Corpws Cenedlaethol Cymraeg Cyfoes
National Corpus of Contemporary Welsh

Cyfarchion gan y Prif Ymchwilydd

Mae'r cennin Pedr yn eu blodau a'r gwanwyn wedi cyrraedd a thîm CorCenCC wrthi yn bwrw ati o hyd i gyflawni'r nodau rydym wedi bod yn gweithio'n galed arnynt ers 2015. Mae'r gwaith yn mynd rhagddo'n dda ac wrth i'r diwedd agosáu, mae ffrwyth ein llafur yn dechrau dod i'r fei, fel coeden ffrwythau lawn blodau ym MharcBute. Wrth i ni ddechrau rhyddhau offer ac adnoddau'r corpws, byddem wrth ein bodd yn derbyn adborth gennych. Gallai hyn gynnwys adfyfyrio ar ein taith hyd yn hyn, eich barn am ddyfodol CorCenCC, hynny yw, beth wnewch chi â'r corpws, sut gallai fe gael effaith ar eich gwaith/arfer ac yn anad dim, yr hyn y mae (y gallai) CorCenCC ei olygu i chi. Mynnwch gip ar ein gwefan a'n cyfrifon cyfryngau cymdeithasol am fwy o fanylion maes o law! Yn ôl i'r presennol am y tro: mae'r rhifyn yma'n rhoi'r newyddion diweddaraf am newidiadau yn y tîm, ambell ddigwyddiad diweddar gyda'r prosiect ac mae'n proffilio ymchil Vigneshwaran Muralidaran, myfyriwr PhD CorCenCC. Mwynhewch y darllen!

Dawn Knight

Cynnwys

T1: Newyddion
T3: Hybu CorCenCC
T3: Bywyd ar y tu mewn
T5: Cynllun ymchwil CorCenCC
T5: Cysylltwch â ni

+ Newyddion Ffarwelio…

Mae'n drist ffarwelio â dau aelod allweddol o dîm CorCenCC a'r ddau wedi bod gyda ni ers i'r prosiect ddechrau. Yn gyntaf, mae'r Cynorthwyydd Ymchwil Steve Neale, sydd wedi chwarae rhan bwysig ym mhob agwedd dechnegol ar y prosiect, gan gyfrannu at WP 2 a 5 yn bennaf (ond hefyd yn gweithio gyda WP 1, 3 a 4), newydd dderbyn swydd fel peiriannydd ymchwil AI yn DeepCognito ym Manceinion. Bydd Steve yn gweithio ar ddeallusrwydd artiffisial a dadansoddi testunau ac yn cychwyn ar ei rôl newydd ym mis Mai. Yn Abertawe, bydd y Cynorthwyydd Ymchwil Mair Rees, aelod o dîm WP1 a thrawsgrifydd ers dechrau'r prosiect, yn ein gadael ni hefyd ar ddiwedd mis Mawrth. Bydd hi'n gweithio fel cyfieithydd dan hyfforddiant yng Nghynulliad Cenedlaethol Cymru am y chwe mis cyntaf ac wedyn bydd yn gyfieithydd llawn amser gyda'r tîm yno. Gobeithiwn felly, mewn ffordd, y gall fod yn ymwneud â'r prosiect yn anuniongyrchol o hyd gan fod yr uned gyfieithu yn un o'n rhanddeiliaid. Bydd Mair yn cychwyn yn ei swydd newydd o fis Ebrill. Mae Steve a Mair wedi bod yn gaffaeliad mawr i'r prosiect a dymunwn bob lwc iddynt yn eu swyddi newydd! Rydym yn gobeithio y bydd y ddau'n dal i fod yn glwm wrth brosiect CorCenCC ym mha ffordd bynnag sy'n bosibl.

