



## Cyfarchion gan y Prif Ymchwilydd

Felly, dyma ni – a chroeso i rifyn olaf ond un cylchlythyr CorCenCC! Bydd y gwaith ar y rhan o brosiect CorCenCC sy'n derbyn cyllid yn dod i ben ddiwedd mis Awst 2019 ond mae'n dda gen i gyhoeddi ein bod wedi derbyn estyniad heb gyllid i'r Gwaith tan



## Cynnwys

- T1:** Newyddion a digwyddiadau
- T3:** Y diweddaraf am y gwaith maes
- T4:** Ymgyrch ddata
- T5:** Cwrdd â'r tîm
- T7:** Cysylltwch â ni



ddiwedd Mai 2020 er mwyn sicrhau bod popeth yn cael ei wirio a'i profi'n drylwyr, gan wneud yn siŵr ei fod yn barod i'w lansio. Mae amrywiaeth eang o ddigwyddiadau cyhoeddus a stondinau teithiol ar y gweill yn ystod y flwyddyn nesaf felly mynnwch gip ar ein cyfryngau cymdeithasol am fanylion y rhain! Yn y rhifyn ('bach yn fyrrach) yma o'r cylchlythyr, rydym yn rhoi'r newyddion diweddaraf i chi am Gyd-Ymchwilydd y prosiect, Enlli Thomas, gallwch ddarllen y diweddaraf am ddigwyddiadau sy'n ymwneud â CorCenCC a chael eich cyflwyno i ddau aelod gwerthfawr arall o dîm CorCenCC. Mwynhewch y darllen – a byddwch yn barod am rifyn terfynol, sylweddol ym mis Gorffennaf 2019. Mae diwedd dechrau CorCenCC ar y gorwel ond bydd yr antur yn parhau...

*Dr Dawn Knight*

## + Newyddion a digwyddiadau

Dyfarwyd Medal Hugh Owen Cymdeithas Ddysgedig Cymru i'r Athro Enlli Thomas, Athro a Chyfarwyddwr Ymchwil ac Effaith, yr Ysgol Addysg, Prifysgol Bangor, am gyfraniadau i ymchwil addysgol, i gydnabod ei harbenigedd ar y Gymraeg, dwyieithrwydd ac astudiaethau mewn addysgu, dysgu a defnyddio'r Gymraeg.

Dyfarwyd medalau'r Gymdeithas mewn digwyddiad a gynhaliwyd yng Ngholeg Brenhinol Cerdd a Drama Cymru yn ddiweddar i ddatlu llwyddiant yn y byd academiaidd. Yn anffodus nid oedd yr Athro Thomas yn bresennol gan ei bod mewn Cynhadledd Ryngwladol ar y pryd.

Wrth dderbyn y wobr, dywedodd yr Athro Thomas: "Rwy'n falch iawn o fod wedi cael fy enwebu ar gyfer y wobr arbennig hon eleni, ac yn hynod ddiolchgar i'r rhai hynny a fy enwebodd ac i'r Gymdeithas am ddyfarnu'r fedal hyfryd hon i mi gan roi cydnabyddiaeth i'm gwaith. Mae hi wir yn fraint ac yn anrhydedd cael bod wedi gweithio mewn maes sydd yn agos iawn at fy nghalon ers dros 20 mlynedd bellach – sef datblygiad y Gymraeg a dwyieithrwydd mewn plant – ac mae'n braf cael bod yn rhan o'r bwrlwm cenedlaethol wrth i ni lunio strategaethau ac ymyraethau addysgol, yn seiliedig ar dystiolaeth, er mwyn cymell rhagor o ddefnyddwyr o'r Gymraeg erbyn 2050."



