



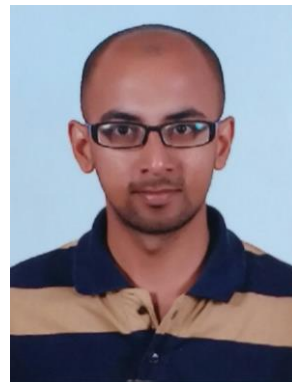
Cyfarchion gan y Prif Ymchwilydd

Croeso i nawfed rhifyn cylchlythyr CorCenCC. Fel y gwyddoch, rydyn ni'n bwriadu casglu o leiaf 10 miliwn o eiriau o Gymraeg (fel y mae hi'n cael ei defnyddio) erbyn diwedd y prosiect. Er mwyn rhoi gwell syniad i chi o'r gwahanol fathau o ddata sy'n mynd i gael eu cynnwys, a natur y cyfranogwyr yr hoffon ni eu denu i gyfrannu, dros dri chylchlythyr byddwn yn cynnig trosolwg o'r 'fframwaith sampl' h.y. cynlluniau casglu data ar gyfer e-iaith (rhifyn 8), iaith lafar (y rhifyn hwn) ac iaith ysgrifenedig (y rhifyn nesaf). Gobeithio bydd hyn yn helpu rhoi darlun ehangach o'r prosiect i chi ac yn eich temtio i gymryd rhan a chyfrannu'ch Cymraeg chi! Hefyd, yn y rhifyn hwn, cewch newyddion diweddaraf y prosiect ac, unwaith eto, byddwn yn cyflwyno aelod arall o'r tîm (Steve Morris).

Mwynhewch! Dr Dawn Knight (Prifysgol Caerdydd)

Newyddion

Mae'n bleser gennym groesawu aelod newydd i dîm CorCenCC, Vigneshwaran Muralidaran! Mae Vignesh newydd ddechrau PhD ym Mhrifysgol Caerdydd, a bydd ei waith doethurol yn cydblethu â rhai o weithgareddau prosiect CorCenCC. Caiff Vignesh ei oruchwylio gan Dr Dawn Knight a'r Athro Irena Spasić. Fe wahoddon ni i Vignesh sôn ychydig amdano fe ei hunan: *Graddiais i gyda gradd Meistr mewn Ieithyddiaeth Gyfrifiadurol o Sefydliad Rhyngwladol Technoleg Gwybodaeth Hyderabad (IIIT-H) yn 2016. Dros y blynyddoedd hyn gweithiais i yn y sefydliad fel Cynorthwydd Ymchwil yng Nghanolfan Ymchwil Technoleg Iaith (LTRC) ac fel cynorthwydd dysgu ar gyfer cwrs Ieithyddiaeth Gyfrifiadurol. Roedd fy ymchwil yn ymwneud â ffurfio fframwaith damcaniaethol ar sail Gramadeg Cystrawennau er mwyn dosrannu ieithoedd Drafidaidd a datblygu dosrannwr llawn ar gyfer Tamileg. Enillais i fy ngradd Baglor mewn Technoleg Gwybodaeth o Amrita Vishwa Vidyapeetham, Coimbatore yn 2010 ac yna gweithiais i fel rhaglennwr-ddadansoddwr mewn diwydiant am dair blynedd cyn ymuno ag IIIT-H. Ar hyn o bryd, fy niddordeb yw archwilio goblygiadau trin iaith fel system sy'n seiliedig ar ffwythiannau ac ar ddefnydd, a sut y gellir manteisio ar hyn mewn Technoleg Iaith.*



Rho dy Gymraeg i ni: Ffocws ar iaith lafar

Fel y gwyddoch, un o amcanion allweddol prosiect CorCenCC yw creu corpws sy'n gytbwys ac sy'n adlewyrchu'r holl ffyrdd y mae'r Gymraeg yn cael ei defnyddio o ddydd i ddydd yn y byd 'go iawn'. Mae hyn yn golygu y bydd y corpws yn cynnwys iaith o ystod o fathau gwahanol, yn trafod pynciau gwahanol, gan wahanol gyfranwyr o bob cefndir. Bydd pedair miliwn o'r deg miliwn o eiriau y gobeithiwn eu casglu yn dod o ffynonellau llafar.

