Cylchlythyr CorCenCC

Corpws Cenedlaethol Cymraeg Cyfoes
National Corpus of Contemporary Welsh

Cynnwys

T1: Cyfarchion (a ffarwél) y Prif Ymchwilydd
T2: Newyddion
T4: Diweddariadau a myfyrdodau'r Pecynnau Gwaith
T13: Cwrdd â'r tîm
T15: Cysylltwch â ni

Cyfarchion (a ffarwel) y PrifYmchwilydd

Dyma ni – cylchlythyr olaf prosiect CorCenCC, sy'n nodi diwedd prif gyfnod cyllido'r gwaith. Mae taith 3 ½ blynedd cyntaf y prosiect bellach ar ben ond, fel soniwyd yn y rhifyn blaenorol, mae gyda ni tan ddiwedd Mai 2020 er mwyn mynd â'r maen i'r wal a chwblhau'r corpws. Wedyn, y cynllun yw rhyddhau CorCenCC ar ddiwedd y cyfnod hwn.

Felly, er mai hwn yw'r cylchlythyr ffurfiol olaf, byddwn yn dal i anfon y newyddion diweddaraf atoch trwy Drydar (@CorCenCC), Facebook (https://www.facebook.com/CorCenCC/) a thrwy'r wefan (www.corcencc.cymru) felly cofiwch ein dilyn o hyd a gobeithio eich gweld mewn sioe deithiol/ digwyddiad prosiect agos yn fuan! Ar ran tîm CorCenCC, hoffwn ddweud diolch enfawr i bawb sydd wedi ein cefnogi ar hyd y daith – mae wedi bod yn bleser mawr cwrdd â chi a gweithio gyda chi, pob un ohonoch chi. O gyd-ymchwilwyr i ymgynghorwyr y prosiect; cymdeithion ymchwil i fyfyrwyr PhD, aelodau bwrdd ymgynghorol y prosiect i'r ymchwilwyr is-raddedig; llysgenhadon i'r miloedd o bobl sydd wedi cyfrannu eu data, mae eich amser a'ch cefnogaeth parhaus wedi bod yn gwbl ryfeddol. Fydden ni erioed wedi cyrraedd ble rydyn ni nawr oni bai am eich help chi. Yn y rhifyn olaf yma o'r cylchlythyr, cewch newyddion a diweddariadau cyffredinol o'r prosiect, yn ogystal ag adolygiadau a myfyrdodau pob un o'r pecynnau gwaith (gan arweinwyr ein pecynnau gwaith). Ar ben hynny, cewch y cyfle i gwrdd â dau aelod arall o deulu estynedig CorCenCC yn ein colofn 'cwrdd â'r tîm' misol. Mwynhewch y darllen a chofiwch gadw mewn cysylltiad!

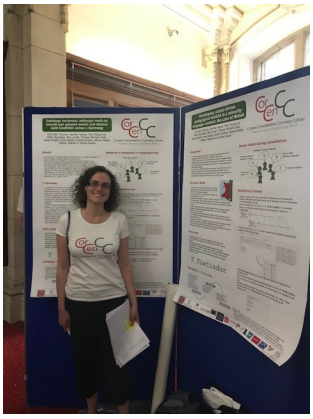
CorCenCC mewn rhifau:

- Derbyniwyd £1.8 miliwn gan yr ESRC/AHRC
- 42 mis o waith (2015-2019, gydag estyniad 9 mis di-gost ar fin cychwyn)
- 4 sefydliad academaidd (1 Prif Ymchwilydd, 7 Cyd-Ymchwilydd, 5 Cynorthwyydd Ymchwil, 2 fyfyriwr PhD, 4 lleoliad haf is-raddedig, 2 wirfoddolwr, 4 aelod staff rhan amser – 3 chyn aelod)
- 6 ymgynghorydd, 12 Aelod o Grŵp Ymgynghorol y Prosiect, 4 Llysgennad
- 10 cyfarfod prosiect ffurfiol, 100edd o gyfarfodydd ychwanegol
- 7 rhestr ohebu, 24 cylchlythyr, niferoedd poenus o uchel o ebyst…
- 14 prif araith a thros 30 o gyflwyniadau mewn 11 o wledydd
- 6 chyhoeddiad hyd yn hyn
- 1500+ o gyfranwyr, tua 100,000 o ymweliadau gwefan (ar draws www.corcencc.cymru ac www.corcencc.org), miloedd o 'hoffi' ac ail-drydariadau

+ Newyddion

CorCenCC @ CL2019, Caerdydd (22 – 26 Gorffennaf)

Cynhaliwyd y 10fed Gynhadledd Ieithyddiaeth Gorpws Ryngwladol (CL2019: www.cl2019.org), a drefnwyd gan Brif Ymchwilydd y prosiect Dawn, ym Mhrifysgol Caerdydd ynghanol tywydd crasboeth mis Gorffennaf. Ymysg y 400 o ieithyddion corpws a oedd yn bresennol roedd aelodau amrywiol o dîm CorCenCC, a gyflwynodd 1 gweithdy, 2 boster a 3 phapur wedi'u seilio ar y prosiect - record mewn cynadleddau! Gwnaeth y cyflwyniadau adrodd a myfyrio ar y datblygiadau hyd yma a rhoi gwybodaeth am gynlluniau a rhyddau'r corpws ar gyfer y dyfodol. Cafodd y tîm sylwadau ac adborth cefnogol iawn. Cafodd pawb amser gwych!

CorCenCC @ACL2019, Fflorens, Yr Eidal

Oedd, roedd y tywydd yn hyfryd, ond yn rhy boeth i rai pobl. Cafwyd llawer o hwyl, gwelwyd llawer o olygfeydd a mwynhawyd digon o fwyd a danteithion blasus. Dyma **57ain Cyfarfod Cyffredinol Blynyddol y Gymdeithas Ieithyddiaeth Gyfrifiadol (ACL2019).** Cafodd ei gynnal o **28ain Gorffennaf - 28il Awst** yn y **Fortezza da Basso** hanesyddol yn **Fflorens, yr Eidal**. Roedd y lle'n fôr o syniadau a dulliau newydd a modern o ran prosesu iaith naturiol a ieithyddiaeth gorpws.





Cyflwynodd tîm CorCenCC, a gynrychiolwyd yn wych gan Paul ac Ignatius, ein papur ar y dull arloesol o adeiladu tagwyr rhannau ymadrodd a semantig Cymraeg gan ddefnyddio modelau ymwreiddio geiriau Cymraeg a raglennwyd ymlaen llaw mewn sefyllfa ddysgu aml-dasg. Mae ein dull newydd yn defnyddio gwybodaeth ieithyddol ddyfnach sydd wedi'i hymwreiddio yn yr iaith wrth aseinio categorïau rhan ymadrodd a semantig gan ddefnyddio rhwydweithiau niwral dwfn. Cyflwynwyd y papur, sy'n dwyn y teitl "***Leveraging Pre-Trained Embeddings for Welsh Taggers***", yn y **4ydd Gweithdy ar Ddysgu Cynrychioladol ar gyfer NLP** ac mae wedi'i gyhoeddi yn ACL Ontology.

Gan Ignatius Ezeani

CorCenCC @Yr Eisteddfod Genedlaethol, Dyffryn Conwy

Er gwaetha gwynt a glaw Dyffryn Conwy, llwyddiant ysgubol oedd yr Eisteddfod Genedlaethol ar gychwyn mis Awst. Hoffem ddiolch a llongyfarch trefnwyr yr Eisteddfod a hefyd tîm Cyfathrebu Prifysgol Caerdydd am gynnal digwyddiad mor egnïol.

Cafodd Steve Morris a Laura Arman gyfle i gyflwyno ar gynnydd ein gwaith yn ogystal ag ar wahanol agweddau ar yr offer corpws gorffenedig yn y Cymdeithasau ac ym Mhabell Caerdydd.

