Cylchlythyr CorCenCC

Rhifyn 12: Gorffennaf 2017

Corpws Cenedlaethol Cymraeg Cyfoes
National Corpus of Contemporary Welsh

Cyfarchion gan y Prif Ymchwilydd

Croeso i rifyn 12 cylchlythyr CorCenCC! Mae'r haf yma unwaith eto, felly rwy'n siŵr y bydd llawer ohonoch chi'n pacio eich bagiau ac yn troi eich ebyst awtomatig ymlaen, yn barod ar gyfer egwyl haeddiannol! Er bod bwrlwm y myfyrwyr a'r arholiadau wedi gostegu yn ein sefydliadau academaidd, mae'r bwrlwm ym mhrosiect CorCenCC i'w weld o hyd yn swyddfeydd pawb sy'n rhan o'r tîm. Mae'r gwaith casglu data ar y gweill a bydd aelodau o'r tîm yn mynychu nifer o ddigwyddiadau cyhoeddus dros yr haf, er mwyn lledu'r gair am y prosiect a dweud wrth aelodau'r cyhoedd: 'Rhowch eich Cymraeg i ni'. Os byddwch chi'n ein gweld ni, dewch i ddweud 'shwmae'!

Mae rhifyn hwn y cylchlythyr yn cynnwys adroddiadau ar newidiadau diweddar o ran staff y prosiect, diweddariad ar weithgareddau mewn cynadleddau diweddar, a chyfle i gwrdd ag aelod arall o'r tîm estynedig – y tro hwn, un o ymgynghorwyr CorCenCC, yr Athro Kevin Scannell o Brifysgol Saint Louis.

Mwynhewch, a gobeithio cewch chi haf hyfryd, Dr Dawn Knight (Prifysgol Caerdydd)

Newyddion

Cyfle am swydd

Bydd y Cynorthwyydd Ymchwil Dr Gareth Watkins yn ymadael â'r prosiect ar ddiwedd mis Medi (byddwn ni i gyd yn drist iawn o weld Gareth yn mynd) ac felly, byddwn yn recriwtio CY newydd i'w ll/leoli ym Mhrifysgol Caerdydd. Swydd 23 mis yw hon, gan ddechrau ar 1 Hydref 2017. Gellir dod o hyd i fanylion am y swydd yma: https://krb-sjobs.brassring.com/TgNewUI/Search/Home/Home?partnerid=30011&siteid=5461 (rhif cyfeirnod y swydd yw 6328BR). Y dyddiad cau yw 11 Awst 2017. A wnewch chi anfon manylion y swydd hon at unrhyw un rydych chi'n credu y gallen nhw fod â diddordeb mewn gwneud cais!

Penodiad newydd

Rydym yn ffarwelio'n drist â Lorena Varona, cynorthwyydd gweinyddol CorCenCC, sydd wedi ymadael â thîm CorCenCC a symud yn ôl gartre i Sbaen. Pob hwyl am y dyfodol Lorena! Bydd Lowri Williams yn cymryd lle Lorena. Gofynnon ni i Lowri ei chyflwyno ei hun i gymuned CorCenCC…

Helo! Fy enw i yw Lowri ac rwy'n wreiddiol o Bwllheli, Gogledd Cymru. Cefais fy magu mewn cartref Cymraeg a derbyniais fy addysg drwy gyfrwng yr iaith fel iaith gyntaf.

Symudais i'r brifddinas fawr i astudio Systemau Gwybodaeth yn yr Ysgol Cyfrifiadureg a Gwybodeg, Prifysgol Caerdydd. Yn dilyn fy ngradd, symudais ymlaen i astudio PhD

yn yr ysgol o dan oruchwyliaeth yr Athro Irena Spasic, sydd hefyd yn aelod o dîm CorCenCC. Mae fy ymchwil yn edrych i mewn i rôl idiomau mewn dadansoddi sentiment mewn testun. Rwyf bellach wedi cyflwyno fy thesis ac yn aros am yr arholiad llafar.

Fel ymchwilydd Cymreig ifanc, mae'r gwaith y mae CorCenCC yn ei wneud yn bwysig. Nid yn unig y bydd y gwaith yn cefnogi'r iaith Gymraeg yn y dyfodol, ond mae yna botensial i'r gwaith fod yn sail gadarn ar gyfer ymchwil i ddatblygu ac integreiddio yr iaith, yn enwedig o bersbectif technoleg gyfrifiadurol ac ieithyddiaeth gyfrifiadurol.

Croeso mawr Lowri – rydym yn edrych ymlaen at weithio gyda ti ar brosiect CorCenCC!

A fyddai gennych chi ddiddordeb mewn gweithio fel trawsgrifydd ar gyfer CorCenCC?