## CorCenCC @ HealTAC 2019

**Ymgasglodd panel o ymchwilyr, cleifion, clinigwyr a chwmnïau yng Nghynhadledd Dadansoddi Testun Gofal Iechyd (HealTAC) yng Nghaerdydd i drafod y gwaith gorau o ran prosesu testun gofal iechyd rhydd pan fydd rhwystrau iaith yn**

**bodoli...** Pan fydd claf a'r gweithiwr iechyd proffesiynol sy'n gofalu amdano ill dau yn meddu ar ieithoedd cyntaf heblaw Saesneg, beth yw'r ffordd orau o gynnal ymgynghoriadau, cadw cofnodion a'u rhannu? Gan i Gynhadledd Dadansoddi Testun Gofal Iechyd y DU eleni gael ei threfnu ym mhrifddinas Cymru, ystyriodd grwpiau paneli'r gynhadledd y mater hwn yn achos y Gymraeg. Bydd y drafodaeth, a gynhaliwyd yng Ngwesty'r Park Plaza ar 24 – 25 Ebrill, yn cefnogi'r gwaith o ddatblygu methodolegau arolygu, prosesau ac adroddiadau newydd i gefnogi'r broses o weithredu 'Mwy na geiriau: Fframwaith strategol dilynol', a gyhoeddwyd yn 2012 gan y Dirprwy Weinidog dros Wasanaethau Cymdeithasol yn Llywodraeth Cymru. Nod y fframwaith oedd sicrhau bod sefydliadau perthnasol yn cydnabod bod



iaith yn rhan gynhenid o ofal ac y cynigir gwasanaethau yn Gymraeg i'r bobl sydd angen gwasanaethau hynny. Yr enw ar hyn yw'r 'Cynnig Rhagweithiol'.

Bwriad y fframwaith strategol dilynol 2016-2019 oedd cadw momentwm ac adeiladu ar y strategaeth flaenorol, gan gynnwys cefnogaeth ymarferol i weithredu'r 'Cynnig Gweithredol'. Ystyriodd y panel ddulliau technegol a all hwyluso'r ddarpariaeth o wasanaethau yn y Gymraeg, gan gynnwys adnoddau terminolegol, lleoleiddio meddalwedd a chyfieithu peirianyddol, yn ogystal ag amrywiaeth o feysydd cymhwyso, megis presgripsiynau, iechyd meddwl ac alergeddau. Cafwyd cyfraniadau gwerthfawr gan ymchwilyr lleol,



clinigwyr a chyfranogwyr rhyngwladol sydd â phrofiad o ymdrin â sawl iaith, gan gynnwys Dr Antoine Pironet o Gofrestrfa Canser Gwlad Belg a gyflwynodd eu gwaith ar brosesu data o adroddiadau patholeg dwyieithog.

Trefnwyd y panel gan Gareth Morlais o Lywodraeth Cymru ac aelodau tîm CorCenCC, Dawn Knight, Laura Arman ac Irena Spasić. Mae Prifysgol Caerdydd yn rhan o prosiect CorCenCC (Corpws Cenedlaethol Cymraeg cyfoes), sy'n creu adnodd iaith mawr i siaradwyr Cymraeg, dysgwyr Cymraeg, ymchwilyr y Gymraeg ac unrhyw un sy'n ymddiddori yn y Gymraeg. Prosiect a yrrir gan y gymuned yw CorCenCC, a gall siaradwyr Cymraeg o bob math o gefndiroedd ac o bob gallu gymryd rhan. Dywedodd Dawn Knight, o dîm CorCenCC: "Oherwydd pryderon ynghylch preifatrwydd, nid ydym yn casglu naratifau gofal iechyd ar hyn o bryd, ond gellir defnyddio seilwaith casglu a rheoli data CorCenCC yn hawdd mewn amgylchedd diogel er mwyn cefnogi'r gwaith o ddatblygu offer cloddio testun a allai hwyluso ddarparu gwasanaethau gofal iechyd yn Gymraeg."

Ychwanegodd Irena Spasić, cyd-sylfaenydd y rhwydwaith HealTex: "Rydym wedi bod yn datblygu offer ac adnoddau a all awtomeiddio'r gwaith o brosesu testun Cymraeg. Gall ein meddalwedd FlexiTerm echdynnu terminoleg sy'n benodol i barth o ddogfennau Cymraeg a gall wneud hynny'n syth. Roedd ein harbrefion cynnar gyda chorpora cyfochrog yn dangos y gellir echdynnu terminolegau sy'n benodol i'r parth yn awtomatig a'u mapio rhwng y Gymraeg a'r Saesneg. Gall swyddogaethau o'r fath wella perfformiad o ran cyfieithu peirianyddol ar gyfer testunau arbenigol fel y rhai a geir mewn gofal iechyd."