Roedd pobl o bob cefndir proffesiynol wedi dod i glywed am y prosiect un ai am y tro cyntaf neu er mwyn cadw i fyny gyda'r datblygiadau diweddaraf. Mae sgyrsiau difyr yn dal i ddigwydd o gylch y math o ddata sy'n cael eu cynnwys yn y corpws a sut y gall y data rhain fod o ddefnydd i wahanol rhanddeiliaid a'r cyhoedd.

Ym Mhabell Caerdydd roedd cyfle i gyflwyno manylion am ddyluniad y prosiect a'r data sydd wedi ei gasglu. Er mai bechan oedd y gynulleidfa, roedd diddordeb gan gyfranwyr newydd. Ar yr ail ddiwrnod o gyflwyno, y pecyn cymorth pedagogaidd gafodd sylw, ac roedd llawer o ddiddordeb brwd gan diwtoriaid ac athrawon yn Y Tiwtiadur a'i systemau o greu ymarferion addas i ddysgwyr o bob lefel yn awtomatig. Fel mae mwy o ganlyniadau ar gael, bydd y prosiect yn cyflwyno mewn gwahanol leoedd fel bod gair yn cyrraedd digonedd o bobl o holl fanteision y Corpws i ddefnyddwyr.

+ Diweddariadau a myfyrdodau'r Pecynnau Gwaith

Pecyn Gwaith 1 (Steve Morris):

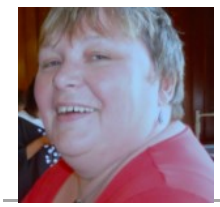
Ble i ddechrau?! Mae'n wir fod amser yn hedfan ac o edrych yn ôl dros y tair blynedd a hanner diwethaf o safbwynt Pecyn Gwaith 1 (WP1), mae'n gwbl ryfeddol fod ein tîm bach bendigedig o dri Chy-northwyydd Ymchwil (ac un ohonynt yn rhan amser) wedi llwyddo i gyflawni cymaint mewn cyfnod o amser mor fyr. Mae'n hawdd anghofio, pan ddechreuon ni yn ôl ym mis Mawrth 2016, nad oedd fframwaith samplo gyda ni, doedd dim confensiynau trawsgrifio, dim ffurflenni moeseg, roedd llwyth o ewyllys da gan ein rhanddeiliaid/ partneriaid ond dim cynllun clir o ran casglu data ac roedd pawb yn newydd i fyd rhyfedd ieithyddiaeth gorpws (diolch byth fod MOOC Prifysgol Caerhirfryn wedi cael ei lansio tua'r un pryd!). O'r tîm gwreiddiol, mae CY Abertawe Jenny Needs (dde)

wedi bod gyda'r prosiect ers y cychwyn cyntaf. Yma, mae hi'n rhannu ei myfyrdodau am weithio ar WP1: *'"Mae rhan sylweddol o'm gwaith ar WP1 wedi ymwneud â recordio iaith lafar ar draws Cymru, ac wrth edrych yn ôl ar fy "nodiadau maes", dw i'n cofio pa mor ddiddorol oedd y teithiau gwahanol. Dw i wedi gyrru dros 3,000 o filltiroedd, er mwyn ymweld ag 16 allan o 22 awdurdod lleol Cymru a chwrdd â dros 1,000 o siaradwyr Cymraeg! Y peth cynta hoffwn i ei ddweud yw pa mor ddiolchgar ydw i i bob un person a gytunodd i gyfrannu enghreifftiau o'r ffordd maen nhw'n siarad at y corpws. Dyw cael eich recordio ddim yn rhywbeth sy'n apelio at lawer o bobl, ond dw i'n falch iawn i chi weld pwysigrwydd yr adnodd rydyn ni'n ei adeiladu a dewis bod yn rhan ohono. Fyddai'r corpws ddim yn bodoli hebddoch chi! Dw i wedi cael y fraint o glustfeinio ar amrywiaeth anhygoel o bethau sy'n digwydd drwy'r Gymraeg, gan gynnwys grwpiau trafod barddoniaeth a nofelau, cystadlaethau siarad cyhoeddus, beirniadaethau sioeau ceffylau a defaid, a chyflwyniadau cyhoeddus ar bob math o bynciau. Y peth gorau yn fy marn i, fodd bynnag, yw bod y gwaith maes wedi rhoi'r cyfle i fi fod yn dyst i gannoedd ohonoch chi yn siarad y Gymraeg yn naturiol ledled y wlad, yn yr ysgol neu'r gweithle, wrth siopa yn y dre, wrth gymdeithasu gyda ffrindiau mewn caffis a thafarndai, wrth fwynhau digwyddiadau megis Eisteddfodau a gwyliau cerddoriaeth a bwyd, ac wrth wneud tasgau pob dydd o gwmpas y tŷ gyda'ch teuluoedd. Daliwch ati! Mae hi wir yn ffantastig y bydd y corpws yn dangos pa mor amlwg yw'r defnydd o'r Gymraeg ym mhob elfen o fywydau eu siaradwyr. Diolch yn fawr iawn i chi i gyd am rannu â ni fewnwelediad i'r ffordd mae'r Gymraeg yn rhan o'ch bywydau chi. Rhywbeth arall sydd wedi bod yn galonogol y tu hwnt, wrth i fi deithio'r wlad, oedd yr ymateb gan bobl sy'n dysgu Cymraeg. Roedd hi'n bwysig iawn i ni fod y corpws yn adlewyrchu'r ffaith bod siaradwyr newydd yn rhan annatod o frithwaith y byd Cymraeg. Er gwaetha'r ffaith bod rhai ohonoch chi'n teimlo'n ddi-hyder, roeddech chi'n deall sut byddai dysgwyr yn elwa o'r corpws, ac felly roedd llawer ohonoch chi'n fodlon cyfrannu at y prosiect er budd dysgwyr y dyfodol. Diolch yn fawr iawn i chi! Hoffwn ddiolch yn fawr iawn hefyd i bawb sydd (wedi bod) yn rhan o dîm trawsgrifio CorCenCC. Mae'r gwaith yn gallu bod yn heriol ar adegau, dw i'n gwybod, ond gobeithio eich bod chithau hefyd yn/wedi mwynhau clywed enghreifftiau o'r holl ffyrdd gwahanol mae'r Gymraeg yn cael ei defnyddio yng Nghymru heddiw. Heb eich gwaith caled chi, fyddai'r holl enghreifftiau hyn ddim yn cael eu hychwanegu at y corpws, felly dych chi'n hollbwysig i lwyddiant y prosiect! (Dim pwysau!)"*









Mae llawer o'r hyn mae Jenny wedi'i ddweud yma yn cael ei ategu gan dîm cyfan WP1. Yn gyntaf, dylem ddweud DIOLCH enfawr i'n holl bartneriaid, pencampwyr, cyfranwyr, cyfranogwyr, trawsgrifwyr ac i gyfeillion y Gymraeg. Afraid dweud, oni bai am eich cefnogaeth chi, fyddai 'na ddim CorCenCC. Gosodon ni'r her i'n hunain o gasglu canran o ddata o bob awdurdod lleol yng Nghymru yn unol â nifer y siaradwyr Cymraeg ym mhob un yn ôl Cyfrifiad 2011. Ymrwymiad y Cynorthwywyr Ymchwil a sut maen nhw wedi ymgysylltu â'r cymunedau lleol hyn sy'n cyfrif am ba mor agos