Fel y gwyddoch, rydyn ni wedi bod yn brysur yn recordio pobl yn siarad Cymraeg ledled y wlad. Mae gwaith trawsgrifio'r recordiadau wedi dechrau, ond rydyn ni bellach yn chwilio am ragor o drawsgrifwyr – a fyddai gennych chi ddiddordeb? Mae'r gwaith yn hyblyg (cewch weithio pryd bynnag sy'n eich siwtio chi, a gwneud cymaint/cyn lleied o oriau a fynnwch chi) felly mae'n hawdd ei ffitio o gwmpas gweithgareddau eraill, ac mae'r recordiadau yn ddiddorol ac yn amrywiol – un diwrnod efallai byddwch chi'n trawsgrifio darlith neu bregeth, a'r diwrnod nesaf sgwrs fywiog lawr yn y dafarn! Os oes gennych chi ddiddordeb mewn ymuno â'n tîm o drawsgrifwyr, ebostiwch trawsgrifio@corcencc.org am ragor o wybodaeth, os gwelwch yn dda.

Fideo lansiad CorCenCC – ar gael ar-lein erbyn hyn!

Mae'n bleser gennym gyhoeddi bod fersiwn llawn o lansiad CorCenCC o Fawrth 2017 ar gael i'w wylio ar-lein erbyn hyn. I wylio'r digwyddiad, ewch i: http://www.swansea.ac.uk/cy/riah/prosiectau-ymchwil/corcencc/

Digwyddiadau diweddar

Tafwyl, Gŵyl Fach y Fro, Parti Ponty, a Chynhadledd y Ganolfan Dysgu Cymraeg Genedlaethol

Wrth i dymor y gwanwyn droi'n dymor yr haf, wrth i'r nosweithiau tywyll encilio ac wrth i bawb estyn am grys T, trowsus byr ac eli haul, daeth hefyd tymor y gwyliau Cymraeg. Roedd CorCenCC yn ddigon ffodus i brofi rhai o'r rhain, a thrwy hynny cwrdd â nifer fawr o siaradwyr Cymraeg. Roedd y siaradwyr yma yn awyddus iawn i glywed am y prosiect, ac yn fodlon iawn cyfrannu eu lleisiau at ein corpws. O heulwen ddisglair Tafwyl a Gŵyl Fach y Fro i law adfywiol Parti Ponty, roedd brwdfrydedd pobl dros yr iaith a thros brosiect CorCenCC yn amlwg iawn!

Daw hefyd tymor y cynadleddau adeg yma o'r flwyddyn. Ym mis Gorffennaf, cafodd aelod o'r tîm gyfle i fynychu Cynhadledd y Ganolfan Dysgu Cymraeg Genedlaethol, ac i recordio yna. Un o brif amcanion prosiect CorCenCC yw creu offer pedagogaidd, hynny yw adnoddau i helpu dysgwyr ac addysgwyr wrth iddynt ymgymryd â'r dasg (pleser?) o ddysgu ac addysgu'r iaith. Dyma enghraifft wych, felly, o ddarpar ddefnyddwyr y corpws yn cyfrannu at greu'r adnodd yn y lle cyntaf.

Diolchiadau lu i bawb sydd wedi cyfrannu at y gwaith pwysig hwn wrth fynychu'r digwyddiadau hyn. Rydyn ni'n edrych ymlaen at gwrdd â mwy byth o bobl a'u recordio yn y misoedd nesaf.

Gareth Watkins

Cynhadledd Ieithyddiaeth Corpws 2017

Ar ddydd Llun 24 Gorffennaf, neidiodd nifer o aelodau tîm CorCenCC ar y trên – yn syth i'r gogledd-ddwyrain – er mwyn mynychu'r gynhadledd Ieithyddiaeth Corpws (CL2017) ym Mhrifysgol Birmingham. Roedd y gynhadledd yn cynnwys diwrnod o weithdai a phedwar diwrnod ar gyfer y brif gynhadledd, ac roedd hi'n gyfle i ni ledaenu rhai o'r

syniadau, datblygiadau a gweledigaeth ar gyfer CorCenCC ymysg academyddion ac arbenigwyr eraill ym maes Ieithyddiaeth Corpws.

Cyflwynwyd y papur cyntaf yn ystod y gweithdy, a rhoddodd arolwg o'r prosiect, gan amlinellu prif amcanion CorCenCC ynghyd â rhai o'r heriau allweddol y mae'r ymchwil yn eu hwynebu. Dawn Knight a Paul Rayson a gyflwynodd y papur, ond roedd cyfanswm o 19 o gyd-awduron (prif aelodau tîm y prosiect) ynghlwm – record ar gyfer CL2017, yn bendant!