Noddwyd y gynhadledd gan EPSRC drwy rwydwaith 'Testun Gofal Iechyd y DU', y dinasoedd iechyd cysylltiedig, cronfa ddata SAIL a'r Sefydliad Ymchwil i Arloesi mewn Data ym Mhrifysgol Caerdydd, yn ogystal ag Averbis Text Analytics a DeepCongito. Trefnir y gynhadledd nesaf gan Kings College London yn 2020.

## Cynhadledd y Ganolfan Dysgu Cymraeg Genedlaethol, 11/5/19

Cynhaliwyd cynhadledd flynyddol y Ganolfan Dysgu Cymraeg Genedlaethol ym Mae Caerdydd ar 11 Mai, ac roedd Enlli Thomas a Jenny Needs yno yn rhedeg gweithdai gyda thiworiaid Cymraeg i Oedolion. Diolch yn fawr iawn i'r Ganolfan am y gwahoddiad! Dyma oedd y cyfle cyntaf i aelodau PG4 gael arddangos y gwaith sydd wedi'i wneud hyd yma ar y pecyn cymorth pedagogaidd, ac i dderbyn adborth fydd yn arwain gweddill y gwaith datblygu.

Pleser oedd cael gweld ymateb gadarnhaol y tiwtoriaid at yr offer a'r posibilïadau mae'r corpws yn eu



cynnig i diwtoriaid a dysgwyr.

Mae gan y corpws y potensial i drawsnewid y ffordd mae dysgwyr yn dod i adnabod patrymau a geirfa'r Gymraeg, drwy gynnig data dilys a chaniatáu i ddysgwyr ddarganfod atebion i'w cwestiynau eu hunain. "Oes treigladau ar ôl *rhai*?" ... "Pa arddodiad sy'n dod ar ôl *tueddo*?" ... os ydych chi'n gofyn cwestiynau o'r fath, mae'r corpws i chithau hefyd!

Rydyn ni'n ddiolchgar y tu hwnt i'r tiwtoriaid a fynychodd ein gweithdai am yr holl adborth adeiladol a roddon nhw, ynghyd â'u syniadau o ran darpar ddatblygiadau ar gyfer y dyfodol. Mae'n hanfodol bod y corpws a'r offer pedagogaidd cysylltiedig yn



cwrdd ag anghenion a disgwyliadau'r defnyddwyr targed. Rydym ni eisiau creu adnoddau sydd wir yn cael eu defnyddio ac sy'n dod yn rhan annatod o wersi Cymraeg ledled y wlad.

Dim ond un o'n grwpiau targed yw tiwtoriaid a dysgwyr Cymraeg i Oedolion. Dros y misoedd nesaf, byddwn ni hefyd yn cwrdd ag athrawon ysgolion cynradd ac uwchradd – cyfrwng Cymraeg a Saesneg – er mwyn sicrhau bod rhanddeiliaid ym mhob maes yn cael y cyfle i ddylanwadu ar luniad yr offer terfynol. Edrychwn ymlaen at gwrdd â chi!

Yn y cyfamser, hoffem ni ofyn am eich adborth ar yr enw arfaethedig ar gyfer y pecyn cymorth pedagogaidd – *y Tiwtiadur*. Beth dych chi'n feddwl? Anfonwch ebost at [corcencc@caerdydd.ac.uk](mailto:corcencc@caerdydd.ac.uk) i fynegi'ch barn.

## ***Y Tiwtiadur***

### + Gwaith maes i lenwi'r bylchau

Er mwyn llunio corpws cytbwys, mae tîm WP1 wedi ymdrechu'n ddibaid dros 3 mlynedd y prosiect i gasglu data o bob congl o Gymru. Mae cyfle nawr i'r tîm lenwi'r bylchau a chanolbwyntio ar yr ardaloedd rheini sydd heb gyrraedd y nod eto, a siroedd Dinbych a Flint oedd yn gyntaf ar y rhestr!

