



### Cyfarchion gan y Prif Ymchwilydd

*Mae'r haf wedi cyrraedd yn swyddogol ac mae tymor y cynadleddau wedi cychwyn. Ar ôl ychydig o amser o ymgyfarwyddo â'n rolau ac ehangu ar gynlluniau ar gyfer y prosiect, rydym yn dechrau ymdeithio a byddwn yn mynychu ystod o wahanol gynadleddau a digwyddiadau dros yr haf. Digwyddodd yr un cyntaf ar ddiwedd Mai, pan fynychodd Paul, Scott (o WP3) a minnau gynhadledd LREC (Language Resources and Evaluation Conference) 2016 yn Portorož. Yr ail un oedd seminar gwadd academiaidd yng Nghaerhirfryn ar 9<sup>fed</sup> Mehefin. Mae cylchlythyr y mis hwn yn cynnwys adroddiadau byrion o'r digwyddiadau hyn, yn ogystal â manylion ynglŷn â ble y gallwch ddisgwyl gweld aelodau o dîm CorCenCC dros yr haf. Yn rhifyn y mis hwn byddwn hefyd yn cyflwyno aelod arall o dîm CorCenCC i chi - Aran Jones, Prif Weithredwr SaySomethingin.com ac aelod o Grŵp Ymgynghorol y Prosiect.*

*Mwynhewch! Dr Dawn Knight (Prifysgol Caerdydd)*

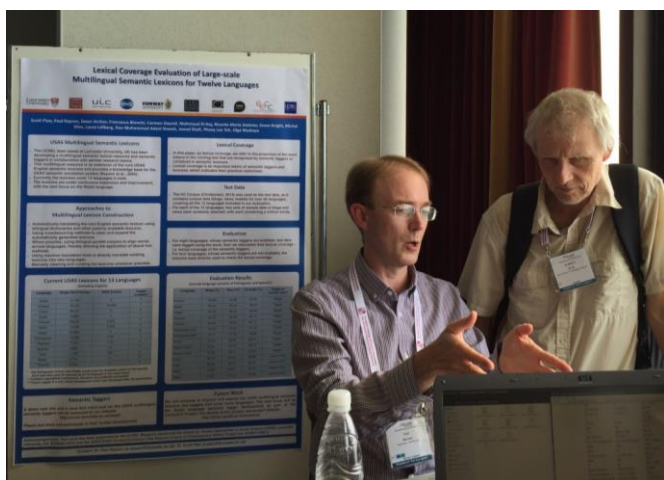
## Cynadleddau a digwyddiadau

### 1. LREC 2016

Cynhaliwyd 10<sup>fed</sup> gynhadledd ddwyflynyddol LREC (Language Resources and Evaluation Conference) rhwng 23<sup>ain</sup> a 27<sup>ain</sup> Mai yn Portorož, Slofenia. Roedd yr arfordir Adriatig-Ganoldirol yn gefnlen brydferth i 5 niwrnod o weithdai ysgogol, posterï a phapurau ar amrywiaeth o wahanol themâu; o ddadansoddi teimladau ac adnabod emosiynau a chorpora ar gyfer dadansoddi iaith, i ddysgu iaith drwy gymorth cyfrifiaduwr a methodolegau gwerthuso.



*Scott Piao, Mahmoud El-Haj, Jawad Shafi, Michal Křen, Dawn Knight a Paul Rayson*



Roedd Paul Rayson, Scott Piao a Dawn Knight yn gyd-awduron y cyflwyniad poster dwy awr hwn o'r enw 'Lexical Coverage Evaluation of Large-scale Multilingual Semantic Lexicons for Twelve Languages'. Yn y cyflwyniad poster, adroddwyd ar adeiladu lecsiconau semantig amlieithog ar raddfa fawr ar gyfer deuddeg iaith, sy'n defnyddio tacsonomi unedig Prifysgol Caerhirfryn ac sy'n darparu sail gwybodaeth geirfaol amlieithog ar gyfer USAS (system anodi semantig awtomatig UCREL).

Cymerodd tri awdur arall, Mahmoud El-Haj, Michal Křen a Jawad Shafi, ran yn y cyflwyniad hefyd, gan amlinellu eu gwaith ar lecsiconau semantig Arabeg, Tsieceg ac Wrdd yn eu tro.