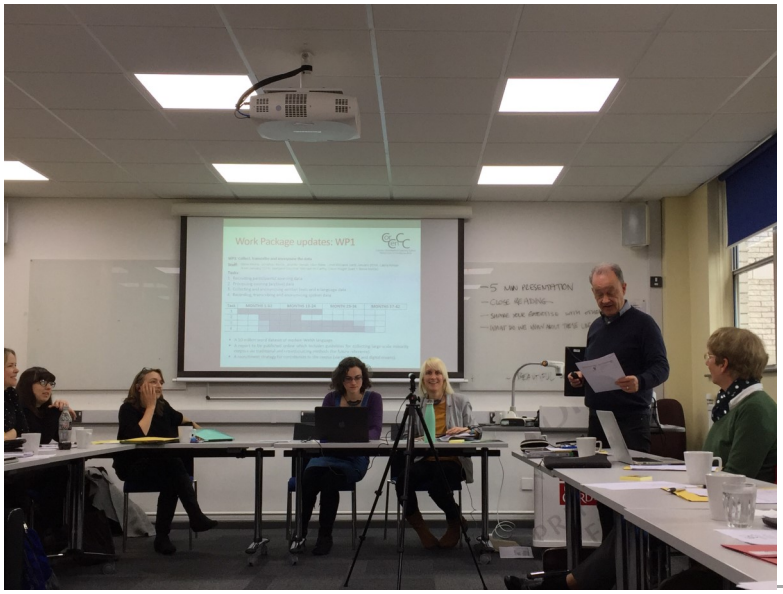
rydym ni wedi dod at wireddu hyn.

Yn ogystal â Jenny, rydym wedi bod yn ffodus i gael grŵp brwdfrydig a dawnus ar WP1 yn Abertawe a Chaerdydd, gan gynnwys Mair Rees yn Abertawe a Gareth Watkins, ac wedyn Lowri Williams ac yn olaf Laura Arman yng Nghaerdydd. Yn sicr, gallech chithau ddweud storïau tebyg i rai Jenny am eich gwaith gyda'r pecyn gwaith yma a byddai'n anodd iawn dewis uchafbwyntiau penodol. Mae'r llun yma o Mair a'i chreadigaeth 'Cor-pws' yn nodweddiadol o'r arloesi a'r hyblygrwydd oedd bob amser mor amlwg yn y tîm. Daeth 'Cor-pws' i fodolaeth fel ffordd o egluro a hwyluso'r taflenni caniatâd sy'n cael eu defnyddio gyda'n cyfranogwyr iau. Ymddangosodd hi hefyd yng nghynhadledd CL2017 ym Mhrifysgol Birmingham – gwnawn yn siŵr ei bod yn mynd i gartre da pan ddaw'r prosiect i ben!

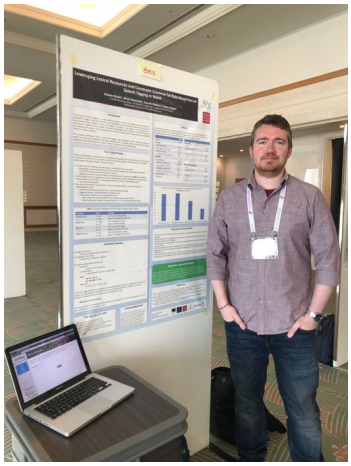
Erbyn hyn, rydym mewn sefyllfa gadarn ar ôl casglu 90% o'r data llafar, bron i 85% o'r data ysgrifenedig a 190% o'r data iaith electronig. Wrth gwrs, mae'na heriau o hyd i'w hwynebu cyn Mai 2020:

- Mae angen i ni ychwanegu at y data llafar ac ysgrifenedig (mae cynlluniau ar y gweill ar gyfer y rhain);
- Mae angen i ni weithio hyd eithaf ein gallu i sicrhau bod yr holl waith trawsgrifio / rheoli ansawdd yn cael ei gwblhau – **os ydych yn nabod unrhyw un a all fod â diddordeb mewn helpu ac ennill 'bach o arian ychwanegol, rhowch wybod i ni**;
- Mae angen i ni gwblhau'r cynlluniau ar gyfer 'cartre terfynol' CorCenCC a sut y byddwn ni'n ei gynnal yn y dyfodol.

Yng nghynhadledd ddiweddar CL2019 yng Nghaerdydd, roedden ni'n gallu siarad am y gwaith rydyn ni wedi'i wneud hyd yn hyn fel Pecyn Gwaith 1 yn ogystal â dechrau dangos sut mae CorCenCC yn gweithio i'r gynulleidfa oedd yno. Yn dilyn hyn, cynhaliwn sesiynau yn yr Eisteddfod Genedlaethol yn Llanrwst ym mis Awst ac mae cynlluniau ar gyfer sioeau teithiol a sesiynau ymgysylltu hyd at (a siŵr o fod ar ôl) Mai 2020. Felly, nid y diwedd mo hwn. Fodd bynnag, mae'n amser priodol i adfyfyrio ar yr hyn a gyflawnwyd, ei gofio a'i wireddu. Mae wedi bod yn fraint enfawr i arwain y tîm arbennig yma a'ch gwaith chi sydd wrth galon CorCenCC. Rwy'n credu ei bod yn deg dweud hefyd ein bod wedi cael cyfle i chwerthin a rhannu digon o hiwmor ar hyd y daith... ond efallai dylen ni gadw rhai o'r storïau hynny ar gyfer rhifyn arall o'r cylchlythyr!!!



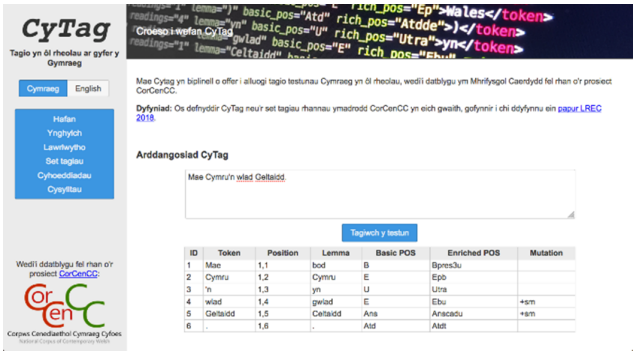
Pecyn Gwaith 2 (Dawn Knight)



Roedd y gwaith ar WP2 yn cynnwys y Cydymaith Ymchwil Steven Neale a'r ymgynghorydd prosiect Kevin Donnelly yn bennaf (dan arweiniad PI Dawn Knight). Gofynnwyd i'r tîm:

- Adeiladu a hyfforddi tagiwr Cymraeg
- Datblygu set o dagiau priodol ar gyfer yr iaith Gymraeg
- Tagio'r holl ddata yn CorCenCC

Gwaith hyd yma: Mae meddalwedd pwrpasol CorCenCC ar gyfer tagio rhannau ymadrodd (POS), *CyTag*, ar waith bellach ac mae wrthi'n ennill ei blwyf. Mae *CyTag* yn liferu deunyddiau ffynhonnell agored i helpu gyda'r broses o benderfynu ar rannau ymadrodd. Mae'n gweithio yn bennaf trwy ddefnyddio'r wybodaeth sydd yng ngeiriadur Kevin Donnelly, *Eurfa* – y geiriadur ffynhonnell agored mwyaf sydd ar gael i'r Gymraeg yn rhad ac am ddim – i baratoi rhestr o'r tagiau posibl i bob gair mewn testun Cymraeg. Ategir hynny trwy restrau penodol o enwau lleoedd, enwau cyntaf a chyfenwau o Wikipedia. Ar ôl paratoi rhestr o'r geiriau sy'n bosibl, gallwn ni ddefnyddio set o reolau pwrpasol i fireinio'r tagiau posibl i bob gair – yn ôl tagiau neu nodweddion y geiriau o'u cwmpas – nes dod o hyd i'r un cywir. Er enghraifft, yn y frawddeg 'mae Cymru yn wlad Geltaidd', gallwn ni dybio bod 'yn' yn elfen sy'n cysylltu 'Cymru' â 'wlad' am ein bod yn gwybod bod 'wlad' yn 'gwlad' wedi'i dreiglo'n feddal ac mae gennym ni reol i ddewis y tag ar gyfer 'yn' lle mae'r gair canlynol yn enw sydd wedi'i dreiglo'n feddal. Mae CyTag wedi'i gwerthuso gan ddefnyddio set safon aur o 611 o frawddegau (14,876 o docynnau), wedi'u hanodi â llaw â thagiau rhannau ymadrodd ar gyfer y Gymraeg. Roedd canlyniadau'r broses werthuso o gywirdeb cymharol â'r hyn a ddisgwylir gan dagwyr rhannau ymadrodd mewn ieithoedd eraill (dros 95%). Lansiwyd y wefan ar gyfer CyTag ym mis Mawrth 2018. Cyfeiriad y wefan yw http://cytag.corcencc.org, lle gallwch ddod i wybod mwy am y tagiau a rhoi cynnig ar ein rhaglen arddangos ar-lein! Ar hyn o bryd, mae CyTag yn cynnwys rhannwr testun, holltwr brawddegau, tocynnwr a'r tagiwr POS ei hun.