Croeso: Joshua Davies @ Prifysgol Bangor (Pecyn Gwaith 4)

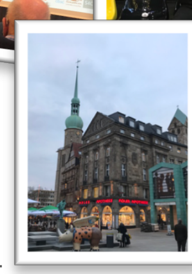
Mae Joshua Davies yn diwtor cyfrifiadureg yng Ngholeg Rhyngwladol Bangor ac yn fyfyriwr PhD sy'n astudio Prosesu Iaith Naturiol ym Mhrifysgol Bangor. Mae ei ymchwil yn ymwneud â modelau iaith wedi'u seilio ar gywasgiad a'u defnyddiau. Cyn cychwyn ar ei Ph.D. (a gyfarwyddir gan Dr Bill Teahan), cwblhaodd Joshua B.Sc. mewn Cyfrifiadureg ym Mhrifysgol Bangor. Ar ben hynny mae ganddo brofiad gyda datblygu gwe ôl-brosesu gyda chwmni cychwynnol yng Nghymru. Mae Joshua'n cyfrannu at ddatblygu Pecyn Gwaith 4 trwy ddatblygu ac integreiddio cyfres o offer ar gyfer y pecyn pedagogaidd.

Digwyddiadau: Cyfarfod Prosiect Cyfan CorCenCC (09/03/19)

Ar 9 Mawrth, daeth 24 aelod o dîm CorCenCC – o Dîm Rheoli CorCenCC i'r PAG (Grŵp Ymgynghori'r Prosiect) – at ei gilydd yng Nghaerdydd ar gyfer Cyfarfod olaf Tîm Cyfan Prosiect CorCenCC. Canolbwyntiodd rhan gyntaf y cyfarfod ar ffarwelio â rhai aelodau, croesawu aelodau newydd a chlywed y newyddion diweddaraf o'r deuddeg mis diwethaf ar draws yr holl Becynnau Gwaith. Roedd hyn wedi cynnwys arddangosiad o offer ymholi CorCenCC ar waith!

Canolbwyntiwyd yn y prynhawn ar ein defnyddwyr ac aelodau o'r bwrdd ymgynghorol yn y dyfodol: adfyfyriwyd ar draweffaith a defnyddiau posibl CorCenCC a'r ffordd orau i ddefnyddwyr y dyfodol ymgysylltu â'r corpws. Ar sail y trafodaethau hyn, rydym yn cynllunio ambell ddigwyddiad ymgysylltu a lledaenu gwybodaeth i'r cyhoedd dros y misoedd nesaf yma – felly cofiwch gadw llygad am y rhain! Roedd yn gyfarfod gwerthfawr ac ysgogol a hoffwn i, fel Prif Ymchwilydd, ddweud diolch o galon i bawb a ddaeth ac i bawb sydd wedi cefnogi'r prosiect dros y blynyddoedd – oni bai amdanoch chi, fydden ni ddim wedi cyrraedd ble rydym ni arni nawr.

Digwyddiadau: Symposiwm IVACS @ Dortmund (29/03/19)



Yn ddiweddar, aeth Dawn Knight a Steve Morris i symposiwm blynyddol IVACS a gynhaliwyd eleni yn y Technische Universität Dortmund. Roedd yn gyfle i roi'r newyddion diweddaraf am ddatblygiadau ar y prosiect i rai o'r sawl oedd yn bresennol yn ogystal â lledu'r gair i ambell un oedd heb glywed am CorCenCC o'r blaen. Thema'r gynhadledd oedd Ieithyddiaeth Corpws a'r Ystafell Ddosbarth ac roedd hyn yn rhoi cyfle i ni sôn am ein cynlluniau yn WP4 ond hefyd i ddysgu am sut mae cymunedau ieithyddol tebyg fel yr Wyddeleg yn defnyddio corpora i fynd i'r afael â heriau yn yr ystafell ddosbarth yn eu cyd-destunau eu hunain. Soniodd llawer o'r cynrychiolwyr am sut roedd clywed am adfyfyrdodau Dawn a Steve ar sicrhau a rheoli prosiectau academaidd mawr wedi eu helpu a'u calonogi wrth lunio eu cynlluniau ymchwil eu hunain i'r dyfodol. Edrychwn ymlaen at gwrdd â nifer o gyfeillion IVACS yng nghynhadledd CL2019 yn nes ymlaen eleni ym mis Gorffennaf.

+ Hybu CorCenCC: Dydd Gŵyl Dewi Sant

Mae tîm CorCenCC wedi bod yn brysur yn hybu'r prosiect ac yn annog cyfraniadau ar ddiwrnod ein nawddsant! Dyma rai o'n lluniau o weithgareddau Dydd Gŵyl Dewi yng Nghaerdydd.