Roedd yr haul yn tywynnu ar Rhuthun a'r Mart ddydd Iau a llawer o drigolion ac ymwelwyr y dref farchnad yn fodlon cyfrannu'n hael at y prosiect wrth sgwrsio â ffrindiau hen a newydd. Hoffem ddiolch yn benodol i Siop Elfair ar Stryd Clwyd am fod gymaint o gymorth ac am gynnig paned ar adeg strategol! Roedd croeso cynnes i'w gael yn y gornel hon o Sir

Ddinbych a digonedd o eiriau i'w ychwanegu at y targed. Ymlaen wedyn i ddydd Gwener yn Ninbych lle roedd yr un brwdfrydedd i'w weld mewn rhan arall o'r sir! Casglwyd data di-ri gyda chymorth y grŵp Iaith a Chwarae yn enwedig. Diolch i bawb a gyfrannodd yn barod yn Ninbych.

Yn yr Wyddgrug ar ddydd Gwener mae'r dref yn llawn bwrlwm yr ocsiwn, ac mewn caffi cyfagos y cafwyd y rhan fwyaf o ddata Sir Fflint y prynhawn hwnnw. Er hynny, roedd cynrychiolaeth gref o amaethwyr Sir Ddinbych yno hefyd! Yna, diolch i Siop y Siswrn, y Mold Alehouse, a'r Saith Seren (Wrecsam) am eu cydweithrediad a'u cwmni am weddill y prynhawn. Gadawyd sawl hysbyseb i'r prosiect yn y tafarndai hyn, felly edrychwch allan amdanynt ar eich ymweliad nesaf! Ffordd hawdd o gyfrannu i'r project: Yn aml, mae negeseuon testun yn adlewyrchu'r iaith lafar yn agos iawn – efallai mai dyma'r iaith ysgrifenedig lleiaf ffurfiol a mwyaf cyfoes.

Mae CorCenCC yn dal i gasglu negeseuon testun drwy WhatsApp, felly danfonwch neges yn dweud "helo!" at +447542348512 ar WhatsApp am gyfarwyddiadau syml ar sut i ddewis a chyfrannu eich negeseuon testun Cymraeg at y prosiect!



Fela dwi ynde. D powb arall ru fath? 😊



## + Ymgyrch ddata

Wrth i brosiect CorCenCC ddirwyn i ben, dyma apêl arall am ddata sydd eu hangen arnon ni o hyd. Aeth yr apêl yma i fwletin mis Mehefin Cymdeithas Cyfieithwyr Cymru ond roeddem am ei rhannu gyda holl ddarllenwyr ein cylchlythyr ni hefyd. Apêl bellach yw hon am ddeunydd gwahanol sydd ei angen arnom er mwyn sicrhau bod y corpws yn cynrychioli pob math o Gymraeg cyfoes a'n helpu i gyrraedd ein targed o 10 miliwn. Y deunydd y mae gennym ddiddordeb arbennig ynddo yw (gan gofio, wrth gwrs, y caiff unrhyw gyfraniadau eu hanonymeiddio'n llwyr):