At hynny, mae set tagiau rhannau ymadrodd (POS) CorCenCC wedi'i hen gwblhau ac yn cynnwys 145 o dagiau 'gwerthfawr' ar draws 13 o gategorïau 'sylfaenol' sy'n cydymffurfio ag EAGLES. Mae'r set lawn o dagiau ar gael ar-lein yma: https://cytag.corcencc.org/tagset?lang=cy

Cafodd papur gan Steven Neale, Kevin Donnelly, Gareth Watkins a Dawn Knight sy'n disgrifio CyTag a set tagiau CorCenCC ei dderbyn i'w gyhoeddi yn y gynhadledd Adnoddau a Gwerthuso Ieithoedd (LREC) ym Miyazaki, Japan yn ôl ym mis Mai 2018. Roedd yn cyflwyno trosolwg technegol trylwyr o CyTag a gwerthusiad manwl o'i chywirdeb. I gael rhagor o wybodaeth am y cyhoeddiad hwn, ewch i: https://cytag.corcencc.org/publications?lang=cy **Tasgau'r dyfodol:** Tasg derfynol tîm WP2 fydd tagio pob un o 10 miliwn o eiriau'r corpws CorCenCC, unwaith y bydd y set ddata wedi'i chwblhau gan dîm WP1. Mwy yn y man!

Pecyn Gwaith 3 (Paul Rayson)

Ym mhecyn gwaith 3 sydd wedi'i arwain gan Paul Rayson, Scott Piao ac Ignatius Ezeani (Prifysgol Caerhirfryn), ein prif nod oedd llunio meddalwedd a dulliau ar gyfer dadansoddi ystyron testunau Cymraeg yn awtomatig. Mae hynny'n ychwanegu at gynnyrch pecyn gwaith 2 am fod tagio rhannau o leferydd o gymorth mawr ynghylch diamwyso testunau. Defnyddir y tagiwr semantaidd i dagio holl gasgliad CorCenCC cyn diwedd y prosiect a bydd y tagiau semantaidd yn helpu i chwilio'r casgliad o ran astudio ieithyddol a thrin a thrafod yr offer addysgol at ddibenion dysgu. Mae'r offer anodi sydd wedi'u creu yn CorCenCC ar gael i bawb, maen nhw wedi'u cynnwys yn sustem cymharu ac anodi casgliadau Wmatrix (http://ucrel.lancs.ac.uk/wmatrix/), ac mae modd eu defnyddio ar gyfer dod o hyd i destunau Cymraeg eraill.

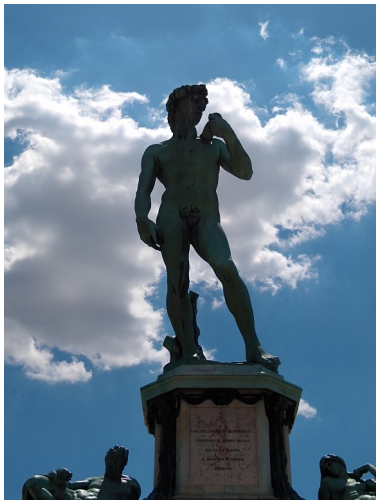
Ein gorchwyl cyntaf oedd llunio set y tagiau semantaidd ar gyfer y Gymraeg, a gwnaethon ni hynny gyda chymorth ein partneriaid yn Abertawe a Chaerdydd. Man cychwyn y gwaith oedd sustem USAS Caerhirfryn (http://ucrel.lancs.ac.uk/usas/), sy'n cynnwys categorïau semantaidd ar ffurf tagiau megis I1.3 (Arian: Prisiau/Money: Price), K5.1 (Chwaraeon/Sports), ar gyfer geiriau, priod-ddulliau ac ymadroddion Cymraeg eraill (a elwir yn fynegiadau aml eu geiriau, hefyd) yn ôl cynllun dosbarthu semantaidd. Mae yn y cynllun hwnnw 232 o gategorïau semantaidd o dan 21 prif gategori fel sydd wedi'i ddangos isod:

Tag	Definition	Tag	Definition
A	*TERMAU CYFFREDINOL A HANIAETHOL* (GENERAL AND ABSTRACT TERMS)	N	RHIFAU A MESUR (NUMBERS AND MEASUREMENT)
B	*B Y CORFF A'R UNIGOLYN* (THE BODY & THE INDIVIDUAL)	O	SYLWEDDAU, DEFNYDDIAU, GWRTHRYCHAU AC OFFER (SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT)
C	*CELF A CHREFFT* (ARTS AND CRAFTS)	P	ADDYSG (EDUCATION)
E	*GWEITHREDIADAU, CYFLYRAU A PHROSESAU EMOSIYNOL* (EMOTIONAL ACTIONS, STATES & PROCESSES)	Q	GWEITHREDIADAU, CYFLYRAU A PHROSESAU IEITHYDDOL (LINGUISTIC ACTIONS, STATES AND PROCESSES)
F	*BWYD A FFERMIO* (FOOD & FARMING)	S	GWEITHREDIADAU, CYFLYRAU A PHROSESAU CYMDEITHASOL (SOCIAL ACTIONS, STATES AND PROCESSES)
G	*LLYWODRAETH A'R CYHOEDD* (GOVENMENT AND THE PUBLIC DOMAI)	T	AMSER (TIME)
H	*PENSAERNÏAETH, ADEILADAU, TAI A'R CARTREF* (ARCHITECTURE, BUILDINGS, HOUSES & THE HOME)	W	Y BYD A'N HAMGYLCHEDD (THE WORLD AND OUR ENVIRONMENT)
I	*ARIAN A MASNACH* (MONEY & COMMERCE)	X	GWEITHREDIADAU, CYFLYRAU A PHROSESAU SEICOLEGOL (PSYCHOLOGICAL ACTIONS, STATES AND PROCESSES)
K	*ADLONIANT, CHWARAEON A GEMAU* (ENTERTAINMENT, SPORTS AND GAMES)	Y	GWYDDONIAETH A THECHNOLEG (SCIENCE AND TECHNOLOGY)
L	*BYWYD A PHETHAU BYW* (LIFE AND LIVING THINGS)	Z	ENWAU A GEIRIAU GRAMADEGOL (NAMES AND GRAMMATICAL WORDS)
M	*SYMUD, LLEOLIAD, TEITHIO A CHLUDIANT* (MOVEMENT, LOCATION, TRAVEL AND TRANSPORT)		

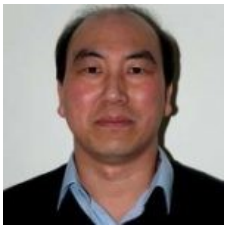
Rydyn ni wedi defnyddio amryw ffyrdd o ddysgu cyfrifiadur i 'ddeall' testunau Cymraeg (gorchwyl eithaf anodd) gan gynnwys paratoi â'n dwylo a'n meddalwedd eiriaduron y gall cyfrifiadur eu darllen gan ddidoli pob gair ac ymadrodd yn ôl set tagiau USAS.