Caiff data iaith llafar CorCenCC eu casglu o 7 prif gyd-destun defnyddio iaith yng Nghymru, o'r rhai sy'n gysylltiedig ag iaith fwy ffurfiol megis cyd-destunau sefydliadol a'r teledu/radio i'r rhai sy'n gysylltiedig ag iaith lai ffurfiol megis cyd-destunau preifat a chymdeithasol. Cafodd ein 'fframwaith sampl' ar gyfer iaith lafar – y fframwaith a fydd yn ein harwain wrth i ni ddewis beth i'w recordio – ei lunio'n ofalus iawn, gan ystyried y niferoedd o bobl sy'n cynhyrchu mathau penodol o iaith yn rheolaidd (e.e. mae mwy o bobl yn defnyddio'r Gymraeg yn breifat ac yn gymdeithasol nag yn broffesiynol) yn ogystal â maint cynulleidfa mathau penodol o iaith (e.e. iaith y teledu a'r radio). Ein nod yw adlewyrchu amrywiaeth a chyfoeth yr iaith Gymraeg, nid yn unig drwy gasglu swm bychan o bob math/genre, ond drwy gasglu swm *prïodol* o bob un, er mwyn creu 'ciplun' o'n hiaith fel y mae hi'n cael ei defnyddio heddiw.

Sail ein 7 cyd-destun iaith lafar oedd y modelau casglu data a ddefnyddiwyd ar gyfer y BNC (British National Corpus) a CANCODE (Cambridge and Nottingham Corpus of Discourse in English). Cafodd y genres a'r is-genres o fewn pob cyd-destun eu teilwra'n arbennig ar gyfer y Gymraeg, ac maen nhw felly yn cynnwys iaith a ddefnyddir mewn Eisteddfodau, iaith o addysg cyfrwng Gymraeg ac o ddsbarthiadau Cymraeg 2^{ll} iaith, a tharged cymharol fawr ar gyfer iaith o raglenni teledu i blant. Yn ogystal â cheisio sicrhau cydbwysedd priodol y gwahanol genres, mae CorCenCC hefyd yn bwriadu cynrychioli siaradwyr Cymraeg eu hunain yn briodol. Mae hyn yn golygu y bydd ein cynorthwywyr ymchwil yn teithio ledled Cymru – i ardaloedd gwledig yn ogystal â threfi a dinasoedd – a recordio siaradwyr o bob oedran a chyda phob math o gefndiroedd ieithyddol. P'un a ydych chi wedi bod yn siarad Gymraeg trwy gydol eich oes, p'un a ddysgoch chi'r iaith drwy addysg cyfrwng Cymraeg neu Saesneg, neu p'un a ddysgoch chi'r Gymraeg yn fwy diweddar neu yn wir yn ei dysgu nawr, y neges i bob unigolyn yw: **rho dy Gymraeg i ni!** Mae pob siaradwr Cymraeg yn cyfrannu at y mosäig sy'n ffurfio'r gymuned Gymraeg, a byddai CorCenCC wir yn hoffi gallu adlewyrchu hyn.

Mae'r fframwaith samplu ar gyfer data iaith llafar isod. Eto, dim ond canllaw yw hwn ar gyfer yr hyn yr hoffen ni ei gasglu – 'delfryd'. Mewn gwirionedd mae'n debyg y bydd dosbarthiad y data yn CorCenCC yn eitha gwahanol o'i gymharu â hwn, ond mae'n gweithredu fel man cychwyn/sail ddefnyddiol i ni adeiladu arno. Cymerwch gip arno ac, os hoffech chi gyfrannu unrhyw rai o'r mathau hyn o ddata iaith i'r corpws, cysylltwch â ni! **Mae dwy brif ffordd i chi gyfrannu data iaith llafar: (1) cysylltu â ni i ddarganfod pryd bydd cynorthwywyr ymchwil CorCenCC yn recordio yn eich ardal chi, neu (2) recordio chi eich hun drwy ddefnyddio ein 'ap' newydd sbon!**

Genre	Is-genre	%	Geiriau
Cyhoeddus/Sefydliadol		10%	400,000
Areithiau gwleidyddol a darlithoedd cyhoeddus	e.e. darllodiadau etholiadol, areithiau gan Aelodau Seneddol/Aelodau'r Cynulliad/cynghorwyr, areithiau gan Gymdeithas yr Iaith/Cyfeillion y Ddaear/ac ati, darlithoedd cyhoeddus	3.75%	150,000
Iaith seneddol	e.e. trafodion seneddol	1.25%	50,000
Iaith ffurfiol yn yr eglwys/capel	e.e. pregethau, gweddïau	3.75%	150,000
Iaith ffurfiol o'r Eisteddfod Genedlaethol	e.e. beirniadaethau Cerddoriaeth/Dawns/Llefaru/Drama, cyflwyniadau wrth stondinau prifysgolion/yn y Lolfa Lên/ym Maes D/ac ati	1.25%	50,000
Cyfryngau		15%	600,000
Rhaglenni teledu (BBC/S4C)	Plant	3.375%	135,000
	Adloniant	1.275%	51,000
	Ffeithiol	1.25%	50,000
	Drama	0.9%	36,000
	Chwaraeon	0.9%	36,000
	Cerddoriaeth	0.275%	11,000
	Newyddion	0.275%	11,000
Rhaglenni radio (BBC Radio Cymru)	Adloniant	1.375%	55,000
	Ffeithiol	1.375%	55,000
	Newyddion	1.1%	44,000
	Cerddoriaeth	0.825%	33,000
	Crefydd a Moeseg	0.275%	11,000
	Chwaraeon	0.275%	11,000
Rhaglenni radio masnachol/cymunedol	Comedi	0.275%	11,000
		1.25%	50,000