Datblygu rhan iaith Gymraeg y system anodi yw canolbwynt Pecyn Gwaith 3 (WP3) ar brosiect CorCenCC. Ceir copi o'r papur cyfrwng Saesneg cyhoeddiedig yma:

[www.lrec-conf.org/proceedings/lrec2016/summaries/257.html](http://www.lrec-conf.org/proceedings/lrec2016/summaries/257.html)

## 2. Cyfres seminarau UCREL

Ar 9<sup>fed</sup> Mehefin teithiodd prif ymchwilydd y prosiect, Dawn Knight, i fyny i Brifysgol Caerhirfryn er mwyn rhoi papur fel rhan o gyfres seminarau academaidd UCREL (University Centre for Computer Corpus Research on Language). Rhoddodd yr anerchiad gwadd hwn – o'r enw 'A community driven approach to linguistic corpus construction' – arolwg manwl o gynlluniau prosiect CorCenCC. Yn y papur, trafodwyd rhai o'r materion ymarferol, methodolegol a thechnegol sy'n gysylltiedig â dylunio ac adeiladu CorCenCC a gofynnwyd am adborth ar amlinelliad drafft y fframwaith samplu ar gyfer y corpws a fydd yn ymddangos fel rhan o'r trafodaethau hyn. Amlinellwyd hefyd rai o'r strategaethau a thechnegau a gaiff eu defnyddio wrth i ni geisio denu a recriwtio cyfranogwyr i gyfrannu data at y corpws, o ledled Cymru.

Roedd y cyflwyniad yn un o dri anerchiad a roddwyd yn ystod y seminar hon, a chafodd ei ragflaenu gan anerchiadau ar elfennau llafar ac ysgrifenedig y BNC14 (British National Corpus 2014) sydd wrthi'n cael ei ddatblygu gan Brifysgol Caerhirfryn, mewn cydweithrediad â Gwasg Prifysgol Caergrawnt. Cafodd Dawn gyfle i siarad â chyflwynwyr yr anerchiadau hyn, Robbie Love ac Abi Hawtin, i rannu arfer gorau ar gynllunio ac adeiladu corpora – taith defnyddiol y tu hwnt (diolch, Gaerhirfryn!).



*Dawn Knight*

## 3. Digwyddiadau eraill: Mehefin i Awst 2016

- **IVACS - Inter-varietal Applied Corpus Studies conference 2016 (16-17eg Mehefin, Prifysgol Sba Caerfaddon)**

Bydd tîm CorCenCC yn cyflwyno dau bapur yng nghynhadledd ddwyflynyddol IVACS – mae manylion y rhain yn dilyn (edrychwch allan am adroddiad ar y gynhadledd yn rhifyn nesaf y cylchlythyr hwn):

- **Knight, D., Neale, S., Watkins, G., Spasic, I., Morris, S. a Fitzpatrick, T.** Crowdsourcing corpus construction: contextualizing plans for CorCenCC (Corpws Cenedlaethol Cymraeg Cyfoes - The National Corpus of Contemporary Welsh).
- **Needs, J., Rees, M., Morris, S., Knight, D. a Fitzpatrick, T.** CorCenCC (Corpws Cenedlaethol Cymraeg Cyfoes – The National Corpus of Contemporary Welsh): Challenges and applications in a minoritised language context.

- **Tafwyl (2<sup>il</sup>-3<sup>ydd</sup> Gorffennaf, Castell Caerdydd)**

Bydd aelodau o dîm CorCenCC yn mynychu gŵyl Tafwyl, yn dosbarthu taflenni am y prosiect ac yn recriwtio darpar gyfranogwyr i'r corpws. Yn ystod y penwythnos, byddan nhw hefyd yn rhoi cwpl o gyflwyniadau byr er mwyn hysbysu aelodau o'r cyhoedd am brif nodau'r prosiect, ac er mwyn manylu ar sut y gallan nhw fynd ynghlwm yn y gwaith. Cynhelir y cyflwyniad cyntaf ym mhabell Llenyddiaeth Cymru ar Ddydd Sadwrn 2<sup>il</sup> Gorffennaf o 13.00-13.45, a chaiff ei gyflwyno gan Steve Morris a Mair Rees. Cynhelir yr ail anerchiad yn yurt Prifysgol Caerdydd – dyddiad/amser/cyflwynydd i'w cadarnhau. Gweler: <http://www.tafwyl.org/>

➤ **Cynhadledd Cymraeg i Oedolion (8<sup>fed</sup> Gorffennaf, Caerdydd)**

Bydd Dawn Knight yn cyflwyno papur byr o'r enw 'Corpora and Pedagogy: developing the community-driven National Corpus of Contemporary Welsh' yng nghynhadledd flynyddol Cymraeg i Oedolion yng Nghaerdydd.