- Llythyron anffurfiol. Mae'r rhain wrth reswm yn wahanol i negeseuon testun ac ebyst (gw. isod) ond mae dod o hyd i'r math yma o ddata yn ddigon heriol. Pe bai gennych unrhyw lythyron (cyfoes) o'r fath yr ydych chi'n fodlon eu rhannu gyda ni, byddai'n gyfraniad gwerthfawr i'r 4 miliwn o eiriau ysgrifenedig rydym wrthi'n eu casglu ar hyn o bryd.
- Hysbysebion / Arwyddion. Mae'n bosib eich bod wedi cael eich comisiynu i gyfieithu hysbysebion neu arwyddion penodol. Byddai cael gwybod ble gallen ni wneud cais am gynnwys y rhain (heb dorri cyfrinachedd wrth gwrs) yn hynod ddefnyddiol neu efallai eich bod yn gwybod am hysbysebion / arwyddion Cymraeg (heb eu cyfieithu) y byddai modd i ni eu defnyddio.
- Ebyst / Negeseuon testun. Rydym wedi derbyn nifer o ebyst proffesiynol a phersonol i'w cynnwys yn y 2 filiwn o eiriau digidol yn CorCenCC... ond mae bob amser lle i fwy! Os ydych chi'n fodlon rhannu unrhyw ebyst (ffurfiol neu anffurfiol), bydden ni wrth ein boddau yn eu derbyn. Yn yr un modd, negeseuon testun neu WhatsApp. Gallwch gyfrannu trwy decstio "helo" bach drwy WhatsApp i +44 7542 348512 . Wedyn, cewch chi ddolen i'r ffurflen ganiatâd a chyfarwyddyd pellach.
- Dyddiaduron. Unwaith eto, mae dod o hyd i'r math yma o ddata yn heriol. Ydych chi'n cadw dyddiadur bob dydd, neu pan fyddwch chi ar eich gwyliau, er enghraifft? Pe bai gennych unrhyw ddyddiaduron (cyfoes) yr ydych chi'n fodlon eu rhannu gyda ni, bydden ni'n ddiolchgar iawn.

Os ydych chi'n medru anfon deunyddiau (neu gyfrannu trwy'r ap neu ystyried gwneud gwaith trawsgrifio), anfonwch air atom [trwycorcencc@caerdydd.ac.uk](mailto:trwycorcencc@caerdydd.ac.uk) neu ewch i'n gwefan [www.corcencc.cymru](http://www.corcencc.cymru).



## + Cwrdd â'r tîm. 1

### Yr Athro Margaret Deuchar – ymgynghorydd prosiect CorCenCC



Ar ôl cwblhau B.A. mewn Ieithoedd Modern (Ffrangeg ac Almaeneg) yng Nghaergrawnt a PhD (ar Iaith Arwyddion Prydain) ym Mhrifysgol Stanford yng Nghalifornia ym 1978, fy swydd gyntaf oedd darlithyddiaeth ym Mhrifysgol Caerhirfryn. Yno, dysgais sosioieithyddiaeth yn bennaf, a pharheais â'm hymchwil ar iaith arwyddion, gan gwblhau llyfr o'r enw British Sign Language (a gyhoeddwyd ym 1984). Ar ôl symud i Brifysgol Sussex, dechreuais ymddiddori mewn caffael dwyieithog ieithoedd llafar, a chefais fy ngrat ESRC cyntaf i gynnal astudiaeth achos o fy merch yn caffael Saesneg a Sbaeneg ar yr un pryd. Arweiniodd hyn at lyfr a ysgrifennwyd ar y cyd â Suzanne Quay, sef fy myfyriwr PhD yn fy ngweithle nesaf, sef Prifysgol Caergrawnt.

Yn ystod y broses o ysgrifennu'r llyfr, cefais gynnig swydd ym Mhrifysgol Bangor, a chyrrhaeddais yno ym 1994, gan ddod ar draws y Gymraeg am y tro cyntaf. Fel myfyriwr ieithoedd modern gyda ffocws ar y Cyfandir wrth astudio Ffrangeg ac Almaeneg, ro'n i'n ystyried y DU yn achos diobaith o unieithrwydd ar y cyfan. Ond cefais sioc pan gyrhaeddais ym Mangor ychydig cyn Eisteddfod 1994 yn Abergele: yn y digwyddiad hwn, rhyfeddais i weld bod yr holl arwyddion, cyhoeddiadau a digwyddiadau yn y Gymraeg yn unig. Ym Mangor, cofrestrais yn llawn brwdfrydedd ar gyfer dosbarthiadau Cymraeg, gan barhau nes i mi basio fy lefel A yn 2000. I baratoi ar gyfer hyn, gwirfoddolais mewn ysgol gynradd cyfrwng Cymraeg ym Mae Colwyn, ac ymunais â'r gangen leol o Ferched y Wawr, lle cefais fy ethol yn Gadeirydd yn nes ymlaen, gydag aelod caredig yn paratoi sgript ar gyfer yr hyn yr oedd angen i fi ei ddweud mewn cyfarfodydd.