Mae'n bwysig i'r prosiect ymwneud â digwyddiadau lleol mewn egwyddor ac ar lefel ymarferol. Dyletswydd unrhyw brosiect academaidd yw dod at ganlyniad sydd o ddefnydd i gymuned benodol neu i gynnydd yn ein gwybodaeth o'r byd o'n hamgylch. Er mwyn sicrhau bod y prosiect yn ddefnyddiol i'r 'gymuned' darged, felly, mae angen cysylltiad cryf a chyson rhwng staff y prosiect a'r rheini sydd am fuddio o'r prosiect. Yn achos CorCenCC, golyga hyn drafod y prosiect yn gyson â'r cyhoedd er mwyn denu cyfraniadau at y prosiect yn ogystal â chodi ymwybyddiaeth o'r prosiect a'r adnodd fydd ar gael wedi iddo ddod i ben.

Hoffai tîm CorCenCC ddiolch o galon i'n holl gyfranogwyr sydd wedi gwneud y pethau bychain (a mawrion!) a chyfrannu data i'r corpws. Diolch hefyd i bencampwyr y prosiect ac i'n cefnogwyr oll. Mae'r gwaith o ddenu cyfraniadau'n parhau, er mwyn i'r corpws gorffenedig fod yn adnodd gytbwys sy'n cynrychioli pob math o siaradwr Cymraeg gan gynnwys pobl o wahanol ardaloedd, o wahanol oedrannau, ac o wahanol lefelau defnydd yr iaith o ddydd i ddydd. Fel rhan o ymrwymiad y tîm i ymgysylltu â'r gymuned, mae gweithgareddau a digwyddiadau lleol yn oll bwysig o ran denu gwahanol bobl i gyfrannu.

Pwysig hefyd yw hyrwyddo'r prosiect fel bod pobl yn ymwybodol o'r adnodd wedi ei gwblhau. Mae modd gwneud hyn drwy safleoedd cymdeithasol fel ar Twitter neu ar Facebook, ond ni fydd y prosiect yn cyflawni ei ddyletswydd i'r gymuned darged (sef unrhyw un sydd â diddordeb mewn astudio neu ddysgu'r iaith Gymraeg) drwy ddibynnu ar y cyfryngau cymdeithasol yn unig. Mae'n hanfodol i'r prosiect a phrosiectau tebyg gael cefnogaeth a defnydd gan siaradwyr Cymraeg gan ddeall eu hanghenion yn drylwyr. Parheir y gwaith ymgysylltu hyd at ddiwedd y prosiect ym mis Awst.

Bydd cyfleoedd i gwrdd â thîm CorCenCC a thrafod prosiect y corpws yn yr Eisteddfod Genedlaethol yn ogystal â digwyddiadau lleol hyd at ddiwedd mis Awst eleni.

Os ydych chi'n meddwl cyfrannu data at y prosiect, ond yn ansicr o sut i wneud, dilynwch ein cyfrifon ar y gwefannau cymdeithasol https://twitter.com/CorCenCC a https://www.facebook.com/CorCenCC/ , lawrlwythwch yr ap neu gyrrwch neges ar CorCenCC@caerdydd.ac.uk!

+ Bywyd ar y tu mewn: profiad gwaith CorCenCC

Yn ystod yr wythnosau diwethaf, rydym wedi croesawu Gareth Smith i dîm CorCenCC yng Nghaerdydd wrth iddo wneud ychydig o brofiad gwaith ar y prosiect. Gofynnon ni i Gareth adfyfyrio ar ei brofiad a rhoi ei farn am y prosiect yn gyffredinol: 'Helo – Gareth ydw i ac eleni dwi'n astudio am MA – "Astudiaethau Cymreig a Cheltaidd" yn Ysgol y Gymraeg, Prifysgol Caerdydd. Ar ddechrau'r semester cyntaf cawsom gyflwyniad gan gynrychiolwyr CorCenCC yn esbonio natur ac amcanion y prosiect a gwnaeth hynny sbarduno fy niddordeb yn eu gwaith. Ar y cwrs MA, buom yn archwilio amryw