➤ **Cynhadledd WISERD (13<sup>eg</sup> – 14<sup>eg</sup> Gorffennaf, Prifysgol Abertawe)**

Bydd Tîm Rheoli CorCenCC, Dawn Knight, Tess Fitzpatrick a Steve Morris yn cyd-gyflwyno papur yng nghynhadledd flynyddol Sefydliad Ymchwil Gymdeithasol ac Economaidd Cymru (WISERD) a gynhelir ym Mhrifysgol Abertawe. Sefydlwyd WISERD yn 2008 ac mae'n 'tynnu ynghyd ymchwilwyr y gwyddorau cymdeithasol o nifer o ddisgyblaethau, gan gynnwys cymdeithaseg, economeg, daearyddiaeth a gwyddor gwleidyddiaeth'. Ceir rhagor o wybodaeth am y digwyddiad hwn yma: <http://www.wiserd.ac.uk/training-events/annual-conference-2016/>

➤ **Eisteddfod Genedlaethol (29<sup>ain</sup> Gorffennaf i 6<sup>ed</sup> Awst, Y Fenni)**

Bydd aelodau o dîm CorCenCC yn mynychu'r Eisteddfod Genedlaethol er mwyn hysbysu'r cyhoedd am brif nodau'r prosiect, a manylu ar sut y gallan nhw fynd ynghlwm yn y gwaith. Os aiff popeth yn iawn, rydym hefyd yn gobeithio lansio'r ap torfoli ar gyfer casglu data yn swyddogol yn y digwyddiad. Cynhelir cyflwyniadau anffurfiol ar y prosiect ym mhabell Prifysgol Caerdydd am 11yb ar 2<sup>il</sup> Awst ac ym mhabell Prifysgol Abertawe am 11yb ar 3<sup>ydd</sup> Awst (ond byddwch yn dod o hyd i ni ar y safle drwy gydol y ddau ddiwrnod). Am ragor o fanylion, gweler: <https://eisteddfod.cymru/>

## Cwrdd â'r tîm

*Bob mis byddwn yn rhoi sylw i aelod gwahanol o'r tîm CorCenCC estynedig yn ein cylchlythyr. Bydd hyn yn rhoi cyfle i bawb ddweud ychydig wrthyf chi am eu cefndir; beth maen nhw am ei weld o CorCenCC a sut maen nhw'n credu y gallai gyfrannu at eu gwaith eu hunain, neu yn fwy cyffredinol, at waith eraill yng Nghymru. Y mis yma mae'r sbotolau ar Aran Jones, Prif Weithredwr SaySomethingin.com, sy'n rhan o Grŵp Ymgynghorol Prosiect CorCenCC.*

### Proffil: Aran Jones

Wedi cyrraedd byd llawen corpora yn hwyr braidd, roedd yn syndod i mi pan sylweddolais nad oedd un ar gael eisoes ar gyfer y Gymraeg - ac felly mae gweld y prosiect hwn yn datblygu yn wirioneddol cyffrous.



Does gen i ddim cefndir academiaidd mewn ieithyddiaeth (mi wn yn iawn nad yw gradd lenyddiaeth yn cyfrif, hyd yn oed o Aberystwyth!). Cefais fy magu, serch hynny, yng Nghymru, Lloegr, yr Almaen, Portiwgal, Sri Lanka a Malaysia, ac wedi gweithio wedyn yn Zimbabwe a Dubai - profiad arweiniodd at ddiddordeb byw mewn ieithoedd. Roeddwn i wastad yn gwybod y byddai rhaid i mi ddysgu Cymraeg rhywben - iaith Nain a Taid - felly rhoddais gais ar ddysgu pob iaith newydd des i ar ei thraws.

Roedd y canlyniadau yn ddychrynlyd o wael.

Ac eto, rhyw ffordd neu'i gilydd, dyma fi erbyn hyn yn siarad Cymraeg - a sbardunodd blas llwyddiant annisgwyl (wedi cymaint o fethiannau) ddiddordeb newydd mewn methodoleg. Dyna arweiniodd yn y pen draw at ddatblygu'r cwrs arlein llawn-iaith-fformiwläig SaySomethinginWelsh. Mae tua 40,000 o bobl wedi cael mynediad at ein cynnwys, mewn bron i 60 o wledydd gwahanol, ac rydym wedi





cychwyn prosiect dysgu cyfunol bach gyda Phrifysgol Bangor yn ddiweddar - mae'n teimlo'n bell iawn o Zimbabwe, a'r dwy flynedd ges i yno yn methu dysgu fawr mwy nac 'Arhoswch yn dda!' (sef 'Hwyl fawr!') a 'Gwych' ('Musara zvakanaka' a 'Zvakanaka', os oes diddordeb ganddoch chi, felly twyll byddai honni i mi ddysgu dwy ymadrodd llawn).