Yn ail i'r llawr oedd Mair Rees a Gareth Watkins (gweler y llun ar y chwith). Rhoddon nhw ddiweddariad ar waith Pecyn Gwaith 1 ac amlinelliad o'r glasbrintiau ar gyfer fframwaith samplo CorCenCC. Ar fore dydd Iau, cyflwynodd Steve Neale ei bapur ar ap torfoli CorCenCC (WP5), ac yn y prynhawn roedd cyflwyniad gan Jennifer Needs (WP4) a ofynnodd 'How will you Make Sure the Material is Suitable for Children?: User-Informed Design of Welsh Corpus-Based Learning/Teaching Tools'. Yn olaf ond nid y lleiaf, cyflwynodd Paul Rayson a Scott Piao eu gwaith ar dagiwr semantig CorCenCC (WP3 – gweler y llun uchod) ar ddiwrnod olaf y gynhadledd. Cyfanswm o 5 papur! Gwaith trawiadol – ond mae pawb wedi blino erbyn hyn!

Llwyddodd y gynhadledd i ysgogi, ymgysylltu ac ymgyfoethogi pawb a gymerodd ran. Cawson ni ein diddanu gan berfformaid byw'r grŵp 'Corpus Interruptus' (gweler y llun ar y chwith)… a doedd hyd yn oed glaw a gwynt Canolbarth Lloegr ddim wedi digalonni'r cyfranogwyr. Wrth i'r gynhadledd gloi, roedd gennym gyhoeddiad cyffrous: bydd y gynhadledd Ieithyddiaeth Corpws nesaf (a fydd yn digwydd yn 2019) yn cael ei chynnal yng Nghanolfan Ymchwil Iaith a Chyfathrebu Prifysgol Caerdydd! Gan y bydd y

gynhadledd hon yn cyd-daro â blwyddyn olaf prosiect CorCenCC, ein bwriad yw i'r gynhadledd gynnwys cymaint o ddigwyddiadau a gweithgareddau sy'n gysylltiedig â CorCenCC ag sy'n bosibl, gan gynnwys digwyddiadau Cymraeg arbennig – felly cadwch eich llygaid ar agor am ragor o fanylion. Edrychwn ymlaen at weld nifer ohonoch chi yn CL2019!

Dewin a Doti a Corpws

Aeth Gareth Watkins a Mair Rees draw i Sain Ffagan ddydd Llun 12fed o Fehefin i ymuno yn hwyl Gŵyl Dewin a Doti a drefnwyd gan y Mudiad Meithrin. Dewin yw cymeriad y Mudiad a Doti y ci yw ei ffrind ffyddlon. Maent yn cael eu defnyddio i gyflwyno a hyrwyddo'r Gymraeg i blant bach.

Mae'r ŵyl deithiol hon yn digwydd yn flynyddol ac yn cael ei chynnal mewn canolfannau hyd a lled Cymru. Mae'n gyfle i ddod â phlant a staff cylchoedd Meithrin a Ti a Fi yr ardal at ei gilydd i gael hwyl wrth ddefnyddio'u Cymraeg. Fel rhan o'r ŵyl roedd y bytholwyrdd Martyn Geraint yn cyflwyno sioe o hoff ganeuon y plant. Bu Martyn yn ddigon caredig i ganiatáu i CorCenCC recordio ei sesiwn er mwyn ei hychwanegu at y corpws.

Ond ymunodd un aelod arall o dîm CorCenCC yn yr hwyl hefyd: Corpws y gath! Cymeriad yw hwn sy'n helpu staff CorCenCC i esbonio gwaith y prosiect i blant ac sy'n ein cynorthwyo i ofyn am ganiatâd plant bach (ynghyd â chaniatâd eu rhieni) i gyfrannu'u Cymraeg llafar at y corpws. Dyma lun o Corpws yn cwrdd â'i ffrind Doti!

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 wrthych chi am eu cefndir; beth maen nhw'n dymuno ei weld gan 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, yr Athro Kevin Scannell sydd dan y chwyddwydr. Mae'n ymgynghorydd ar brosiect CorCenCC ac yn gweithio ym Mhrifysgol Saint Louis.

Proffil: Kevin Scannell

Yn gyntaf hoffwn ddweud pa mor falch rydw i fy mod yn gallu cyfrannu at brosiect CorCenCC fel ymgynghorydd, ac rwyf wedi mwynhau darllen y darnau "Cwrdd â'r Tîm" eraill yn fawr iawn!

