

Yr Amliadur:
Rhestri Amllder ar gyfer Cymraeg Cyfoes
Frequency Lists for Contemporary Welsh

D. Knight, S. Morris, B. Tovey-Walsh,
T. Fitzpatrick, & L. Anthony



Prifysgol Caerdydd | University of Cardiff

2020

I ddyfynnu *Yr Amliadur*:

Knight, D., Morris, S., Tovey-Walsh, B., Fitzpatrick, T. ac Anthony, L. (2020). *Yr Amliadur: Rhestrï Amllder ar gyfer Cymraeg Cyfoes*. Prifysgol Caerdydd, <http://doi.org/10.17035/d.2020.0120164107>

To cite *Yr Amliadur*:

Knight, D., Morris, S., Tovey-Walsh, B., Fitzpatrick, T. and Anthony, L. (2020). *Yr Amliadur: Frequency lists for contemporary Welsh*. Cardiff University, <http://doi.org/10.17035/d.2020.0120164107>

Cynnwys | Contents

1	Cefndir Background	1
1.1	CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes	1
1.2	CorCenCC: The National Corpus of Contemporary Welsh . . .	2
2	Rhagarweiniad Introduction	3
2.1	Nodiadau rhagarweiniol ar y rhestrï amllder	3
	Y rhestrï amllder ar sail geiriau	5
	Y rhestrï amllder ar sail lemâu	5
2.2	Introductory notes on the frequency lists	6
	The word frequency lists	7
	The lemma frequency lists	8
3	Dehongli'r Rhestrï Interpreting the Lists	9
3.1	Dehongli'r rhestrï: Cyfyngiadau posibl a nodiadau rhybyddiol .	9
	Lemateiddio	9
	Prosesu data	10
	Ffurfiâu ansafonol	12
	Ystyriaethau eraill	14
3.2	Interpreting the lists: Potential limitations and cautionary notes	15
	Lemmatization	15
	Data processing	16
	Non-standard forms	17
	Other considerations	20
4	Llyfryddiaeth Bibliography	21
	Atodiadau Appendices	22

Atodiad Appendix A Rhannau ymadrodd CorCenCC	
The CorCenCC part-of-speech tagset	22
Atodiad Appendix B Trefniant yn ôl yr wyddor Gymraeg Welsh alphabetical ordering	44
Atodiad Appendix C Rhestrï amllder Frequency lists	45
Cyflwyniad	45
Introduction	48
C.1 100 gair uchaf yn CorCenCC	
Top 100 words in CorCenCC	51
C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)	
Top 1000 words in CorCenCC (alphabetic list)	59
C.3 100 lema uchaf yn CorCenCC	
Top 100 lemmas in CorCenCC	72
C.4 1000 lema uchaf yn CorCenCC (rhestr yn ôl yr wyddor)	
Top 1000 lemmas in CorCenCC (alphabetic list)	80
C.5 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig)	
Top 100 lemmas in CorCenCC (open-class words only)	93
C.6 1000 gair uchaf yn CorCenCC (geiriau dosbarth agored yn unig; rhestr yn ôl yr wyddor)	
Top 1000 lemmas in CorCenCC (open-class words only; alphabetic list)	98
C.7 500 enw uchaf yn CorCenCC	
Top 500 noun lemmas in CorCenCC	111

C.8	500 berf uchaf yn CorCenCC	
	Top 500 verb lemmas in CorCenCC	118
C.9	500 ansoddair uchaf yn CorCenCC	
	Top 500 adjective lemmas in CorCenCC	125
C.10	50 adferf uchaf yn CorCenCC	
	Top 50 adverb lemmas in CorCenCC	132
C.11	50 ebychiad uchaf yn CorCenCC	
	Top 50 interjection lemmas in CorCenCC	133
C.12	100 gair dosbarth agored uchaf yng nghydran ysgrifenedig Cor- CenCC	
	Top 100 open-class words in the written component of Cor- CenCC	134
C.13	100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC	
	Top 100 open-class words in the spoken component of Cor- CenCC	142
C.14	100 gair dosbarth agored uchaf yng nghydran electronig Cor- CenCC	
	Top 100 open-class words in the electronic component of Cor- CenCC	150

1 Cefndir | Background

1.1 CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes

Mae corpws CorCenCC yn cynnwys 14,338,149 o docynnau (tua 11.2 miliwn o eiriau) o ffynonellau Cymraeg ysgrifenedig, llafar ac electronig (testunau digidol, ar-lein) a dynnwyd o amrywiaeth o genres, amrywiadau iaith (rhanbarthol a chymdeithasol), a chyd-destunau. Mae cyfranwyr CorCenCC yn cynrychioli'r hanner miliwn o siaradwyr Cymraeg yn y wlad. Prosiect cymunedol oedd creu CorCenCC a roddodd gyfle i ddefnyddwyr y Gymraeg fod yn rhagweithiol wrth gyfrannu at adnodd Cymraeg sy'n adlewyrchu sut mae'r Gymraeg yn cael ei defnyddio ar hyn o bryd.

Er mwyn sicrhau bod CorCenCC mor gynrychiadol o Gymraeg cyfoes â phosibl, cynlluniodd tîm y prosiect fframwaith samplu pwrpasol. Casglwyd dyfyniadau o ffynonellau gan gynnwys, er enghraifft, cyfnodolion, negeseuon e-bost, pregethau, arwyddion ffordd, rhaglenni teledu, cyfarfodydd, cylchgronau, a llyfrau, a thrwy ap torfoli penodol (gw: <https://www.corcencc.cymru/ap/>) roedd modd i siaradwyr Cymraeg yn y gymuned recordio a lanlwytho samplau o'u defnydd eu hunain o'r iaith i'r corpws. Felly, mae'r corpws a gyhoeddwyd yn cynnwys data gan siaradwyr Cymraeg o bob math o gefndir, gallu, a chyd-destun, gan gipio sut mae'r Gymraeg wir yn cael ei defnyddio ar draws y wlad heddiw. I ddod o hyd i wybodaeth fanwl am brosiect CorCenCC, ewch i <https://www.corcencc.cymru>.

Mae'r data yn CorCenCC yn cynrychioli amrywiaeth eang o gyd-destunau, genres a thestunau. Cyn belled ag y bo modd, cafodd y data eu hanonymeddio gan ddefnyddio cyfuniad o dechnegau â llaw ac wedi'u hawtomeiddio, a'u tagio'n llawn o ran categorïau rhan ymadrodd (RhY) a semantig. Cyflawn-

wyd y tagio RhY a semantig gan ddefnyddio offer CyTag (Neale, Donnelly, Watkins, & Knight, 2018) a CySemTag (Piao, Rayson, Knight, & Watkins, 2018; Piao, Rayson, Knight, Watkins, & Donnelly, 2017) sydd ar gael ar wefan GitHub CorCenCC: <https://github.com/CorCenCC>.