Ar y pryd, roedd yr holl addysgu ym Mangor yn fy adran innau (Ieithyddiaeth) yn Saesneg, ond roedd grantiau ar gael ar gyfer modiwlau cyfrwng Cymraeg newydd. Llwyddais i ennill un o'r rhain, ac roedd modd i fi gynnig modiwlau newydd i israddedigion mewn Ieithyddiaeth Gymraeg, trwy gyflogi darlithydd cyfrwng Cymraeg allanol. Roedd modd i fyfyrwyr ddewis gwneud eu gwaith yn Saesneg neu'r Gymraeg eisoes, a chyflwynodd myfyriwr MA newydd her i fi trwy ofyn a allwn gyfathrebu â hi'n gyfan gwbl trwy'r Gymraeg. Yn fuan wedi hynny, cefais fy ngrat cyntaf i weithio ar gyfathrebu dwyieithog Cymraeg/Saesneg, a recriwtiais gynorthwydd ymchwil oedd yn siarad Cymraeg. Gyda'i hanogaeth hi, cytunais gynnal ein holl drafodaethau yn y Gymraeg ac, yn raddol, dysgais y derminoleg ieithyddol briodol wrth i ni drafod fframweithiau damcaniaethol a chywirdeb trawsgrifiadau.

Yn 2005, recriwtiais ddau fyfyrwr PhD, gyda chymorth grant AHRC, i weithio ar gyfnewid codau rhwng y Gymraeg a'r Saesneg, a daeth un o'r myfyrwyr hyn (Peredur Webb-Davies) yn ddarlithydd ieithyddiaeth cyfrwng Cymraeg parhaol. Yn 2007, roedd modd i fi, gyda'm cydweithwyr (gan gynnwys Enlli Thomas), sefydlu Canolfan ESRC dros Ymchwil i Ddwieithrwydd, gyda grant sylweddol gan yr ESRC. Yn ogystal â chyfarwyddo'r Ganolfan, cefais gyfle i deithio i Batagonia yn 2009 i gasglu data am y Gymraeg/Sbaeneg - profiad hynod ddiddorol. Mae'r data hwn, ynghyd â data Cymraeg/Saesneg a Sbaeneg/Saesneg, bellach ar gael ar [bangor.org.uk](http://bangor.org.uk), a chyhoeddwyd llyfr a ysgrifennwyd ar y cyd â Peredur Webb-Davies a Kevin Donnelly am y corpws Siarad Cymraeg/Saesneg yn 2018. Yn bennaf oherwydd fy ngwaith gyda chorpws Siarad, manteisiais ar y cyfle i weithredu fel Ymgynghorydd ar y prosiect CorCenCC ar ôl i Dawn gysylltu â fi yn 2012.

Ers 2013, rwyf wedi bod yng Nghaergrawnt. Rwyf hyd yn oed ymhellach o Gaerdydd nag yr oeddwn i ym Mangor, a dyw cydweithio o bell ddim wedi bod yn hawdd. Fodd bynnag, rwyf wedi mynychu pob cyfarfod posibl CorCenCC, ac rwyf wedi fy mhlesio gan yr amrywiaeth o arbenigedd sydd gan aelodau'r tîm. Edrychaf ymlaen at barhau fy nghysylltiad â'r prosiect, ac rwy'n dymuno pob llwyddiant iddo yn y dyfodol.