Des i yn rhan bell o'r prosiect hwn pan oedd Tess Fitzpatrick ac Alison Wray yn ddigon caredig i roi gwahoddiad i mi siarad am fy ngwaith yng Nghaerdydd. Yr adeg hynny, roeddwn i'n dal i ystyried 'Does gen i ddim syniad' fel ateb rhesymol i'r cwestiwn 'Pam ydi eich methodoleg yn llwyddo?' - ac felly estynnodd Tess ac Alison yr un gwahoddiad yn amyneddgar blwyddyn ar ôl blwyddyn hyd nes i mi gynnig atebion gwell (fyddai byth wedi cyrraedd fy mhen heblaw am yr oriau hirion yn darllen papurau ar ieithyddiaeth mewn panigs braidd ar y trên i Gaerdydd).

Dwi'n gwyllo'r prosiect yn datblygu gyda chryn dipyn o angerdd hunanol - bydd yn lleihau fy maich gwaith yn sylweddol. Mae llawer o'r hyn dwi'n gwneud yn dibynnu ar sgysiau model - mae cyfansoddi nhw fy hun yn waith digon diflas, gyda rhai cyfyngiadau naturiol (fel bydd unrhyw un sydd wedi rhannu'r profiad o glywed drymiwr 'death metal' tal, barfog, yn datw's i gyd yn cwyno'n groch nad ydi o ddim eisiau gwybod sut i hwyllo ar y blydi môr yn deall yn iawn). Bydd corpws i'r Gymraeg, yn enwedig un bydd gyda chymaint o gynnwys llafar â dwi'n gobeithio gweld yn hwn, yn gychwyn ar daith newydd difyr iawn.



Bydd modd i ni wirio popeth sydd gyda ni eisoes yn erbyn y corpws, er mwyn gweld os oes tyllau gwael neu gynnwys di-angen - ond tu hwnt i hynny, bydd modd i ni ddefnyddio fo er mwyn dylanwadu ar ein gwaith gydag estyn geirfa ac ymarfer gwrando, ac er mwyn rhoi arweiniad gwell o lawer i ddysgwyr sydd eisiau gwybod mwy am dafodieithoedd.

Bydd, heb os, yn gwella safon ein gwaith, ac rydym yn ddiolchgar iawn o flaen llaw. [Peidiwch â stopio rwan! Daliwch ati! Brysiwch!].

*Aran Jones, [aranjones@gmail.com](mailto:aranjones@gmail.com)*

## CorCenCC ar-lein

Mae'r wybodaeth ddiweddaraf am ddatblygiadau'r prosiect hefyd ar gael drwy Facebook [www.facebook.com/CorCenCC/](https://www.facebook.com/CorCenCC/); Twitter <https://twitter.com/corcencc> (gallwch ein trydar @CorCenCC). Gallwch hefyd gysylltu â ni drwy anfon neges i gyfeiriad ebost y prosiect: [corcencc@caerdydd.ac.uk](mailto:corcencc@caerdydd.ac.uk) neu ewch i'n gwefan, sef: <http://sites.cardiff.ac.uk/corcencc/>

Mae CorCenCC yn brosiect ymchwil a ariennir gan ESRC/AHRC (Grant Rhif ES/M011348/1). Mae tîm CorCenCC yn cynnwys y **Prif Ymchwilydd** - Dawn Knight; **y Cyd-Ymchwilywr** - Tess Fitzpatrick, Steve Morris, Irena Spasic, Paul Rayson, Enlli Thomas, Mark Stonelake a Jeremy Evas; **y Cynorthwywr Ymchwil** - Steven Neale, Jennifer Needs, Mair Rees, Scott Piao a Gareth Watkins; **Ymgynghorwyr** - Kevin Donnelly, Kevin Scannell, Laurence Anthony, Tom Cobb, Michael McCarthy a Margaret Deuchar; **Grŵp Ymgynghorol y Prosiect** - Karen Corrigan, Llion Jones, Maggie Tallerman, Mair Parry-Jones a Gwen Awbery, Emyr Davies (CBAC-WJEC), Gareth Morlais (Llywodraeth Cymru), Owain Roberts (Llyfrgell Genedlaethol Cymru), Aran Jones (Saysomethingin.com) ac Andrew Hawke (Geiriadur Prifysgol Cymru).

Os oes gennych unrhyw sylwadau neu gwestiynau am gynnwys y cylchlythyr hwn, cysylltwch â Dr Dawn Knight: [KnightD5@caerdydd.ac.uk](mailto:KnightD5@caerdydd.ac.uk)