Cyllidwyd yr ymchwil y seiliwyd y set ddata yma, yr offer meddalwedd cysylltiedig, a'r adnoddau corpws ar-lein arni gan Gyngor Ymchwil Economaidd a Chymdeithasol y DU (ESRC) a Chyngor Ymchwil y Celfyddydau a'r Dyniaethau (AHRC) fel prosiect Corpws Cenedlaethol Cymraeg Cyfoes: Dull cymunedol o greu corpws ieithyddol (Rhif Grant ES/M011348/1).

1.2 CorCenCC: The National Corpus of Contemporary Welsh

The CorCenCC corpus contains 14,338,149 tokens (circa 11.2 million words) from written, spoken and electronic (online, digital texts) Welsh language sources, taken from a range of genres, language varieties (regional and social), and contexts. The contributors to CorCenCC are representative of the over half a million Welsh speakers in the country. The creation of CorCenCC was a community-driven project, which offered users of Welsh an opportunity to be proactive in contributing to a Welsh language resource that reflects how Welsh is currently used.

To make CorCenCC as representative of contemporary Welsh as possible, the project team designed a bespoke sampling framework. Extracts were collected from sources including, for example, journals, emails, sermons, road signs, TV programmes, meetings, magazines and books. Conversations were recorded by the research team, and a specially designed crowdsourcing app (see: <https://www.corcenc.org/app/>) enabled Welsh speakers in the community to record and upload samples of their own language use to the corpus. The

published corpus therefore contains data from Welsh speakers from all kinds of backgrounds, abilities and contexts, capturing how Welsh is truly used today across the country. For detailed information about the CorCenCC project, visit <https://www.corcenc.org>.

The data in CorCenCC represents a wide range of contexts, genres and topics. This data has, as far as possible, been anonymised using a combination of manual and automated techniques, and has been fully tagged in terms of part-of-speech (POS) and semantic categories. The POS and semantic tagging was carried out using the CyTag (Neale et al., 2018) and CySemTag (Piao et al., 2018; Piao et al., 2017) tools, available from CorCenCC's GitHub website: <https://github.com/CorCenCC>.

The research on which this dataset, the accompanying software tools, and online corpus resources are based was funded by the UK Economic and Social Research Council (ESRC) and Arts and Humanities Research Council (AHRC) as the *Corpws Cenedlaethol Cymraeg Cyfoes* (The National Corpus of Contemporary Welsh): A community driven approach to linguistic corpus construction project (Grant Number ES/M011348/1).

2 Rhagarweiniad | Introduction

2.1 Nodiadau rhagarweiniol ar y rhestri amlder

I'r rhestri amlder geiriau a lemâu a gyflwynir yma ac yn y taenlenni cysylltiedig (ar gael yn <http://doi.org/10.17035/d.2020.0120164107>), y tocynnau alffabetig yn unig a gafodd eu cyfrif tuag at gyfanswm y tocynnau yn y corpws neu mewn is-gorpws. Caewyd y categorïau canlynol o docynnau alffabetig allan o'r rhestri hefyd:

- Data wedi'u hanonymeiddio: Defnyddiwyd geiriau generig i gymryd lle gwybodaeth a all fod yn sensitif wrth anonymeiddio'r testunau mewnbyn. Lle bo modd, tagiwyd y geiriau hyn yn y corpws (fel enwb am enw benywaidd neu cyfenw) gyda'r rhan ymadrodd *Anon*.
- Nodiadau trawsgrifwyr: Yn nhrawsgrifiadau'r data llafar, cynhwysai'r trawsgrifwyr nodiadau am nodweddion llafar nad oeddynt yn eirfaol (fel pesychu, ochneidio, neu ddyllyfu gê) ac am synau cefndirol yn y recordiadau. Tynnwyd y rhain allan gan eu bod yn rhan o'r meta-destun, nid o gynnwys y corpws ei hun.
- Geiriau nad ydynt yn Gymraeg
- Enwau priod

Er nad ydynt yn cael eu cynnwys yn y rhestr (sy'n cynnwys geiriau Cymraeg yn unig), mae tocynnau alffabetig fel geiriau estron ac enwau priod yn cael eu cyfrif tuag at brif gyfanswm y tocynnau yn y corpws. Felly, caiff ffigurau amllder fesul miliwn o docynnau eu cyfrifo ar sail cyfanswm nifer y tocynnau alffabetig yn y corpws.

Er mwyn creu'r rhestr hyn, cywirwyd rhai gwallau tagio treiddiol a thagiâu rhan ymadrodd amwys yn nata'r corpws â llaw. Yn ogystal, tacluswyd y ffeil ddata er mwyn dileu ambell i ddarn dieisiau (e.e. darnau o god JavaScript yn y samplau iaith electronig) nad oedd eisioes wedi'u cywiro gan prosesau awtomatig. Fel canlyniad i'r camau hyn, bydd y rhestr sampl a'r set ddata amllder lawn yn y daenlenni gysylltiedig yn amrywio i ryw raddau o restru cyffelyb pe baent yn cael eu creu o fersiwn cyhoeddus cyntaf set ddata CorCenCC (neu o unrhyw fersiwn wedi'i diweddarau y cyhoeddir yn y dyfodol).

Y rhestri amllder ar sail geiriau

Seilir rhestri C1 a C2 ar amllder geirffurfiau unigol. Dibynna pob cofnod ar ffurf ysgrifenedig tocyn yn unig. Golyga hyn fod homograffau yn cael eu grwpio gyda'i gilydd: er enghraifft, mae'r ategydd berfol *wedi* a'r arddodiad *wedi* yn cael eu cyfrif fel un cofnod geirffurf. Mae amrywiolion fel talfyriadau a gogwyddeiriau yn cael eu cyfrif ar wahân. O ganlyniad, felly, byddai gan y gair *cariad* a'i ffurf luosog *cariadon* gofnodion ar wahân yn y rhestri hyn. Yn yr un ffordd, mae'r ffurfiau wedi'u treiglo *gariad*, *chariad*, a *nghariad* i gyd yn cael eu cyfrif ar wahân i'r geirffurf *cariad*.

Mae rhestri C6 a C7–14 wedi'u seilio ar eirffurfiau yn hytrach na lemâu. Caiff amrywiolion eu gwahanu fel yn y rhestri geiriau rheolaidd. Fodd bynnag, dim ond y geirffurfiau hynny â chategoriâu rhan ymadrodd penodol sy'n cael eu cyfrif ar gyfer y rhestri hyn. Mae rhestri C7–11 i gyd yn cyfrif un rhan ymadrodd yn unig. Geiriau dosbarth agored yn unig a gynhwysir yn rhestri C6 a C12–14. Caiff unrhyw homograffau dosbarth caeedig eu gadael allan o'r cyfrif cronus a briodolir i'r geirffurfiau yn y rhestri hyn.