Cylchlythyr CorCenCC

At hynny, rydyn ni wedi troi at ffynonellau torfol lle y gofynnwyd i leygwyr ddidoli geiriau, a defnyddio dulliau dysgu peiriannol i hyfforddi meddalwedd yn ôl rhai cannoedd o frawddegau prawf uchel eu hansawdd oedd wedi'u gwirio gan arbenigwyr Cymraeg eu mamiaith sy'n ymwneud â'r prosiect. Mae'n anodd iawn llunio offeryn anodi semantaidd manwl gywir ond rydyn ni wedi cyflawni llawer yn ystod tair blynedd a hanner prosiect CorCenCC. Mae dros 143,000 o nodion yn ein geiriadur semantaidd Cymraeg ac rydyn ni wedi trafod dros 91% o destunau Cymraeg cyfredol. Mae cywirdeb ein hoffer dysgu peiriannol tua 94-5% ynghylch dewis tagiau yn ôl cyd-destun.



Ein bwriad yw parhau i gynnal ymchwil i ddulliau tagio semantaidd gwell a defnyddio ffyrdd traws ieithyddol o ehangu'r sustem, yn arbennig ar gyfer ymadroddion aml eu geiriau, sy'n rhan bwysig o'r dadansoddi. At hynny, byddwn ni'n ceisio defnyddio'r offer awtomatig i ddadansoddi mathau eraill o ddata Cymraeg. Rydyn ni wedi dangos a chyflwyno ein hymchwil yn rheolaidd drwy gydol y prosiect gan gynnwys yn ysgolion haf *Natural Language Processing* UCREL ac mewn cynadleddau megis:



- *Association for Computational Linguistics (ACL)*, Gorffennaf 2019, Fflorens, yr Eidal.
- *Corpus Linguistics Conference 2019*, Mehefin 2019, Prifysgol Caerdydd.
- *LREC (Language Resources and Evaluation) 2018 Conference*, Mai 2018, Miyasaki, Siapan.

- *European Chapter of the Association for Computational Linguistics 2017 (EACL) Conference*, Ebrill 2017, Falensia, Sbaen.
- *1st International Conference on Corpus Analysis in Academic Discourse (CAAD)*, Falensia, Sbaen.
- *Corpus Linguistics Conference 2017*, Gorffennaf 2017, Prifysgol Birmingham.
- *LREC (Language Resources and Evaluation) 2016 Conference*, Mai 2016, Slofenia.

Pecyn Gwaith 4 (Enlli Thomas)

Amcanion cyffredinol WP4

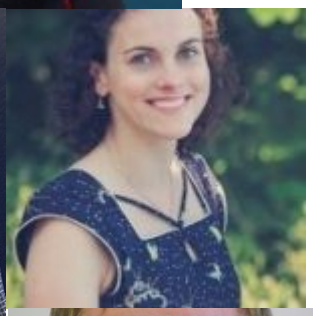
Un maes lle gall corpora fod yn arbennig o ddefnyddiol yw ym maes dysgu / addysgu iaith. Gellir defnyddio corpora i dynnu sylw at y geiriau, ymadroddion a phatrymau mwyaf cyffredin mewn iaith. Gallant ddangos pa eiriau sy'n tueddu i fynd gyda'i gilydd, a pha rai sy'n digwydd ym mha fathau o destun (e.e. testunau ysgrifenedig ffurfiol, sgyrsiau llafar, e-byst proffesiynol, neu negeseuon testun personol). Gall defnyddwyr corpws chwilio am eiriau penodol a'u gweld mewn brawddegau enghreifftiol. Felly mae data corpws yn darparu ffynhonnell gyfoethog o iaith i'r dysgwr sy'n dangos sut mae ieithoedd yn cael eu defnyddio'n ymarferol, mewn gwahanol feysydd a bydd defnyddio CorCenCC yn yr ystafell ddosbarth iaith yn helpu i ddangos sut mae'r Gymraeg yn cael ei defnyddio mewn gwirionedd.

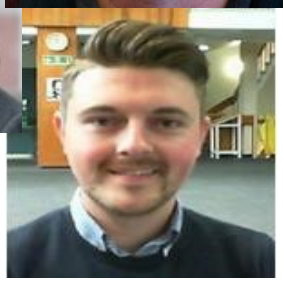
Un agwedd arloesol ar brosiect CorCenCC, dan arweiniad tîm WP4, fu datblygu offeryn pwrpasol - a elwir yn Y Tiwtiadur - i'w ddefnyddio o fewn a thu allan i ddosbarthiadau Cymraeg, o'r ysgol gynradd i addysg oedolion, sy'n darparu ymarferion a gweithgareddau'n seiliedig ar gorpws y gellir eu cynllunio gan yr athro iaith neu'r disgybl.

Datblygu Y Tiwtiadur

Yn dilyn archwiliad cychwynnol o'r mathau o offer ar-lein a oedd yn bodoli eisoes, gan gynnwys y rhai sydd ar gael i ddysgwyr y Gymraeg er mwyn osgoi dyblygu, roedd datblygiad Y Tiwtiadur ar y gweill. Cawsom ein hysbrydoli'n arbennig gan waith un o ymgynghorwyr CorCenCC - yr Athro Tom Cobb - sef y LexTutor (https://lextutor.ca/), ac aethom ati i gynllunio ar gyfer datblygu teclyn tebyg ar gyfer y Gymraeg. Roedd y datblygiad hwn yn cynnwys tri phrif gam: cam un - ymgynghori; cam dau - datblygu cynnyrch; a cham tri - y cyfnod arddangos. Yn ystod y cam ymgynghori, treialwyd holiadur gyda nifer fechan o ymarferwyr ym maes dysgu Cymraeg i archwilio pa adnoddau y mae dysgwyr ac athrawon eisoes yn eu defnyddio a beth yn ddelfrydol yr hoffent ei weld yn cael ei ddatblygu fel rhan o becyn cymorth addysgol CorCenCC. Yn seiliedig ar eu hadborth, datblygwyd holiadur newydd ar gyfer cynulleidfa fwy. Dosbarthwyd hwn i athrawon a thiwtoriaid Cymru mewn dwy gynhadledd, a'i rannu'n ddiweddarach fel fersiwn ar-lein.

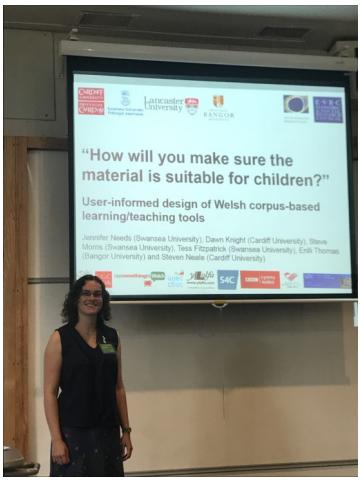
Cylchlythyr CorCenCC





Yn ogystal â'r holiadur, gwnaethom gwrdd ag athrawon a thiwtoriaid wyneb yn wyneb, i godi ymwybyddiaeth o'r corpws a'r pecyn cymorth addysgol, ac i barhau i gasglu barn a fyddai'n help i lywio datblygiad pellach y pecyn cymorth. Fe wnaeth yr holiadur - ynghyd â grwpiau ffocws dilynol - ein helpu i nodi'r blaenoriaethau ar gyfer pecyn cymorth addysgol CorCenCC. Roedd y trafodaethau a gawsom yn hynod ddefnyddiol, ac roedd y cyfnewid syniadau a ddigwyddodd yn ystod y grwpiau ffocws a'r adborth a gafwyd trwy'r holiaduron yn gwella ein meddwl am sut i lywio'r gwaith wrth inni symud ymlaen. At ei gilydd, dychwelwyd 44 holiadur gan athrawon, hyfforddwyr, darlithwyr a thiwtoriaid o amrywiaeth eang o gyd-destunau - o'r rhai sy'n dysgu mewn ysgolion cynradd (cyfrwng Cymraeg a Saesneg) i'r rhai sy'n dysgu oedolion - a chnyhaliwyd 10 grŵp ffocws, gan gyrchu barn 55 o athrawon / tiwtoriaid a 14 o ddysgwyr Cymraeg i Oedolion.

