google Play or the Apple Appstore.)

Electronic data is of enormous importance to us as we can easily search it to find raw data for the Dictionary. Over the years we have collected hundreds of millions of words of text from various periods. Naturally, the CorCenCC project is of great interest to us as a source of contemporary Welsh, and I was very happy to be invited to serve on the Advisory Group. For our part, the Dictionary has contributed some basic data from the Dictionary itself to improve the part-of-speech tagger and also to enrich the Welsh Wordnet project. It's great to see everyone pulling together so effectively, and the project leaders deserve credit for all their careful planning. We look forward to seeing the first fruits of the project.

+ Contact us

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 website at: www.corcencc.org



Arts & Humanities
Research Council

CorCenCC is an ESRC/AHRC funded research project (Grant Number ES/M011348/1). The CorCenCC team includes **PI** - Dawn Knight; **CI**s - Tess Fitzpatrick, Steve Morris, Irena Spasić, Paul Rayson, Enlli Thomas, Alex Lovell and Jonathan Morris; **RA**s - Steven Neale, Jennifer Needs, Mair Rees, Scott Piao and Lowri Williams; the **PhD students** - Vigneshwaran Muralidaran and Bethan Tovey; **Consultants** - Kevin Donnelly, Kevin Scannell, Laurence Anthony, Tom Cobb, Michael McCarthy and Margaret Deuchar; **Project Advisory Group** - Colin Williams, 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