Y rhestri amllder ar sail lemâu

Mae'r rhestri wedi'u lemateiddio (C3 a C4) yn rhoi cyfrifon amllder ar gyfer lemâu yn hytrach na geirffurfiau. Yn y rhestri hyn, caiff amrywiolion eu grwpio gyda'i gilydd, a homograffau â rhannau ymadrodd gwahanol yn cael eu gwahanu lle bo modd. Mewn rhai achosion, golyga hyn y bydd geirffurfiau sy'n cael eu cyfrif ar wahân yn y rhestri geiriau yn cael eu grwpio gyda'i gilydd: caiff ffurfiau lluosog ac unigol enwau, er enghraifft, eu cyfrif fel eu bod yn perthyn i'r un lema. Mewn achosion eraill, bydd tocynnau a gafodd eu grwpio fel un cofnod yn rhestri geiriau yn ymddangos fel cofnodion ar wahân yn

y rhestrï o lemâu. Dyma'r achos yn bennaf o ran homograffau gyda rhannau ymadrodd gwahanol, fel *achos* (enw) ac *achos* (cysylltair). Rhestrir yr homograffau hyn gydag un amllder cyfunol yn y rhestrï amllder geirïau, ond maent yn perthyn i gofnodion ar wahân yn y rhestrï amllder lemâu.

Wrth gasglu'r rhestrï amllder lemâu, nid ydym wedi cyfrif unrhyw do-cynnau na fedrai CyTag ddadamwyso eu rhan ymadrodd. Cymerwyd hyn i ystyriaeth wrth gyfrifo'r holl ystadegau, felly gall cyfanswm y cyfrif ar gyfer unrhyw eirffurf penodol yn y rhestrï amllder geirïau fod ychydig yn uwch na'r un geirffurf ar ôl cael ei rannu'n lemâu ar wahân.

Mae rhestr C5 yn cynnwys lemâu hefyd, ond mae'n cyfrif y rhai sydd â rhan ymadrodd dosbarth agored yn unig.

2.2 Introductory notes on the frequency lists

For both the word and the lemma frequency lists presented here and in the associated spreadsheets (available at <http://doi.org/10.17035/d.2020.0120164107>), only alphabetic tokens have been counted towards the total of tokens in the corpus or in a subcorpus. All tokens and lemmas have been converted to lowercase for the purpose of generating these lists. The following categories of alphabetic tokens were also excluded from the lists:

- Anonymized data: Generic words were used to replace potentially sensitive information in anonymizing the input texts. Where possible, these words (such as *enwb* for a female name or *cyfenw* for a surname) were tagged in the corpus with the part-of-speech *Anon*.
- Transcribers' notes: In the transcripts of spoken data, transcribers included notes about non-lexical features of speech (such as coughing,

sighing, or yawning) and about background noises in the recordings. These are excluded since they are part of the meta-text, not of the corpus content proper.

- Non-Welsh words
- Proper nouns

Although they are excluded from the lists (which contain Welsh words only), alphabetic tokens such as foreign words and proper nouns are counted towards the grand total of tokens in the corpus. Figures given for frequency per million tokens are therefore calculated based on the total number of alphabetic tokens in the corpus.

In order to generate these lists, some pervasive tagging errors and ambiguous part-of-speech tags in the corpus data were manually corrected. In addition, the data file was somewhat refined in order to remove a few unwanted passages (e.g. of JavaScript code in the e-language samples) which had not already been deleted during automatic processing. These sample lists and the associated spreadsheets will therefore differ from similar lists if they were generated from the first public release of the CorCenCC dataset (or from any updated versions that are released in the future).

The word frequency lists

Lists C1 and C2 are based on the frequency of individual word-forms. Each entry relies solely on the written form of a token. This means that homographs are grouped together: for example, the preverbal particle *wedi* and the preposition *wedi* are counted as a single word-form entry. Variant forms such as abbreviations and clitics are counted separately. As a result, for example, the

words *cariad* (“love”, “lover”) and its plural form *cariadon* (“lovers”) would have separate entries in these lists. Similarly, the mutated forms *gariad*, *chariad*, and *nghariad* are all counted as separate from the word-form *cariad*.

Lists C6 and C7–14 are also based on word-forms rather than lemmas. Variant forms are separated as in the regular word lists. However, only word-forms with particular part-of-speech categories are counted for these lists. Lists C7–11 all count only a single part-of-speech. Lists C6 and C12–14 feature only open-class words. Any closed-class homographs are excluded from the cumulative count ascribed to the word-forms in these lists.

The lemma frequency lists

The lemmatized lists (C3 and C4) present frequency counts for lemmas rather than for word-forms. In these lists, variant forms are grouped together and homographs with different parts of speech are separated where possible. In some cases, this means that word-forms counted separately in the wordlists will be grouped together: plural and singular forms of nouns, for example, are counted as belonging to the same lemma. In other cases, tokens which were grouped as a single entry in the word lists will appear as separate entries in the lists of lemmas. This is primarily the case for homographs with different parts of speech, such as *achos* (noun meaning “cause”) and *achos* (conjunction meaning “because”). These homographs are listed with a single, combined frequency in the word frequency lists, but belong to separate entries in the lemma frequency lists.

In compiling the lemma frequency lists, we have not counted any tokens whose part of speech could not be disambiguated by CyTag. This has been taken into account in calculating all statistics, so the total count of a given

word-form in the word-form frequency lists may be somewhat higher than the combined total for the same word-form when split into separate lemmas.

List C5 also lists lemmas, but only counts those with an open-class part of speech.

3 Dehongli'r Rhestri | Interpreting the Lists

3.1 Dehongli'r rhestri: Cyfyngiadau posib a nodiadau rhybyddiol

Defnyddiwyd tagiwr rhan ymadrodd CyTag prosiect CorCenCC i labelu pob tocyn yn y corpws gyda gwybodaeth yn cynnwys tag rhan ymadrodd a lema. Mae gan CyTag sgoriau trachywiredd tagio a galw i gof ardderchog yn bell dros 90% (Neale et al., 2018). Eto i gyd, mae ei ddibynadwyedd yn disgyn wrth brosesu tocynnau y mae'n anodd eu dadamwysu (Neale et al., 2018). Dylid cofio hyn wrth ddehongli amlderau geiriau neu lemâu sy'n agored i gael eu cam-dagio. Mae'r crynodeb isod yn amlygu rhai meysydd lle y gall y data rhestri amllder yn y ddogfen hon a'r taenlenni cysylltiedig fod yn llai dibynadwy neu gynnwys canlyniadau a all ymddangos yn afreolaidd o ganlyniad i berfformiad y tagiwr.