Trafodol		10%	400,000
Derbyn/Darparu nwyddau, gwybodaeth a gwasanaethau (wyneb-yn-wyneb neu dros y ffôn)	Sector cyhoeddus, e.e. cynghorau, swyddfa Comisiynydd y Gymraeg, swyddfeydd post, Cyllid a Thollau EM, llyfrgelloedd	3.5%	140,000
	Manwerthu, e.e. siopau, ocsiynau	1.5%	60,000
	Bwyd, e.e. bwytai, caffis	1.5%	60,000
	Twristiaeth a theithio, e.e. gwybodaeth i dwristiaid, llety, ymholiadau trên cenedlaethol	1.5%	60,000
	Hamdden, e.e. cyfryngau, clybiau, adloniant, harddwch	1.5%	60,000
	Cyllid, y gyfraith ac iechyd	0.5%	20,000
Proffesiynol		10%	400,000
Cymraeg y Gweithle	Cyfarfodydd staff	2.5%	100,000
	Cyfweliadau	2.5%	100,000
	Sgyrsiau rhwng desgiau	2.5%	100,000
	Cyfathrebu rhwng busnesau (wyneb-yn-wyneb neu dros y ffôn)	1.25%	50,000
	Galwadau ffôn mewnol	1.25%	50,000
Pedagogaid		10%	400,000
Cyfathrebu mewn dosbarthiadau cynradd	Cyfathrebu mewn dosbarthiadau cyfrwng Cymraeg gan amrywiaeth o grwpiau oedran	3.5%	140,000
Cyfathrebu mewn dosbarthiadau uwchradd	Gwersi Gwyddoniaeth a Mathemateg cyfrwng Cymraeg	1%	40,000
	Gwersi Celf, Drama a Cherddoriaeth cyfrwng Cymraeg	0.375%	15,000
	Gwersi Daeryddiaeth, Hanes ac Astudiaethau Crefyddol cyfrwng Cymraeg	0.75%	30,000
	Gwersi Cymraeg cyfrwng Cymraeg	0.875%	35,000
Cyfathrebu mewn dosbarthiadau uwchradd ac Addysg Bellach (AB)	Gwersi Cymraeg (ail iaith) mewn ysgolion/ colegau cyfrwng Saesneg	0.75%	30,000
Cyfathrebu mewn dosbarthiadau Addysg Bellach (AB)	Gwersi cyfrwng Cymraeg mewn amrywiaeth o bynciau, e.e. Sgiliau sylfaenol, Astudiaethau diwylliannol, Gofal iechyd, TG	1%	40,000
Cyfathrebu mewn dosbarthiadau Addysg Uwch (AU)	Gwersi AU cyfrwng Cymraeg mewn amrywiaeth o bynciau, e.e. leithyddiaeth, Celfyddydau, Meddygaeth	0.5%	20,000
Cyfathrebu mewn dosbarthiadau i oedolion	Addysg gymunedol cyfrwng Cymraeg	0.5%	20,000
	Gwersi Cymraeg i oedolion	0.75%	30,000
Cymdeithasu		22.5%	900,000
Sgyrsiau gyda ffrindiau/teulu	Sgyrsiau yn y cartref gyda grwpiau o ffrindiau	7%	280,000
	Sgyrsiau gyda ffrindiau/teulu mewn llefydd cyhoeddus	6.75%	270,000
	Sgyrsiau gyda ffrindiau yn ystod gweithgareddau hamdden	6.75%	270,000
Cyfathrebu anffurfiol yn yr Eisteddfod Genedlaethol	Sgyrsiau mewn mannau cymdeithasol megis pabell Merched y Wawr, Maes D, lolfa'r cystadleuwyr, y Lle Celf	1.25%	50,000
Cyfathrebu anffurfiol yn Eisteddfod yr Urdd	Sgyrsiau mewn mannau cymdeithasol megis yr ardal fwyd, Caffi Mistar Urdd	0.75%	30,000
Preifat		22.5%	900,000
Sgyrsiau gyda'r teulu	Sgyrsiau rhwng rhiant/rhieni a phlentyn/phlant	4.25%	170,000
	Sgyrsiau gyda phartneriaid (priodasol)	4.25%	170,000
	Sgyrsiau rhwng aelodau eraill o'r teulu	4%	160,000
	Galwadau ffôn i aelodau o'r teulu	2%	80,000
Sgyrsiau gyda ffrindiau	Sgyrsiau gyda chyd-breswylwyr	3%	120,000
	Sgyrsiau gyda ffrindiau agos mewn lleoliadau preifat	3%	120,000
	Galwadau ffôn i ffrindiau agos	2%	80,000
		100%	4,000,000