agweddau ar sosioieithyddiaeth mewn cydberthynas â'r Gymraeg ac yn syth sylweddolais pa mor hanfodol yw'r corpws iaith arfaethedig ar gyfer ymchwil ym meysydd megis tafodieitheg a defnydd iaith. Mae'r "profiad gwaith" yn un elfen o'r cwrs MA sy'n cynnig cyfle i'r myfyrwyr ymgymryd ag wythnos o waith gyda sefydliad o'u dewis. Mae'n ffordd wych o ryngweithio, datblygu sgiliau a chael blas ar agwedd penodol o'r byd gwaith.

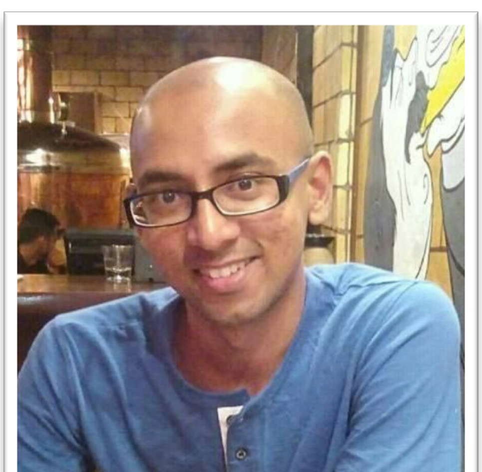
Cyn dychwelyd i astudio, bues i'n gweithio am nifer o flynyddoedd fel dadansoddwr busnes a thechnoleg felly penderfynais fod cydweithio gyda thîm CorCenCC yn gyfle i gyfuno 'niddordebau mewn iaith a thechnoleg. Mae 'nghyfraniad i'r prosiect wedi cwmpasu amrywiaeth o dasgau yn cynnwys trawsgrifio a gwirio darnau llafar a negeseuon testun a sicrhau preifatrwydd data – ac hyn oll wedi estyn fy nealltwriaeth o'r ymdrech a'r her sydd wrth wraidd sefydlu corpws llwyddiannus. Er bod fy mhrofiad gwaith yn dod i derfyn, bydda i'n parhau fy nghysylltiad gyda thîm CorCenCC trwy 'ngwaith fel trawsgrifydd.

A phwy a ŵyr? Efallai bydda i'n gwneud defnydd o'r corpws am waith ymchwil yn y dyfodol…***Diolch am dy waith caled Gareth. Mae dy gyfraniadau wedi bod yn amhrisiadwy. Pob lwc gyda'r MA!***

+ Cynllun ymchwil CorCenCC

Vigneshwaran Muralidaran, Myfyriwr PhD ym Mhrifysgol Caerdydd

Mae ieithyddiaeth corpws yn darparu dull empirig o ddadansoddi'n ieithyddol, trwy dynnu mewnweledIad o gorpws iaith naturiol - casgliad systematig, strwythuredig, wedi'i gynllunio o destunau sy'n digwydd yn naturiol mewn iaith benodol. Mae ieithydd corpws yn chwilio am eiriau neu ymadroddion penodol yn y corpws, amlder digwyddiad eitem benodol mewn testun, ac yn amlygu achosion targed a'u cyd-destun er mwyn astudio gwahanol fathau o ffenomena ieithyddol. Rhyddheir corpora gyda rhyngwyneb meddalwedd priodol er mwyn hwyluso'r ymholiadau hyn trwy restrau geiriau, rhestrau amlder, geiriau allweddol mewn cyd-destun penodol, a dywediadau a gydleolir. Gellid defnyddio corpws hefyd i ddatblygu offer Prosesu Iaith Naturiol (NLP) ar gyfer ieithoedd, megis tagio Rhannau Ymadrodd, Dosbarthu Cystrawennol a Dadansoddi Semanteg.