Er gwaethaf ambell ragrybudd fel hyn, mae gan y data amllder ar raddfa fawr wedi'u lemateiddio yma ar gyfer Cymraeg cyfoes botensial i hwyluso dadansoddi ac ymchwilio ar lefel teulu gair, gair pen, neu fôn y gair.

Lemateiddio

Fel arfer, mae cyfrif amllder ar gyfer lemâu yn cynhyrchu rhestri sy'n llai dibynadwy na rhestri amllder geirffurfiau gan fod un yn gofyn am dagio rhan

ymadrodd trachywir a'r llall yn dibynnu a ffurf ysgrifenedig gair yn unig (Kilgarriff, 1997). Mae'n annhebygol y bydd neilltuo geirffurfiau i'r lema cywir trwy ddulliau cyfrifiadurol yn 100% gywir (Evison, 2010; O'Keeffe, McCarthy, & Carter, 2007). Yn nata CorCenCC, neiltuodd CyTag nifer o dagiau rhan ymadrodd a lema i docynnau nad oedd yn medru eu dadamwysu. Ar ben hynny, mewn rhai achosion, mae'n bosibl i docyn wedi'i ddadamwysu gael ei neilltuo i'r tag rhan ymadrodd anghywir. Cafodd CyTag anhawster arbennig gyda thocynnau fel *yn*, sydd yn digwydd yn aml ac yn anodd ei ddadamwysu (Neale et al., 2018). Mae *yn* yn gweithredu fel arddodiad, ategydd berfol, ac yn draethiadol:

- (1) *Mae Siân yn y tŷ.*
bod.PRS.3.UN Siân yn.ARDD y tŷ
- (2) *Mae Siân yn gadael.*
bod.PRS.3.UN Siân yn.ATEG.BERF. gadael
- (3) *Mae Siân yn oer.*
bod.PRS.3.UN Siân yn.TRAETH. oer

Mae'r enghreifftiau o'r defnydd o *yn* yma i gyd yn gyffredin yn y Gymraeg ac roedd yn anodd gwahanu rhyngddynt yn fanwl gywir wrth ddefnyddio CyTag. Mae unrhyw ddatblygu pellach ar CyTag yn debygol o ganolbwyntio at docynnau amllder uchel o'r fath (Neale et al., 2018).

Prosesu data

Cafodd y data yn CorCenCC eu casglu oddi wrth amrywiaeth eang o ffynonellau a'u prosesu gan ddefnyddio technegau â llaw ac wedi'u hawtomeiddio. Ni lwyddodd prosesu'r data electronig, yn benodol, i ddileu nodweddion all-destunol bob tro e.e. cod HTML a JavaScript, penynnau a throedynnau ebost, neu destun a grëwyd gan feddalwedd fel ategion WordPress. O ganlyniad,

mae'n debygol y bydd rhestr o eiriau wedi'u cymryd o'r data electronig nad ydynt yn Gymraeg (heb eu cyflwyno yma) wedi chwyddo cyfrifon y tocynnau sy'n ymddangos yn aml mewn cod datblygu gwe neu yn yr hyn sy'n cael ei alw'n destun 'haearn boeler' (testun a ailadroddir fel sy'n cael ei ddefnyddio mewn troedynnau ebost, ffurfiau gwe, bylchau sylwadau ar-lein, ac yn y blaen).

Tynnwyd darnau yn y data nad oeddynt yn Gymraeg (yn Saesneg yn bennaf) fel rhan o'r biblinell prosesu data, er na wnaed hynny gyda geiriau neu ymadroddion nad oeddynt yn Gymraeg a oedd yn cael eu defnyddio mewn disgwrs Cymraeg yn bennaf. (Lle bo modd, cafodd y rhain eu tagio gyda'r tag rhan ymadrodd "*Gwest: Gweddillol - estron*"). Felly, ar y cyfan, bydd y testun haearn boeler sydd ar ôl yn y corpws yn Gymraeg a gall y ffaith ei fod yno chwyddo'r cyfrifon am y geiriau mae'n eu cynnwys. Mae modd dadlau a ddylid ystyried hyn fel chwyddo amllder y geiriau hyn neu beidio. Yn ein barn ni, gan fod haearn boeler yn nodwedd gyson mewn cyfathrebu trwy gyfrwng cyfrifiadur, y dewis amlwg oedd ei adael yn gyfan yn y corpws. Mae'r ffaith ei bod yn bresennol yn aml yn adlewyrchu'r ffordd y mae haearn boeler yn gwthio ei hun ar ddefnyddwyr cyffredin genres trwy gyfrwng cyfrifiadur yn Gymraeg.

Cafodd y data ysgrifenedig ac electronig eu hanonymeddio'n syml trwy newid eitemau sensitif (fel enwau personol, cyfeiriadau, rhifau ffôn, ac ati) i docynnau generig (e.g. *cyfenu*, *cyfeiriad*). Nid yw'r tocynnau meta-destunol generig hyn bob amser wedi cael eu gwahanu oddi wrth ddefnyddiau go iawn o'r tocynnau yn y testunau. O ganlyniad, bydd y cyfrifon ar gyfer y geiriau hyn wedi chwyddo.

Ffurfiâu ansafonol

Mae'r ffurfiolaeth gramadeg cyfyngiadol ("constraint grammar") a ddefnyddir yn CyTag yn ystyried cyd-destun pob tocyn mae'n ei ddadansoddi fel mewnbwn er mwyn pennu'r tag rhan ymadrodd mwyaf tebygol ar gyfer y tocyn hwnnw. Yn y bôn, gallai ddadansoddi'r tocyn *yn* trwy edrych i weld a yw'r tocyn sy'n ei ddilyn yn ferf, yn ansoddair, yn enw, neu'n fannod wedi'i dilyn gan enw. Gallai set sylfaenol o reolau gramadeg cyfyngiadol i dargedu *yn* gyfarwyddo'r gramadeg i ddewis y rhan ymadrodd "arddodiad" os mai enw (neu ddilyniant o fannod + enw) yw'r tocyn nesaf, i ddewis "traethiadol" os mai ansoddair yw'r tocyn nesaf, ac i ddewis "ategydd berfol" os mai berf yw'r tocyn nesaf. Oni bai fod rheol arall yn disodli neu'n ymyrryd fel arall yn y tair rheol seml yma, dylai'r tocyn *yn* yn yr enghreifftiau a roddwyd uchod (Engh. 1–3) gael ei dagio'n gywir. Fel mae'r enghraifft hon yn dangos, mae tagio unrhyw docyn yn gywir yn dibynnu i ryw raddau ar dagio'r tocynnau sydd o'i amgylch yn gywir. Gan mai nod CyTag yw rhoi tagiau rhan ymadrodd manwl gywir ar gyfer cymaint o dagiau ag sy'n bosibl, rydym wedi ymdrechu i wneud y system yn 'ymwybodol' nid yn unig o ffurfiâu safonol tocynnau fel y rhai a geir mewn geiriaduron, ond hefyd o ffurfiâu ansafonol ac amrywiadau llafar neu ranbarthol.