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 wrthyhych 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 Steve Morris, sy'n Gyd-Ymchwilydd ar brosiect CorCenCC ym Mhrifysgol Abertawe.

Proffil: Steve Morris

Un o gryfderau CorCenCC yw bod pob un ohonon ni wedi dod ato o gyfeiriad gwahanol. Ers dros dri deg o flynyddoedd rwyf wedi bod yn gweithio ym maes Cymraeg i Oedolion [CiO], gan ddysgu fy nghwrs Wlpan cyntaf yn 1982 yma yn Abertawe ac wedyn yn dychwelyd i'm *alma mater* fel darlithydd mewn Addysg Barhaus (Cymraeg) yn 1991. Yr hyn sydd wedi fy ngyrru i (yn ogystal â chynifer o bobl eraill sy'n gweithio yn y maes) yw sicrhau bod cymaint o bobl ag sy'n bosibl yn cael y cyfle i droi'n siaradwyr newydd o'r Gymraeg a thrwy hynny, yn cael bod yn rhan o'r ymdrechion i sicrhau dyfodol cyffrous a chynaliadwy i'r iaith. Rwyf wedi dadlau erioed y dylai CiO fod yn rhan greiddiol o gynllunio ieithyddol yng Nghymru nid dim ond oherwydd ei fod yn faes sy'n ymwneud â mwyafrif y boblogaeth sy'n methu â siarad Cymraeg ond hefyd oherwydd ei fod yn ffordd effeithiol o greu'r siaradwyr newydd hynny sydd eu hangen i helpu rhoi polisïau a strategaethau iaith y llywodraeth ar waith. Yn y cyd-destun yna, mae llawer o'm gwaith wedi troi o gwmpas yr hyn sy fel arfer yn cael ei alw'n 'gynllunio statws'.

Dim ond yn gymharol ddiweddar yn fy ngyrfa – ar ôl arbenigo mewn gwaith ar gymhelliant a llwyddiant ym maes CiO – y des i nabod gwaith cydweithwyr mewn leithyddiaeth Gymhwysol yma ym Mhrifysgol Abertawe, yn enwedig Paul Meara a Tess Fitzpatrick. Mae'r awydd i sicrhau bod ymchwil CiO yn ymwneud â gwir anghenion ymarferwyr a dysgwyr wastad wedi bod yn gryf o fewn y ddisgyblaeth. Golygai natur '*gymhwysol*' y gwaith yma (cyn i agenda 'traweffaith' prifysgolion ymwreiddio o ddifrif) y gallai gael ei ddiystyru weithiau fel rhywbeth nad oedd yn 'ymchwil go iawn'. Rwy'n cofio bod mewn cynhadledd ymchwil CiO lle roedd Paul newydd roi papur ar gaffael geirfa a dyna pryd dechreuon ni drafod meysydd posibl lle gallen ni gydweithio â'n gilydd (meddai Paul: '*Rho broblem i fi rwyf ti eisïau ei datrys, Steve ac fe helpa'i di i ddod o hyd i ateb iddi...*') a sylweddolais mai'r elfen '*gymhwysol*' yma oedd gyda ni yn gyffredin! Trwy hyn, dechreuon ni weithio ar greu geirfâu craidd ar gyfer lefelau A1 ac A2 yn y Gymraeg – fy nghwestiwn gwreiddiol i Paul oedd '*Sut gallwn ni fod yn siŵr bod ni'n addysgu'r eirfa sydd ei hangen ar ein dysgwyr yn absenoldeb corpws cynhwysfawr o'r Gymraeg?*' Dyma fi felly – ar fin troi'n ieithydd cymhwysol a dablo gyda 'chynllunio corpws'... ac yna, dyma Tess Fitzpatrick yn ymddangos.