Mae dosbarthu cystrawennol yn fodiwl NLP sy'n nodi strwythur gramadegol brawddeg, a gellir ei ddefnyddio hefyd i gyflawni tasgau adalw gwybodaeth eraill, megis crynhoi dogfennau, dadansoddi teimladau, peiriannau chwilio a chyfieithu peirianyddol. Fy nod innau, yn fy astudiaethau PhD, yw datblygu dosbarthwr cystrawen ar gyfer y Gymraeg fel rhan o brosiect CorCenCC. Mae'r gwaith ymchwil a datblygu ar gyfer offer adnoddau technoleg iaith i'r Gymraeg yn parhau, ac ni ddatblygwyd dosbarthwr cystrawen addas ar gyfer yr iaith Gymraeg eto. Er bod dosbarthu cystrawennol yn broblem sydd wedi cael llawer o sylw ynddi'i hun fel rhan o NLP, nid oes un fframwaith safonol ar gyfer dadansoddi gramadegol neu anodi cystrawennol oherwydd nodau a dewisiadau damcaniaethol ymchwilwyr. Mae sawl dull damcaniaethol o ymdrin â strwythurau gramadegol ym maes ieithyddiaeth,

megis gramadeg cynhyrchiol, gramadeg seiliedig ar gyfyngiadau, theori strwythurol, gramadeg ffurfiannol, dulliau gwybyddol, a ieithyddiaeth feintiol, i enwi rhai yn unig. Pan gaiff y dulliau damcaniaethol hyn eu modelu'n gyfrifiadurol ar gyfer prosesu iaith naturiol, mae amrywiaeth mawr o gynlluniau anodi ar gyfer gwahanol ieithoedd.

Yn y cyd-destun hwn, mae gen i ddiddordeb mewn deall iaith fel system symbolig sy'n paru ffurf ac ystyr bob cam o'r lecsicon i'r gramadeg. Yn fy ngwaith ymchwil cynharach, nododd fy nghydweithiwr a minnau rai ffeithiau cyffredinol diddorol o ran gweithrediad cystrawen ieithoedd Drafidaidd, y gellid eu hesbonio'n dda trwy fabwysiadu agwedd theoretig seiliedig ar ddamcaniaethau defnydd, fel Gramadeg Ffurfiannol, Gramadeg Gwybyddol. Y syniad sylfaenol yw nad system ffurfiol sy'n ei rheoli ei hun yw gramadeg, yn gwbl annibynnol ar ystyr. Mae i bob cystrawen ieithyddol symbyliad ystyrlon, ffurfiad gwybyddol, ac amlinelliad trwy ddefnydd, ac mae patrymau defnydd o'r fath eto yn cael eu ffurfio'n ystyrlon i greu amlinelliadau ffurfiannol newydd. O set gyfyngedig o amlinelliadau ffurfiannol o'r fath, sy'n seiliedig ar ddefnydd, gellir enghreifftio set ddibendraw o fynegiannau ieithyddol a'u rhoi at ei gilydd i ffurfio brawddegau go iawn. Nodwyd hefyd fod cysylltiadau disgwrs yn chwarae rôl sylweddol wrth ddeall amlinelliadau ffurfiannol ieithoedd Drafidaidd.

Mae fy ngwaith ymchwil cyfredol wedi estyn y gwaith cefndir hwn er mwyn canolbwyntio ar nodi a chymhwyso perthnasedd syniadau Gramadeg Ffurfiannol i ddysgu gramadeg yn awtomatig o destun Cymraeg sylfaenol sydd heb ei anodi â llaw. At ddiben fy astudiaethau PhD, rwy'n ceisio dysgu patrymau defnydd gyda modelau Rhwydwaith Niwral Artiffisial (ANN). Model cyfrifiadurol, anllinol yw ANN, a ysbrydolwyd gan strwythur niwral ymennydd dynol, sy'n gallu dysgu tasgau gwahanol dysgu patrwm fel dosbarthu, gwneud penderfyniadau, rhagfynegi ac ati trwy ystyried amrywiol esiamplau. Dyma rai o'r heriau sydd ynghlwm wrth y prosiect: dewis y bensaernïaeth briodol ar gyfer dysgu, nodi patrymau defnydd heb fawr ddim anodi â llaw, os o gwbl, strategaethau i werthuso'r patrymau allbwn a ddysgwyd gan y system.

+ 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** - Steven Neale, Jennifer Needs, Mair Rees, 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