Felly, cyn belled ag y bo'n rhesymol, mae CyTag yn eithaf trugarog o ran tagio sillafiadau amyriadau ansafonol tocynnau. Yn ymarferol, er enghraifft, mae defnyddwyr cyfoes y Gymraeg yn rhoi acenion yn anghyson yn enwedig mewn genres anffurfiol. (Nid oes neb, hyd y gwyddom, wedi ymchwilio i weld a yw hyn yn cael ei achosi gan anawsterau o ran dod o hyd i'r acenion priodol ar fysellfwrdd cyfrifiadur, gan ddiffyg gwybodaeth defnyddwyr am sillafiadau safonol geiriau gydag acenion, neu gan ffactorau eraill.) Oherwydd

hyn, mae CyTag yn ystyried ffurfiau gyda a heb acenion fel darlleniadau posibl wrth ddadansoddi tocyn. Golyga hyn, er enghraifft, y gall y tocyn *a* gael ei dagio'n gywir os, fel sy'n digwydd yn aml, gaiff ei ddefnyddio fel amrywiad sillafu ar yr arddodiad cyffredin *á*, yn hytrach na'r cysylltair *a*. Gan fod *á* yn achosi treigladau llaes i'r enwau sy'n ei ddilyn, mae'r enghraifft hon yn arbennig o bwysig o ran tagio'n gywir, gan fod y gramadeg cyfyngiadol yn ystyried a all y tocynnau blaenorol achosi treigladau wrth asesu darlleniadau posibl tocyn. Os yw *a* yn cael ei ddefnyddio fel amrywiad ar *á* ond yn cael ei dagio fel pe bai'n perthyn i'r lema *a*. [CYS.], ni chaiff rheolau'r gramadeg cyfyngiadol ar gyfer tocynnau sy'n dilyn sbardunau am y treigladau llaes eu gweithredu'n gywir. Eto i gyd, yn sicr, bydd y dull trugarog yma yn achosi cynnydd yn y nifer o rai tocynnau sy'n cael eu dadamwysu'n wallus. Tra byddai dehongliad caeth yn dosbarthu *dý* (lema *tý*) a *dy* (banodolyn/rhagenw, meddiannol, 2il berson unigol) fel dau docyn gwahanol, yn y bôn byddai CyTag yn ei trin nhw'n homograffig. Er bod hyn yn ddiamau yn cynyddu'r potensial am anghywirdeb ym mherfformiad CyTag, yr oedd yn angenrheidiol er mwyn rhoi cyfrif llawn am arferion ysgrifennu a sillafu defnyddwyr cyfoes y Gymraeg.

Mae'r amrywiadau tafodieithol cyffredin y mae CyTag yn ceisio cyfrif amdanynt yn cynnwys lleihau /ai/ ar ddiwedd geiriau i /a/ neu /ɛ/ ar lafar, sy'n cael ei adlewyrchu gan rai ysgrifenyddwr wrth leihau <-ai>, <-au> ac <-ae> yn orgraffyddol i <-a> neu <-e>. Pan na all adnabod gair gydag <-a> neu <-e> ar y diwedd, mae CyTag yn chwilio am yr un gair gyda'r llythyren olaf yn cael ei disodli gan <-ai>, <-au> ac <-ae>. Er enghraifft, o gael y tocyn *chware*, bydd CyTag yn chwilio am *chwarai*, *chwarau* a *chwarae*, sy'n golygu bod modd iddo briodoli'r tocyn i'r lema cywir *chwarae*. Ar wahân i amrywiadau tafodieithol, mae CyTag yn ystyried hefyd a yw tocyn anhysbys sy'n cynnwys y dilyn-

ant canol <-nn-> yn bodoli gydag <-n-> yn lle hynny. O gael y tocyn *penderfynnu*, bydd CyTag yn adnabod yn gywir sillafu ansafonol (neu wall sillafu) sy'n perthyn i'r lema *penderfynu*. Cawsom fod newid <n> am <nn> yn gyffredin ac yn wall systematig a oedd prin (os o gwbl) yn esgor ar docyn nad oedd modd adnabod ei ffurf safonol yn ddiamwys, tra bo gwallau sillafu eraill a nodwyd yn y data yn aml yn fwy anghyson neu'n esgor ar docynnau nad oedd modd eu trosglwyddo'n hawdd yn ôl i'r ffurff orgraffyddol safonol.

Ystyriaethau eraill

Caiff demonymau ac enwau ieithoedd Cymraeg eu tagio gan CyTag fel enwau cyffredin, nid enwau priod. O ganlyniad, caiff *Cymraeg* ei restru fel yr enw mwyaf aml yn y corpws. Er nad ydynt yn enwau cyffredin a bod yn fanwl gywir, mae geiriau o'r fath yn wahanol i enwau personol, enwau brandiau, a mathau eraill o enwau priod gan fod ganddynt gyfeiryddion sengl, sefydlog sydd yn hysbys yn eang i'r gymuned ieithyddol. Gellir cymharu hyn, dyweder, â'r enw personol gwrywaidd *Gareth*, a all gyfeirio at unrhyw un o nifer o bobl a all fod yn hysbys neu beidio.

Maes arall sy'n effeithio ar ddibynadwyedd y rhestri amllder yw sut mae CyTag yn prosesu dilyniannau aml-eiriau yn Gymraeg. Yn benodol, nid yw CyTag yn tagio arddodiaid aml-eiriau fel *ar gyfer* ac *er mwyn* fel unedau sengl, ond yn hytrach mae'n eu trin fel arddodiad yn cael ei ddilyn gan enw. O ganlyniad, bydd y rhestri wedi'u lemateiddio yn dangos amllder uwch i gydrannau enwol y dilyniannau hyn na'r hyn a ddisgwylid fel arall. Yr enw *cyfer*, sy'n ffurfio ail elfen *ar gyfer* yw'r wythfed lema enwol mwyaf cyffredin yn y corpws, er mai prin iawn yw'r defnydd ohono'n annibynnol. Mewn gwirionedd, mae'n annhebygol y byddai'r rhan fwyaf o ddefnyddwyr cyfoes y Gymraeg

yn adnabod ystyr yr enw pe bai'n cael ei ddefnyddio y tu allan i'r dilyniant *ar gyfer*. Felly mae'n rhaid dehongli'r rhestri'n ofalus, a bod yn ymwybodol o'r swyddogaethau gwahanol y gall geiriau dosbarth agored eu cyflawni mewn dilyniannau aml-eiriau Cymraeg.