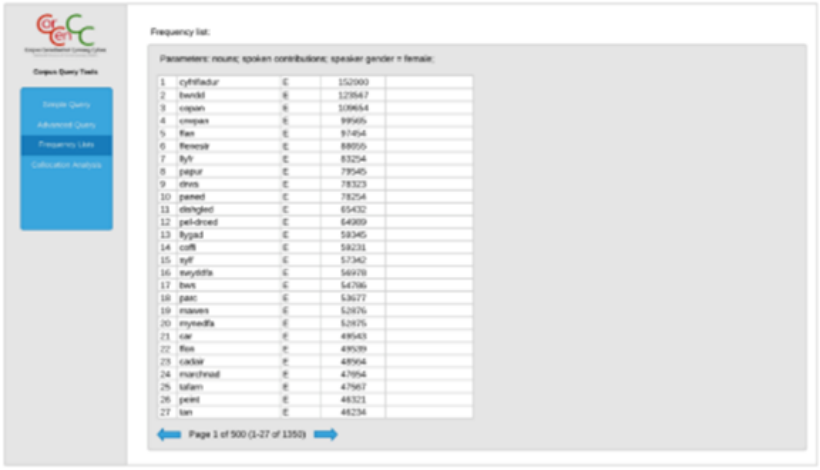
Yn dilyn yr ymgynghoriad, buom yn gweithio gydag aelodau o dîm WP5, Anelia Kurteva, myfyriwr yn Ysgol Cyfrifiadureg a Gwybodeg Prifysgol Caerdydd dan oruchwyliaeth yr Athro Irena Spasić, i ddechrau gweithio ar brototeipiau offeryn, yn seiliedig ar yr adborth terfynol yn ystod yr enghreifftiol o'r ymgynghoriad. Datblygwyd y gwaith hwn ymhellach gan Joshua Davies, dan oruchwyliaeth Bill Teahan, Prifysgol Bangor, yn barod i'w arddangos yng nghynhadledd flynyddol y Ganolfan Genedlaethol ar gyfer Dysgu Cymraeg a gynhaliwyd ym Mae Caerdydd ar 11 Mai, 2019. Hwn oedd y cyfle cyntaf i aelodau WP4 allu dangos y gwaith a gwblhawyd hyd yma ar y pecyn cymorth pedagogaidd, a chael derbyn adborth a fyddai'n arwain gweddill y gwaith datblygu.

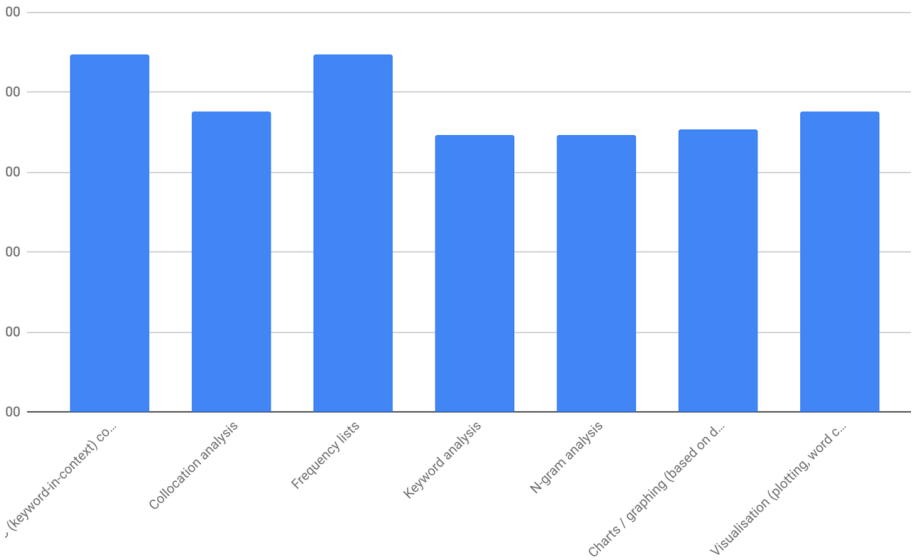


Cynlluniau terfynol: Dim ond un o'n grwpiau targed yw tiwtoriaid a dysgwyr Cymraeg i Oedolion. Dros yr ychydig fisoedd nesaf, byddwn hefyd yn cwrdd ag athrawon ysgolion cynradd ac uwchradd - cyfrwng Cymraeg a chyfrwng Saesneg - er mwyn sicrhau bod rhanddeiliaid ym mhob maes yn cael cyfle i ddylanwadu ar greu'r offer terfynol. Byddwn hefyd yn cysylltu'n uniongyrchol â swyddogion Llywodraeth Cymru sy'n gyfrifol am ddatblygu Cwricwlwm newydd Cymru er mwyn sicrhau hygrededd y corpws ac Y Tiwtiadur fel dull posibl o wella ymgysylltiad a darganfod unigrywiaeth y Gymraeg ymhlith siaradwyr L1 a L2. Ein bwriadau dros yr ychydig fisoedd nesaf fydd i (i) cwblhau'r cynnyrch prototeip, (ii) arddangos y prototpye terfynol ar gyfer adborth, (iii) diwygio'r cynnyrch yn unol â'r adborth, (iv) cwblhau llyfr canllaw ar-lein sy'n esbonio'r rhinweddau pedagogaidd ymarferion sy'n cael eu gyrru gan gorpws, tasgau enghreifftiol ac ymarferion enghreifftiol i'r athro a'r dysgwr.

Pecyn Gwaith 5 (Irena Spasić)

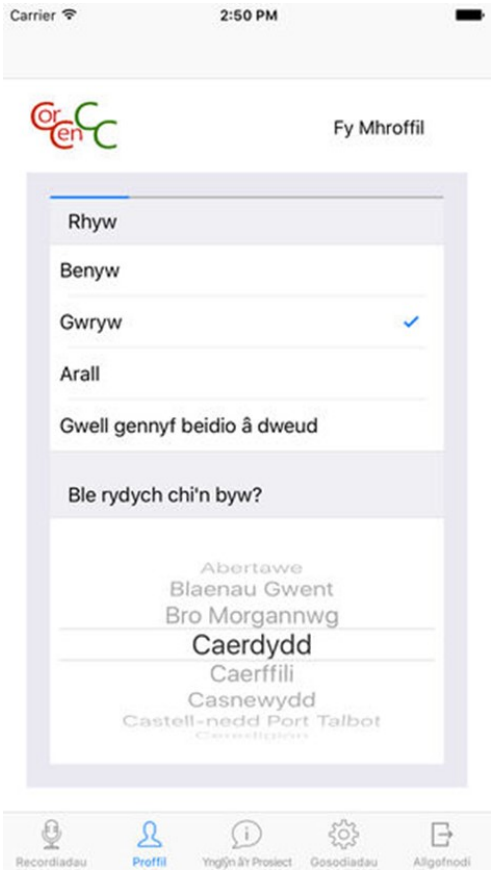
I roi'r testun ar gael i bawb, llunion ni ryngwyneb pen blaen arbennig fel y byddai'n hawdd i ddefnyddwyr archwilio data Cymraeg mewn modd rhyngweithiol. Rydyn ni'n disgwyl y bydd defnyddwyr o sawl lliw a llun megis arbenigwyr ieithyddol, dysgwyr, athrawon, cyhoeddwyr ac unrhyw un arall a chanddo ddiddordeb yn y Gymraeg. I gynnig gwasanaeth cyson ac effeithlon y gallai pawb ei ddefnyddio'n hawdd beth bynnag fo ei offer cyfrifiadurol, dewison ni ryngwyneb sy'n seiliedig ar y we. Yn hynny o beth, rydyn ni'n defnyddio dull sawl cwmni arall yn y brif ffrwd megis *BNCweb*, *Corpus of Contemporary American English (COCA)*, y *Michigan Corpus of Academic Spoken English (MICASE)* a'r *English Native Edited Japanese Essays (ENEJE)*. Er bod rhai ffyrdd cyfredol o gyhoeddi testunau o'r fath ar y we – megis *CQPweb* – penderfynon ni lunio dull a fyddai'n diwallu ein hanghenion ni am ddau reswm pwysig. Y naill oedd y dylai ei swyddogaethau weddu i'r Gymraeg yn sgil modelu ei chystrawennau a'i semanteg trwy rannau ymadrodd yr iaith a thagiau semantaidd fel ei gilydd. Yn benodol, roedden ni am alluogi pobl i chwilio ynglŷn â'r treigladau, un o nodweddion yr iaith lle y bydd cytsain gyntaf gair yn newid neu'n diflannu yn ôl cyflwr gramadegol y gair o'i flaen. Ar ben hynny, i geisio manteisio i'r eithaf ar y data ar gyfer dysgu'r iaith, roedd angen dull modiwlaidd y gallen ni ei gyfuno'n hawdd â'r pecyn cymorth addysgol sydd wedi'i lunio yn rhan o'r prosiect hwn.