## + Cwrdd â'r tîm.2

## Kevin Donnelly – ymgyngorydd prosiect CorCenCC



Rwy'n gweithio ar offer ac adnoddau meddalwedd rhydd ar gyfer y Gymraeg ac ieithoedd eraill ers 2003, gan gynnwys fersiwn Cymraeg o'r bwrdd gwaith GNU/Linux mwyaf blaenllaw, system cyfieithu peirianyddol Cymraeg-Saesneg, a chorpora Cymraeg ([cymraeg.org.uk](http://cymraeg.org.uk)). Mae meddalwedd rhydd ([fsf.org](http://fsf.org)) yn golygu bod gan ddefnyddwyr meddalwedd y rhyddid i redeg, copïo, dosbarthu, astudio, newid a gwella'r meddalwedd. Yn Saesneg defnyddir y gair 'Free' sydd â mwy nag un ystyr – rhydd ac am ddim (er bod y rhan fwyaf o feddalwedd rhydd ar gael am ddim hefyd) ([gnu.org](http://gnu.org)). Yn y Gymraeg gallwn ni wahaniaethu'n glir rhwng ystyrion 'free' yn Saesneg, y gair 'rhydd' sy'n golygu nad yw rhywbeth yn gaeth i rywbeth neu i rywun arall a'r ymadrodd 'am ddim' sy'n golygu nad oes rhaid talu am rywbeth. Mae'r rhan fwyaf o'r rhyngwrwyd yn rhedeg ar feddalwedd rhydd (e.e. y system enwau parth), ac mae llawer o'r cwmnïau technolegol mwyaf wedi'u hadeiladu ar feddalwedd rhydd, ac yn wir yn ei chreu (IBM, Google, Amazon, Facebook).

Rwyf wedi credu erioed mai trwydded rydd yw'r unig ddewis rhesymegol i ieithoedd lleiafrifol ar gyfer offer ac adnoddau, gan ei bod yn caniatáu aildefnyddio a datblygu - nid oes angen ailddyfeisio'r olwyn ar gyfer pob prosiect newydd, ac nid yw colli "cof cymunedol" yn digwydd hyd yn oed os daw prosiect i ben oherwydd diffyg cyllid neu resymau eraill. Yn anffodus, serch hynny, mae bron yn ddieithriad yn wir nad yw cyrff sy'n gweithio ar ieithoedd lleiafrifol yn rhannu offer ac adnoddau'n ddirwystr – maent yn blaenoriaethu cyllid tymor byr neu statws dros les tymor hwy i'r iaith. Mae'r mater hwn yn arbennig o bwysig o gofio'r cynnydd yn nifer y prif ieithoedd megis Saesneg, Sbaeneg a Tsieinëeg yn y byd digidol, sy'n ei gwneud yn anos i ieithoedd lleiafrifol greu eu lle eu hunain yma. Y cwestiwn y bydd eraill ond yn gallu ei ateb wrth edrych yn ôl yn 2050 yw: a yw'r Gymraeg wedi colli ei chyfle?

Pan oedd Dawn yn gweithio ar gael CorCenCC ar y gweill, ac wedi cysylltu â fi i ddweud y byddai'r prosiect yn defnyddio trwydded agored, dyna oedd y trobwynt i mi, ac roeddwn yn naturiol yn awyddus i gyfrannu beth bynnag a allwn i'r adnodd agored newydd mawr hwn. Fel y digwyddodd, bues i'n gweithio gyda Steven Neale ac eraill ar yr is-brosiect tagio, a ddefnyddiodd ddau o'm prosiectau blaenorol.

Un o'm prif brosiectau Cymraeg yw Eurfa ([eurfa.org.uk](http://eurfa.org.uk)) y geiriadur Cymraeg cyntaf (a hyd yn hyn yr yn unig un) sydd ar gael o dan y drwydded feddalwedd rhydd, y drwydded gyhoeddus gyffredinol ([gnu.org/licenses/quick-guide-gplv3.html](http://gnu.org/licenses/quick-guide-gplv3.html)). Defnyddiwyd hwnnw ar gyfer sawl prosiect NLP, a dylai'r fersiwn nesaf, a fydd yn cyrraedd yn ddiweddarach eleni gobeithio, olygu ehangu'r geiriadur yn sylweddol.