3.2 Interpreting the lists: Potential limitations and cautionary notes

The CorCenCC project's CyTag part-of-speech tagger was used to label each token in the corpus with information including a part-of-speech tag and lemma. CyTag has excellent tagging precision, and recall scores well over 90% (Neale et al., 2018). Nonetheless, its reliability drops when processing tokens that are hard to disambiguate (Neale et al., 2018). This should be borne in mind when interpreting frequencies of words or lemmas which are susceptible to mis-tagging. The summary below highlights some areas in which the frequency list data in this document and the associated spreadsheets may be less reliable or may contain results which seem anomalous as a result of the tagger's performance.

Notwithstanding these caveats, this large-scale lemmatized frequency data for contemporary Welsh has potential to facilitate analysis and investigation at the word family, head word, or base word level.

Lemmatization

Calculating frequency for lemmas generally produces lists which are less reliable than word-form frequency lists, since the former require accurate part-of-speech tagging, whereas the latter rely solely on the word's written form (Kilgarriff, 1997). Assigning word-forms to the correct lemma by computational

means is unlikely to be 100% accurate (Evison, 2010; O’Keeffe et al., 2007). In CorCenCC’s data, CyTag assigned multiple part-of-speech tags and lemmas to tokens that it was unable to disambiguate. In some cases, furthermore, a disambiguated token may actually have been assigned the wrong part-of-speech tag. CyTag had particular difficulty with tokens such as *yn*, which are both frequent and hard to disambiguate (Neale et al., 2018). *Yn* functions as a preposition, a verbal particle, and a predicative particle:

- (4) *Mae Siân yn y tŷ.*
be.PRS.3.SG Siân in the house.
“Siân is in the house.”
- (5) *Mae Siân yn gadael.*
be.PRS.3.SG Siân PART.V.PRES leaving.
“Siân is leaving.”
- (6) *Mae Siân yn oer.*
be.PRS.3.SG Siân PART.PRED cold.
“Siân is cold.”

Each of these uses of *yn* is common in Welsh, and it proved difficult to distinguish accurately between them using CyTag. Any further development of CyTag is likely to focus on such high-frequency tokens (Neale et al., 2018).

Data processing

The data in CorCenCC has been collected from a wide variety of sources, and processed using both automated and manual techniques. The automatic processing of the electronic data, in particular, was not always entirely successful in removing extra-textual features such as HTML and JavaScript code, email headers and footers, or text generated by software such as WordPress plugins. As a result, it is likely that lists of non-Welsh words taken from the electronic

data (not presented here) will have inflated counts of tokens that appear often in web development code or in so-called 'boilerplate' text (repeated text such as that used in email footers, web forms, online comment boxes, and so on).

Non-Welsh passages in the data (primarily in the English language) were removed as part of the data processing pipeline, although non-Welsh words or phrases used within primarily Welsh discourse were not removed. (These have, where possible, been tagged with the part-of-speech tag *Gwest*: "Gwed-dillol - estron" / *Other - foreign*). On the whole, therefore, the boilerplate text that remains in the corpus will be in Welsh, and its presence may inflate counts for the words it contains. Whether or not this should be considered an artificial inflation of these words' frequency is open to debate. In our view, since boilerplate is a consistent feature of computer-mediated communication, leaving it intact in the corpus was the appropriate choice. Its repeated presence reflects the way in which boilerplate imposes itself on the average consumer of computer-mediated genres of Welsh.

Both the written and electronic data were anonymized simply by replacing sensitive items (such as personal names, addresses, telephone numbers, and so on) with generic tokens (e.g. *cyfemw* / 'surname', *cyfeiriad* / 'address'). These generic metatextual tokens have not always been separated from actual uses of the tokens in the texts. As a result, the counts for these words will be inflated.

Non-standard forms

The constraint grammar formalism used in CyTag considers the context of each token it analyses as input to determine the most likely part-of-speech tag for that token. In basic terms, it might analyse the token *yn* by looking at whether the following token is a verb, an adjective, a noun, or an article fol-

lowed by a noun. A basic set of constraint grammar rules to target *yn* might instruct the grammar to select the part of speech “preposition” if the following token is a noun (or a sequence of article + noun), to select “predicative particle” if the following token is an adjective, and to select “verbal particle” if the following token is a verb. If no other rules supersede or otherwise interfere with these three simple rules, the token *yn* in the examples given above (Exx. 4–6) should be correctly tagged. As this example illustrates, the correct tagging of any token is at least somewhat reliant on the correct tagging of the tokens that surround it. Since the aim of CyTag is to provide accurate part-of-speech tags for as many tokens as possible, we have attempted to make the system ‘aware’ not only of the standardized forms of tokens as found in dictionaries, but also of non-standard forms and colloquial or regional variants.

As far as is reasonable, therefore, CyTag takes a lenient approach to the tagging of non-standard variant spellings of tokens. In practice, for example, contemporary users of Welsh use accented characters inconsistently, particularly in informal genres. (Whether this is caused by difficulty accessing accented characters on a computer keyboard, by users’ imperfect knowledge of the standard spellings of words with accented characters, or by other factors has not, to our knowledge, been explored in any depth). Because of this, CyTag considers both accented and unaccented forms as possible readings when it analyses a token. This means, for example, that the token *a* can be correctly tagged when, as often happens, it is used as a spelling variant of the common preposition *â*, rather than as the conjunction *a*. Since *â* causes following nouns to undergo aspirate mutation, this example is particularly important for correct tagging, as the constraint grammar considers whether preceding tokens may cause mutation when assessing possible readings of a token. If *a* is being used

as a variant of *á*, but is tagged as though it belongs to the lemma *a*.[CONJ], the constraint grammar's rules for tokens following aspirate-mutation triggers will not be correctly activated. Nonetheless, this lenient approach will almost certainly cause an increase in incorrect disambiguation of some tokens. Whereas a strict interpretation would class *dŷ* (lemma *tŷ*; "house") and *dy* (determiner/pronoun, possessive, 2nd person singular) as two different tokens, CyTag would essentially treat them as homographic. Although this undeniably increases the potential for inaccuracy in CyTag's performance, it was necessary in order to account fully for contemporary Welsh users' writing and spelling habits.