Wrth lunio'r nodweddion a'r swyddogaethau, rydyn ni wedi cymryd i ystyriaeth ganlyniadau arolwg o'r hyn sydd orau gan ddefnyddwyr o safbwynt ieithyddol. Roedden ni am hel sylwadau ynglŷn â'r arferion gorau fel y gallen ni gydymffurfio â nhw a lleddfu anawsterau dysgu'r iaith. Cyn bwysiced â hynny, roedden ni am adnabod y meysydd a allai elwa trwy arloesi. Roedd yr ymatebion i'r arolwg o gymorth ynglŷn â'n helpu i ganolbwyntio ar y nodweddion pwysicaf ar gyfer cam cyntaf y datblygu ac maen nhw wedi bod yn ddefnyddiol iawn ers hynny o ran mireinio ac ehangu'r swyddogaethau cyntaf.

Cylchlythyr CorCenCC



Elfen bwysig arall o'n penderfyniad ni i lunio ein rhyngwyneb ein hunain oedd bod angen galluogi pobl i hidlo pob chwiliad yn ôl metadata CorCenCC. Mae CorCenCC wedi hel data yn ôl egwyddorion penodol gan ddefnyddio, ar y cyfan, fframwaith samplo sy'n rhoi ehangder o gyd-destunau, mathau, ffynonellau a lleoliadau. Felly, roedd yn hanfodol cynnig i ddefnyddwyr allu i hidlo data yn ôl y fframwaith hwnnw fel y byddai modd astudio tafodieithoedd a'u nodweddion. Yn ogystal â chasglu data cyfredol, rydyn ni wedi annog pobl i hel data llafar. I'w helpu i hel data llafar, llunion ni raglen ffôn poced sy'n galluogi Cymry Cymraeg i recordio eu sgyrsiau, ysgrifennu nodiadau amdanynt a'u rhoi i'r casgliad yn ysbryd gwyddoniaeth dinasyddion.

Cynhalion ni gyfres o astudiaethau i werthuso holl agweddau ein rhyngwyneb a pharhau i'w gwella yn ymateb i anghenion defnyddwyr a datblygiadau technolegol. Gyda chymorth Llywodraeth Cymru, rydyn ni'n estyn swyddogaethau'r casgliad mewn ffyrdd nas rhagwelwyd cynt. Ar hyn o bryd, rydyn ni'n ystyried ffordd newydd o astudio'r Gymraeg, sef llywio trwy ofod aml ei dimensiynau lle y cynrychiolir geiriau gan rifolion mewn modd sy'n adlewyrchu perthynas semantaidd neu ramadegol geiriau cyfatebol â'i gilydd. Ymhlith prosiectau eraill sydd wedi deillio o hyn mae *Welsh WordNet* (cronfa ddata lle mae geiriau wedi'u didoli'n gyfystyron yn ôl eu perthynas semantaidd), *Welsh FlexiTerm* (adnabod ymadroddion aml eu geiriau yn syth: https://github.com/ispasic/FlexiTermCymraeg) a *Welsh Stemmer* (cyfleu ffurfiau gwreiddiol geiriau sydd wedi newid trwy dreiglo ac ati: https://users.cs.cf.ac.uk/I.Spasic/wncy).

Mae'r llwyddiant hwn wedi dod trwy ymdrechion y tîm. Ymhlith aelodau allweddol y tîm mae'r Dr Steven Neale (cyfrannog ymchwil ar ôl doethuriaeth), yr Athro Laurence Anthony (ymgynghorydd), yr Athro Irena Spasić (cydymchwilydd) a'r Athro Dawn Knight (prif ymchwilydd). Rydyn ni wedi ymgysylltu â tho newydd o lunwyr trwy gyfraniadau myfyrwyr megis Corey Watson, Anelia Kurteva, David Owen a Vigneshwaran Muralidaran. Mae'n hymdrechion i lunio meddalwedd gynaladwy wedi'u hatgyfnerthu gan beirianwyr meddalwedd ymchwil y *Data Innovation Research Institute*, Ian Harvey a'r Dr Jeffrey Morgan. Mae'r Dr Padraig Corcoran a'r Dr Geraint Palmer wedi ymuno â ni yn ddiweddar i helpu i astudio iaith o safbwynt mathemategol. Yn olaf, ond nid yn lleiaf, hoffen ni ddiolch i'r cyhoedd am roi o'u hamser i gyfrannu at lwyddiant y prosiect trwy ffynonellau torfol.




+ Cwrdd â'r tîm.1

Alex Lovell, Cyd-Ymchwilydd, Adran y Gymraeg, Prifysgol Abertawe

Ers dyddiau'r ysgol y mae diddordeb gen i mewn ieithoedd a'r iaith Gymraeg yn benodol. Ar ôl astudio Lefel A mewn Cymraeg Ail Iaith yn yr ysgol, penderfynais i barhau â'm hastudiaethau yn y Gymraeg ym Mhrifysgol Abertawe yn 2010. A dydw i heb adael Abertawe ers hynny! Ar ôl cwblhau gradd BA yn y Gymraeg yn 2014, dechreuais i radd MA trwy ymchwil a chafodd y radd hon ei huwchraddio i Ph.D. yn 2015. Roedd fy Ph.D. yn archwilio sut orau y gellir cefnogi cyflwyno'r Gymraeg fel ail iaith yn llwyddiannus yng nghyd-destun ysgolion uwchradd cyfrwng Saesneg. Dyma faes ymchwil sydd yn agos iawn at fy nghalon felly mi oedd yn fraint rhoi rhywbeth nôl i'r maes hwn. Ym mis Medi 2017, ymunais i â staff addysgu Adran y Gymraeg Prifysgol Abertawe, ar ôl saith mlynedd o astudio yno. Rwy'n gyfrifol yn bennaf am addysgu ar nifer o fodiwlau iaith ar gyfer myfyrwyr ail iaith, ond rwy' hefyd yn cydlynu a chyfrannu at fodiwlau eraill ym maes iaith, addysg, ieithyddiaeth gymhwysol a chynllunio ieithyddol.