Rwy'n Athro Cyfrifiadureg ym Mhrifysgol Saint Louis yn yr Unol Daleithiau lle rwy'n addysgu ers 1998. Rwy'n un o'r nifer bychan (ond cynyddol!) o Americanwyr sy'n siarad Gwyddeleg (dechreuais i ddysgu'r iaith o ddifri yn fy 20au) a'r dyddiau yma, mae'r rhan fwyaf o'r gwaith rwy'n ei wneud yn ymwneud â datblygu technoleg Wyddeleg o ryw fath neu'i gilydd. Rwyf wedi gweithio ar wirwyr sillafu a gramadeg,

geiriaduron, thesawrysau, peiriannau cyfieithu peirianyddol, corpora o fathau gwahanol ac rwyf wedi gwneud tipyn o ddadansoddi'r iaith yn y cyfryngau cymdeithasol. Yn y maes yma, mae fy nghefndir academaidd braidd yn anarferol; ces i fy hyfforddi fel mathemategwr a datblygu gyrfa academaidd trwy gyhoeddi ymchwil ar fathemateg haniaethol iawn (== ddim-yn-ddefnyddiol-o-gwbl) a ffiseg ddamcaniaethol. Yn ffodus, gwnes i ymbwyllo tua 2005 a dechrau ymroi o ran fy holl amser ymchwil i brosiectau'n ymwneud â'r Wyddeleg ac ieithoedd brodorol/lleiafrifol eraill. Yn y bôn, rwy'n credu y dylai pawb allu defnyddio'r cyfrifiadur 100% yn eu hiaith frodorol; prosiectau corpws fel CorCenCC a'r ymchwil a'r technolegau sy'n deillio ohono fydd yn gwneud hyn yn realiti i'r Gymraeg, y Wyddeleg a llawer o ieithoedd eraill.

Mae bron pob un o'm prosiectau'n cael eu gyrru gan gorpora a dulliau ystadegol ac mae hyn yn mynd yr holl ffordd yn ôl i tua 1999 pan ddechreuais i weithio ar y gwirydd sillafu "GaelSpell" i'r Wyddeleg. A siarad yn blaen, doedd dim syniad gen i beth roeddwn i'n ei wneud ar y pryd, ar wahân i gydnabod pe bawn i am gynhyrchu rhestr eiriau fawr i'r iaith, yna dylwn gychwyn trwy roi corpws mawr o destunau at ei gilydd a'i "lanhau". Yn ffodus, roedd eithaf tipyn o wefannau Gwyddeleg eisoes yn bodoli ar y pryd yn ogystal â chymuned ar-lein fywiog wedi'i chanoli ar *listservs* ebost fel Gaelic-L a Gaeilge-A (nhw oedd "rhwydweithiau cymdeithasol" y cyfnod). Ysgrifennais i ambell sgript gragen gyflym i grafu'r testun o'r holl safleoedd hyn, penderfynu sut i wneud adnabyddiaeth iaith ystadegol seml ac erbyn 2004 roeddwn wedi adeiladu corpws o tua 20 miliwn o eiriau Gwyddeleg.

Fel mae'n digwydd, y Gymraeg oedd yr iaith gyntaf (ar wahân i'r Wyddeleg) i mi adeiladu corpws a we-ymlusgwyd ar ei chyfer, gan gydweithio â Dewi Evans o Goleg y Brifysgol Dulyn ac Andrew Hawke o Eiriadur Prifysgol Cymru ar gorpws mawr o'r Gymraeg trwy we-ymlusgwr ym mis Chwefror 2004. Roedd symud pethau drosodd i'r Gymraeg yn golygu cryn dipyn o ailffactoreiddio'r ymlusgwr i ddileu unrhyw ragdybiaethau "Gwyddeleg" a gwneud y cyfan yn fwy annibynnol o ran iaith.

A dyma sut y dechreuodd ymgais deng mlynedd 'bach yn obsesiynol o ymlusgo cymaint o ieithoedd gwahanol ag sy'n bosibl; rwyf wedi llwyddo i ddefnyddio'r gwe-ymlusgwr ac adeiladu corpora ar gyfer mwy na 2100 o ieithoedd gwahanol fel rhan o brosiect o'r enw "Crúbadán" (http://crubadan.org/).

Eto, rwyf wrth fy modd yn cyfrannu at y prosiect cyffrous hwn ac yn edrych ymlaen at ddod i nabod gweddill y tîm!

Kevin Scannell

CorCenCC ar-lein

Mae'r wybodaeth ddiweddaraf am ddatblygiadau'r prosiect hefyd ar gael drwy Facebook 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 neu ewch i'n gwefan, sef: www.corcencc.cymru

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-Ymchwilwyr** - Tess Fitzpatrick, Steve Morris, Irena Spasić, Paul Rayson, Enlli Thomas, Mark Stonelake a Jeremy Evas; **y Cynorthwywyr Ymchwil** - Steven Neale, Jennifer Needs, Mair Rees, Scott Piao a Gareth Watkins; **y myfyriwr PhD** - Vigneshwaran Muralidaran; **Ymgynghorwyr** - Kevin Donnelly, Kevin Scannell, Laurence Anthony, Tom Cobb, Michael McCarthy a Margaret Deuchar; **Grŵp Ymgynghorol y Prosiect** - Colin Williams, Karen Corrigan, Llion Jones, Maggie Tallerman, Mair Parry-Jones, 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