Common dialectal variants for which CyTag attempts to account include the reduction of word-final /ai/ to /a/ or /ɛ/ in speech, reflected in the reduction of orthographic <-ai>, <-au>, and <-ae> to <-a> or <-e> by some writers. When it is unable to recognise a word with final <-a> or <-e>, CyTag searches for the same word, with the final letter replaced with each of <-ai>, <-au>, and <-ae>. For example, given the token *chware*, CyTag will search for *chwarai*, *chwarau*, and *chwarae*, enabling it to attribute the token correctly to the Welsh lemma *chwarae*. Apart from dialectal variants, CyTag also considers whether an unidentified token featuring the medial sequence <-nn-> exists with a single <-n-> instead. Given the token *penderfynnu*, CyTag will correctly identify a non-standard spelling (or spelling error) belonging to the lemma *penderfynnu* ("to decide"). We found the substitution of <nn> for <n> to be a common and systematic error, which rarely (if ever) resulted in a token whose standardized form could not be unambiguously identified, while other spelling errors identified in the data were often inconsistent or resulted in tokens that could not easily be transformed back into the standard orthographic form.

Other considerations

Welsh demonyms and language names are tagged by CyTag as common nouns, not proper nouns. As a result, *Cymraeg* ('the Welsh Language') is listed as the most frequent common noun in the corpus. Although they are not strictly common nouns, such words differ from personal names, brand names, and other types of proper noun in that they have single, stable referents which are generally widely-known to the language community. Compare this with, say, the masculine personal name *Gareth*, which can refer to any of numerous people who may or may not be well-known.

A final area affecting the reliability of the frequency lists is CyTag's processing of multiword sequences in Welsh. In particular, CyTag does not tag multiword prepositions such as *ar gyfer* ('for the purpose of', 'in preparation for') and *er mwyn* ('in order to', 'for the sake of') as single units, but rather treats them as a preposition followed by a noun. As a result, the lemmatized lists will show a higher frequency for the noun components of these sequences than would otherwise be expected. The noun *cyfer*, which forms the second element of *ar gyfer*, is the eighth most common noun lemma in the corpus, despite being used independently exceedingly rarely. In fact, it is unlikely that most users of contemporary Welsh would recognize the noun's meaning if it were used outside of the sequence *ar gyfer*. The lists must therefore be interpreted carefully, with awareness of the different functions open-class words can perform in Welsh multiword sequences.

4 Llyfryddiaeth | Bibliography

- Evison, J. (2010). What are the basics of analysing a corpus? Yn | In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122–135). Abingdon: Routledge.
- Kilgarriff, A. (1997). Using word frequency lists to measure corpus homogeneity and similarity between corpora. Yn | In *Proceedings of the 5th ACL workshop on very large corpora*, Beijing.
- Neale, S., Donnelly, K., Watkins, G., & Knight, D. (2018). Leveraging lexical resources and constraint grammar for rule-based part-of-speech tagging in Welsh. Yn | In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, European Language Resources Association (ELRA).
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Caergrawnt | Cambridge: Cambridge University Press.
- Piao, S., Rayson, P., Knight, D., & Watkins, G. (2018). Towards a Welsh semantic annotation system. In *Proceedings of the LREC (Language Resources Evaluation) 2018 conference*, Miyazaki, Japan.
- Piao, S., Rayson, P., Knight, D., Watkins, G., & Donnelly, K. (2017). Towards a Welsh semantic tagger: Creating lexicons for a resource poor language. Yn | In *Proceedings of the corpus linguistics 2017 conference*. Prifysgol Birmingham | University of Birmingham, DG | UK.

Atodiad | Appendix A Rhannau ymadrodd CorCenCC
The CorCenCC part-of-speech tagset

Rhan Ymadrodd POS	Tag	Disgrifiad	Description	Enghraifft Example	
Enw (Noun)	Egu	Enw gwrywaidd unigol	Noun, common masc. Singular	ci	
	Ebu	Enw benywaidd unigol	Noun, common fem. Singular	cath	
	Egll	Enw gwrywaidd lluosog	Noun, common masc. Plural	cŵn	
	Ebll	Enw benywaidd lluosog	Noun, common fem. Plural	cathod	
	Egbu	Enw gwrywaidd/benywaidd unigol		Noun, common masc./fem. Singular	meddalwedd

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
Y Fannod Ben- odol (The defi- nite article)	Egbl	Enw gwrywaidd/beny- waidd lluosog	Noun, common masc./fem. Plural	arferion
	Epg	Enw priod gwrywaidd	Noun, proper, masc.	Dafydd
	Epb	Enw priod benywaidd	Noun, proper, fem.	Esyllt
	Ep	Enw priod diryw	Noun, proper, gender unk.	Instagram
	YFB	Y Fannod Benodol	Article, definite	y, yr, 'r
Arddodiad (Preposition)	Arsym	Arddodiad syml	Preposition, uninflected	ar
	Ar1u	Arddodiad rhediadol, pers. 1af. unigol	Inflected preposition, 1st pers. Singular	arnaf
	Ar2u	Arddodiad rhediadol, 2il pers. unigol	Inflected preposition, 2nd pers. Singular	arnat

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Ar3gu	Arddodiad rhediadol, 3ydd pers. gwrywaidd unigol	Inflected preposition, 3rd pers. Masc. Singular	arno
	Ar3bu	Arddodiad rhediadol, 3ydd pers. benywaidd unigol	Inflected preposition, 3rd pers. Fem. Singular	arni
	Ar1ll	Arddodiad rhediadol, pers 1af. lluosog	Inflected preposition, 1st pers. Plural	arnom
	Ar2ll	Arddodiad rhediadol, 2il pers. lluosog	Inflected preposition, 2nd pers. Plural	arnoch
	Ar3ll	Arddodiad rhediadol, 3ydd pers. lluosog	Inflected preposition, 3rd pers. Plural	arnynt
Cysylltair (Conjunction)	Cyscyd	Cysylltair cydradd	Coordinating conjunction	a, neu
	Cysis	Cysylltair isradd	Subordinating conjunction	ers, petai
Rhifair (Numeral)	Rhifol	Rhifair - rhifolyn (0-10)	Numeral, cardinal (0-10)	saith, dau, dwy

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Rhifold	Rhifair - rhifolyn degol (11+)	Numeral, cardinal, decimal (11+)	un deg un
	Rhifolt	Rhifair - rhifolyn traddodiadol (11+)	Numeral, cardinal, traditional (11+)	ugain, dau ar hugain, dwy ar hugain
	Rhitref	Rhifair - trefnolyn (0-10)	Numeral, ordinal (0-10)	seithfed, ped-werydd, ped-waredd
	Rhitrefd	Rhifair - trefnolyn degol (11+)	Numeral, ordinal decimal (11+)	un deg wythfed
	Rhitref	Rhifair - trefnolyn traddodiadol (11+)	Numeral, ordinal traditional (11+)	deunawfed
Ansoddair (Adjective)	Anscadu	Ansoddair unigol	Adjective, positive, singular	byr