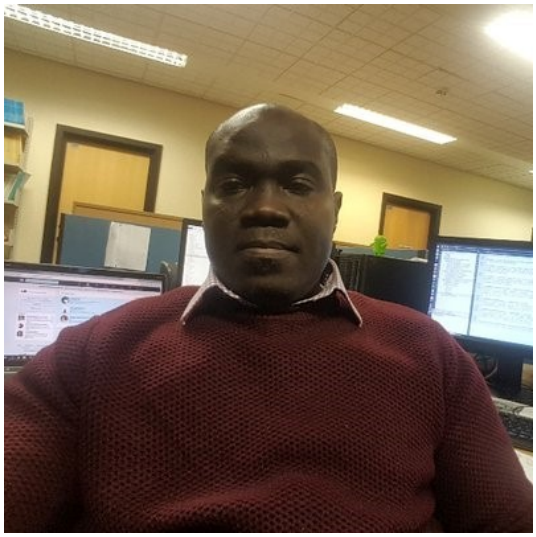
Fel pob athro iaith, rwy'n chwilio am adnoddau dysgu newydd trwy'r amser, nid yn unig i wella safon yr addysgu ond hefyd i wella profiad ac ansawdd y dysgu ar gyfer y myfyrwyr. A dyma pam fy mod wrth fy modd o fod yn rhan o dîm Pecyn Gwaith 4 – mae gan y corpws cenedlaethol hwn y potensial i weddnewid y ffordd rydym yn dysgu ac addysgu'r Gymraeg yn y dosbarth a'r tu hwnt. Trwy ddefnyddio'r offer pedagogaidd sydd yn cael eu datblygu ar hyn o bryd gan dîm Pecyn Gwaith 4, bydd modd i athrawon deilwra'r dysgu a'r asesu i weddu i anghenion eu dysgwyr. Yn ogystal â hyn, bydd modd i ddysgwyr archwilio'r corpws, gan eu hannog nhw i gymryd cyfrifoldeb am eu dysgu eu hunain ac i ddatblygu'n ddysgwyr annibynnol. Fel un sydd yn aml yn troi at y we a geiriaduron er mwyn dysgu sut mae geiriau newydd (i mi!) yn cael eu defnyddio mewn cyd-destunau penodol, rwy'n cyffroi wrth feddwl am y posibiliadau addysgol y bydd y corpws yn eu cynnig i mi a'm myfyrwyr yn y dyfodol.

+ Cwrdd â'r tîm.2

Ignatius Ezeani, Cydymaith Ymchwil, Prifysgol Caerhirfryn

Rwyf i'n newydd i Brosesu Iaith Naturiol (NLP) ac Ieithyddiaeth Gyfrifiannol (CL) Yn 2000, cefais fy ngradd gyntaf mewn Cyfrifiadureg o Brifysgol Nnamdi Azikiwe, Nigeria. Bryd hynny, doedd hi ddim yn fwriad gen i aros ym maes ymchwil na'r byd academaidd felly sefydlais gwmni cymorth TG bach yn nhref prifysgol Awka yn Nhalaith Anambra, Nigeria. Ond o fewn ychydig flynyddoedd i raddio cefais fy sugno'n ôl i'r brifysgol i ymgymryd â rôl Cynorthwyydd Graddedig.

Cylchlythyr CorCenCC



A dyna fi ar lwybr gyrfa academaidd yn addysgu modiwlau cyfrifiadureg a goruchwylio prosiectau myfyrwyr. Ar ôl dychwelyd i amgylchedd academaidd, cofrestrais ac ennill MSc mewn Peirianneg Meddalwedd Uwch o Brifysgol Bournemouth yn y DU yn 2006. Roedd rhaid i fi ddychwelyd i Nigeria ar ôl y rhaglen MSc i barhau â fy nghontract.
Dechreuodd fy nhaith gyda NLP gyda'r gwahoddiad gan gydweithiwr, Dr Ikechukwu Onyenwe, oedd ar y pryd yn fyfyriwr doethurol ym Mhrifysgol Sheffield. Roedd yn arloesi gydag ymchwil ar lunio adnoddau iaith (corpora ac offer) ar gyfer Igbo, iaith fawr yn Nigeria. Ei oruchwylwyr oedd Dr Mark Hepple oedd â diddordeb mewn ymchwil iaith adnodd isel ac roedd Igbo'n digwydd bod yn enghraifft dda. Roedd eu gwaith yn edrych ar gynllunio set data Igbo, gan lunio corpws â thagiau a datblygu tagiwr Igbo. Mae nifer dda o gyhoeddiadau ar wahanol agweddau ar y prosiect wedi'u cyhoeddi.

Yn 2014 ymunais i â'r prosiect, a dagiwyd yn ddiweddarach, IgboNLP. Roeddwn i'n edrych ar her cynbrosesu fwy cynnil ond eithaf problematig - 'adferiad deiacritig'. Sylwon ni oherwydd cynnwys deiacritig uchel Igbo, bod geiriau â deiacritigau coll yn dychwelyd i ffurf sylfaenol sy'n amwys heb gyd-destun. Gallai'r elfennau amwys hyn ymwneud ag ystyr yn unig, ond gallen nhw hefyd fynd y tu hwnt i ddosbarthiadau geiriau. Mabwysiadom ni ymagwedd seiliedig ar gorpws nad oedd angen unrhyw anodi dynol wrth lunio ein ffrwd arbrofol. Mae'r ffrwd yn caniatáu hyd at greu enghreifftiau hyfforddi o ddata crai, a chymhwyso un o'r dulliau adferol a gynigiwyd h.y. modelau ngram, dysgu peiriannol ac ymwreiddio. Gwnaethom ni nifer o gyflwyniadau mewn cynadleddau gyda'r gwaith hwn.

Ymunais i â phrosiect CorCenCC ym mis Hydref 2018 fel Cydymaith Ymchwil ar dîm WP3 yn gweithio ar ddatblygu'r tagiwr Semantig Cymraeg. Dysgais lawer o waith rhagorol Steve Neale a Scot Piao ar CyTag a CySemTagger sydd yn dagwyr seiliedig ar reolau ar gyfer tagio rhannau ymadrodd a semantig Cymraeg. Ar ôl gweithio ar fodelau ymwreiddio, roedd gen i ddiddordeb yn yr echdyniadau ystyr a chysylltiadau semantig hynny a geir drwy fodelau ymwreiddio dwfn. Ond fel y nododd Kevin Donnelly, mae angen paratoi llawer iawn o ddata a dyw hyn ddim yn aml ar gael ar gyfer ieithoedd fel y Gymraeg sy'n gwneud ymagweddau seiliedig ar reolau'n ddewis clir.

Cylchlythyr CorCenCC

Serch hynny rydym ni'n gwthio'r ffiniau ac yn canolbwyntio ar ddod o hyd i ffyrdd i ddefnyddio modelau a ragbaratowyd sy'n cael eu creu'n aml o ddata anstrwythuredig yn defnyddio modelau lled-oruchwyliedig. Rydym ni'n cydweithio gyda thîm WP2 ar baratoi aml-dagiwr niwrol dwfn seiliedig ar ymwreiddio a all gyflawni tagio rhannau ymadrodd a semantig ar yr un pryd. Mae'r gwaith yn mynd yn ei flaen, ond rydym ni wedi gwneud cynnydd sylweddol. Yn wir, caiff papur ei gyflwyno mewn gweithdy ACL2019. Rydym ni hefyd wedi llunio poster a byddwn yn paratoi i arddangos yr offeryn yng nghynhadledd CL2019. Mae ein llwyddiant hyd yma wedi ein hysgogi ac rydym ni'n llawn cyffro wrth edrych ymlaen at ragor o gyfraniadau a llwyddiannau ar brosiect CorCenCC.

+ Cysylltu â ni

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, Alex Lovell a Jonathan Morris; y **Cynorthwywyr 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