Trwy'r gwaith geirfa a seminarau Gregynog, trafododd Tess a fi fwy a mwy o syniadau ar gyfer cydweithio ac fe ddaeth yn glir i mi, er bod cymaint o orgyffwrdd rhwng ein disgyblaethau a photensial mawr i weithio gyda'n gilydd, roedd llawer ohonon ni yn y Gymraeg wedi treulio blynyddoedd yn torri ein cwys ein hunain a pheidio ag ymwneud â'r bobl iaith, wych a brwd yma ar ben arall y campws! Yn fuan ar ôl i mi symud o Addysg Barhaus i Adran y Gymraeg *per se* y cyflwynodd Tess Dawn i mi (oedd ar y pryd hanner ffordd trwy restr o bethau roedd hi eisïau eu gwneud i ddatlu ei phen-blwydd yn 30 oed). Roedd y cyfarfod hanner diwrnod gyda'r tri ohonon ni ym Mar Caffé Bonaparte yng ngorsaf Temple Meads Bryste ar 1 Mehefin 2012 yn lleoliad annhebygol i roi cychwyn i gorpws cenedlaethol Cymraeg cyfoes ond fe weithiodd ac fe blannwyd hedyn CorCenCC.



Gallai'r llun yma o gapel anghydfurfiol yn Abertawe ymddangos fel ffordd ystrydebol o gynrychioli'r iaith mewn cylchlythyr sy'n eich gwahodd i gwrdd â'r tîm... ond mynnwch gip arall. Arferai'r adeilad gwreiddiol sefyll mewn rhan wahanol o'r ddinas ac roedd yn gysylltiedig â'r papur newyddion Cymraeg cyntaf i gael ei gyhoeddi, sef 'Seren Gomer'. Cafodd ei adleoli i'r adeilad yn y llun yn 1962 ac yn 2015 cafodd ei brynu gan gymuned Gristnogol Tseiniaidd Abertawe. Y neges bwysig yma yw bod natur y gymuned Gymraeg wedi newid yn gyson trwy gydol ei hanes ac mae'n newid o hyd: dylen ni groesawu hyn! Mae'n arwydd fod yr iaith yn iach. Dydy'r syniad o siaradwr Cymraeg sy'n uniaith (ar wahân i rai plant ifanc iawn) ddim yn bodoli bellach ac fel arweinydd y pecyn gwaith sy'n canolbwyntio ar gasglu data fydd yn cael eu bwydo i mewn i'r corpws, mae o'r pwys mwyaf ein bod yn gwneud yn siŵr ein bod yn casglu enghreifftiau o Gymraeg cyfoes yn ôl egwyddorion penodol sy'n cynnwys pob amrywiad o'r iaith fel y mae heddiw.

I mi, mae potensial hyn oll yn hynod o gyffrous (testun erthygl arall i'r cylchlythyr siŵr o fod!). Rwy'n byw mewn dinas odidog mewn lleoliad gwych ar lan y môr ac yma, nawr, yn 2017, mae'r iaith Gymraeg yn dal i fod yn



'Rho dy Gymraeg i ni!'

rhan fawr o fywydau nifer ohonon ni sy'n byw yma (a phob man arall yng Nghymru) er bod hynny'n digwydd mewn ffyrdd gwahanol ac mewn peuoedd gwahanol i'r gorffennol. Mae gyda ni addysg trwy gyfrwng y Gymraeg, rydyn ni'n tecstio, yn sgeipio ac yn trydar yn Gymraeg, rydyn ni'n gwyllo'r teledu sy'n dod o loeren yn Gymraeg ac rydyn ni'n dal i siarad yr iaith wyneb-yn-wyneb â'n gilydd ar draws y cenedlaethau.

Mae cerdd Gymraeg enwog yn sôn am sut mae llawer ohonon ni yn hwyrach yn ein bywydau yn gallu gweld yn glir yr elfennau hynny sydd wedi ein mowldio ni. O edrych yn ôl, gallaf weld fy mod wedi bod yn gweithio'n aml ar y rhyngwyneb rhwng y Gymraeg ac leithyddiaeth Gymhwysol heb yn wybod i mi. Mae datblygu, meithrin a sicrhau bod y cyswllt hwn yn datblygu yn waddol penodol a fwriadwyd ar gyfer prosiect CorCenCC ac un y mae gen i ymrwymiad angerddol iddo.

Steve Morris, S.Morris@Abertawe.ac.uk

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: <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-Ymchwilydwr** - Tess Fitzpatrick, Steve Morris, Irena Spasić, Paul Rayson, Enlli Thomas, Mark Stonelake a Jeremy Evas; y **Cynorthwyydwr 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** - 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



Arts & Humanities
Research Council