Bydd corpora CorCenCC ar gael o dan drwydded agored, ond torrwyd y cwys cyntaf yma gan gorpws Siarad, a gasglwyd ym Mangor gan yr Athro Margaret Deuchar a'i thîm, a ddefnyddiodd y GPL. Yn y diwedd, fe wnes i weithio gyda Margaret ar system awtomataidd i losio'u corpora Patagonia (Cymraeg-Sbaeneg) a Miami (Sbaeneg-Saesneg) ([bangortalk.org.uk](http://bangortalk.org.uk)), ac yna yn ôl-ddefnyddio hyn i drin Siarad – defnyddiwyd y fersiwn diweddaraf o hwn ([autoglosser.org.uk](http://autoglosser.org.uk)) fel model ar gyfer CyTag Steven, un o allbynnau CorCenCC. Mewn oes o NLP sy'n seiliedig ar ystadegau, mae Autoglosser/CyTag yn defnyddio dull sy'n seiliedig ar reolau, ac rwy'n tueddi i feddwl ei fod yn opsiwn gwell (er yn fwy ymestynnol) ar gyfer ieithoedd lleiafrifol, nad oes ganddynt lawer o destun o bosibl i ddeillio model ystadegol cadarn ohono.

Yn ddiweddar rwyf wedi bod yn treulio llai o amser ar y Gymraeg, ac yn mynd yn ôl at fy nyddiau fel myfyriwr mewn SOA gyda dau brosiect. Y cyntaf yw Andika!, teclyn i gysodi barddoniaeth Swahili clasurol mewn lythrennau Arabeg,

([kevindonnely.org.uk/swahili](http://kevindonnely.org.uk/swahili)). Gall y testun gynnwys trawslythreniad agos a safonol i lythrennau Rhufeinig, darlleniadau eraill, nodiadau, ac yn y blaen, ac mae'n caniatáu i destun o wahanol fersiynau llawysgrif o'r un gerdd gael ei haenu er mwyn cymharu geiriau (e.e. [kevindonnely.org.uk/swahili/jaafari](http://kevindonnely.org.uk/swahili/jaafari)).

Yr ail yw Qiezi, teclyn i ddarllen testun mewn Tsieinëeg a'i gysodi mewn nifer o fformatau gwahanol at ddibenion addysgu neu ddysgu ([kevindonnely.org.uk/qiezi](http://kevindonnely.org.uk/qiezi)). Mae'r fersiwn rwy'n gweithio arni ar hyn o bryd (nad yw wedi'i chyhoeddi eto) yn rhoi llawer o wybodaeth am nodau unigol, gan gynnwys dull mnemonig i gynorthwyo ddysgu'r nod hwnnw a'i wahaniaethu oddi wrth eraill. Mae gwybod 3,000 o nodau yn caniatáu i chi ddarllen 99% o destun Tsieinëeg modern – hyd yn hyn dw i wedi creu dulliau mnemonig ar gyfer 1,700 o nodau, felly dw i ar y ffordd!

Er bod prosiect CorCenCC yn dirwyn i ben, fy ngobaith yw y bydd yn gadael set gadarn o offer ac adnoddau agored a fydd yn rhoi man cychwyn i waith NLP ar y Gymraeg (ac yn wir ar ieithoedd eraill, gan y gellir porthi'r offer a'r casgliad o adnoddau gweithdrefnau yn ôl yr angen). Mae wedi bod yn llwyddiant mawr y gall tîm prosiect CorCenCC (Dawn, Tess a Steve M) ymfalchïo ynddo.

## + Cysylltu â ni

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/corcenc> (gallwch ein trydar @CorCenCC). Gallwch hefyd gysylltu â ni drwy anfon neges i gyfeiriad ebost y prosiect: [corcenc@caerdydd.ac.uk](mailto:corcenc@caerdydd.ac.uk) neu ewch i'n gwefan, sef:

[www.corcenc.cymru](http://www.corcenc.cymru)



Arts & Humanities  
Research Council

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-Ymchwilydd** - Tess Fitzpatrick, Steve Morris, Irena Spasić, Paul Rayson, Enlli Thomas, Alex Lovell a Jonathan Morris; y **Cynorthwyydd Ymchwil** - Jennifer Needs, Ignatius Ezeani a Laura Arman; y **myfyrwyr PhD** - Vigneshwaran Muralidaran a Bethan Tovey; **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](mailto:KnightD5@caerdydd.ac.uk)