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Anscadbu	Ansoddair cadarnhaol benywaidd unigol	Adjective, positive, fem. Singular	ber
	Anscadll	Ansoddair cadarnhaol lluosog	Adjective, positive, plural	byrion
	Anscyf	Ansoddair cyfartal	Adjective, equative	byrred
	Anscym	Ansoddair cymharol	Adjective, comparative	byrrach
	Anseith	Ansoddair eithaf	Adjective, superlative	byrraf
Adferf (Adverb)	Adf	Adferf	Adverb	adref, felly
Berf (Verb)	Be	Berfenw	Verb noun	ariannu
	Bpresdyflu	Berf presennol/dyfodol, pers. 1af. unigol	Inflected verb, present, 1st pers. Singular	ariannaf
	Bpresdyfzu	Berf presennol/dyfodol, 2il pers. unigol	Inflected verb, present, 2nd pers. Singular	arienni

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Bpresdyf3u	Berf presennol/dyfodol, 3ydd pers. unigol	Inflected verb, present, 3rd pers. Singular	arianna
	Bpresdyf1ll	Berf presennol/dyfodol, pers. 1af. lluosog	Inflected verb, present, 1st pers. Plural	ariannwn
	Bpresdyf2ll	Berf presennol/dyfodol, 2il pers. lluosog	Inflected verb, present, 2nd pers. Plural	ariennwch
	Bpresdyf3ll	Berf presennol/dyfodol, 3ydd pers. lluosog	Inflected verb, present, 3rd pers. Plural	ariannant
	Bpresdyfamhers	Berf presennol/dyfodol amhersonol	Inflected verb, present, im- personal	ariennir
	Bpresdyf3perth	Berf presennol/dyfodol, 3ydd pers. perthynnol	Inflected verb, present, 3rd pers. Singular relative	sy, sydd
	Bgorb1u	Berf gorberfaith, pers. 1af. unigol	Inflected verb, pluperfect, 1st pers. Singular	arianaswn

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
	Bgorb2u	Berf gorberffaith, 2il pers. unigol	Inflected verb, pluperfect, 2nd pers. Singular	arianasit
	Bgorb3u	Berf gorberffaith, 3ydd pers. unigol	Inflected verb, pluperfect, 3rd pers. Singular	arianasai
	Bgorb1ll	Berf gorberffaith, pers. 1af. lluosog	Inflected verb, pluperfect, 1st pers. Plural	arianasem
	Bgorb2ll	Berf gorberffaith, 2il pers. lluosog	Inflected verb, pluperfect, 2nd pers. Plural	arianasech
	Bgorb3ll	Berf gorberffaith, 3ydd pers. lluosog	Inflected verb, pluperfect, 3rd pers. Plural	arianasent
	Bgorbambhers	Berf gorberffaith, amher- sonol	Inflected verb, pluperfect, impersonal	arianasid
	Bamhen1u	Berf amhenodol, pers. 1af. unigol	Inflected verb, imperfect, 1st pers. Singular	ariannwn

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Bamhen2u	Berf amhenodol, 2il pers. unigol	Inflected verb, imperfect, 2nd pers. Singular	ariannet
	Bamhen3u	Berf amhenodol, 3ydd pers. unigol	Inflected verb, imperfect, 3rd pers. Singular	ariannai
	Bamhen1ll	Berf amhenodol, pers. 1af. lluosog	Inflected verb, imperfect, 1st pers. Plural	ariannem
	Bamhen2ll	Berf amhenodol, 2il pers. lluosog	Inflected verb, imperfect, 2nd pers. Plural	ariannech
	Bamhen3ll	Berf amhenodol, 3ydd pers. lluosog	Inflected verb, imperfect, 3rd pers. Plural	ariannent
	Bamhenamhers	Berf amhenodol, amher- sonol	Inflected verb, imperfect, impersonal	ariennid
	Bgorff1u	Berf gorffennol, pers. 1af. unigol	Inflected verb, past, 1st pers. Singular	ariennais

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
	Bgorff2u	Berf gorffennol, 2il pers. unigol	Inflected verb, past, 2nd pers. Singular	ariennaist
	Bgorff3u	Berf gorffennol, 3ydd pers. unigol	Inflected verb, past, 3rd pers. Singular	ariannodd
	Bgorff1ll	Berf gorffennol, pers. 1af. lluosog	Inflected verb, past, 1st pers. Plural	arianasom
	Bgorff2ll	Berf gorffennol, 2il pers. lluosog	Inflected verb, past, 2nd pers. Plural	arianasoch
	Bgorff3ll	Berf gorffennol, 3ydd pers. lluosog	Inflected verb, past, 3rd pers. Plural	arianasant
	Bgorffamhers	Berf gorffennol, amhersonol	Inflected verb, past, imper-sonal	ariannwyd
	Bgorffisef	Berf gorffennol, sefydlog	Inflected verb, past, invariant	ddaru

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Bgorch2u	Berf gorchmynnol, 2il pers. unigol	Inflected verb, imperative, 2nd pers. Singular	arianna
	Bgorch3u	Berf gorchmynnol, 3ydd pers. unigol	Inflected verb, imperative, 3rd pers. Singular	arianned
	Bgorch1ll	Berf gorchmynnol, pers. 1af. lluosog	Inflected verb, imperative, 1st pers. Plural	ariannwn
	Bgorch2ll	Berf gorchmynnol, 2il pers. lluosog	Inflected verb, imperative, 2nd pers. Plural	ariennwch
	Bgorch3ll	Berf gorchmynnol, 3ydd pers. lluosog	Inflected verb, imperative, 3rd pers. Plural	ariannt
	Bgorchamhers	Berf gorchmynnol, amher-sonol	Inflected verb, imperative, impersonal	arianner
	Bdibdyf1u	Berf dibynnol dyfodol, pers. 1af. unigol	Inflected verb, subjunctive, 1st pers. Singular	ariannwyf

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
	Bdibdyf2u	Berf dibynnol dyfodol, 2il pers. unigol	Inflected verb, subjunctive, 2nd pers. Singular	ariennych
	Bdibdyf3u	Berf dibynnol dyfodol, 3ydd pers. unigol	Inflected verb, subjunctive, 3rd pers. Singular	arianno
	Bdibdyf1ll	Berf dibynnol dyfodol, pers. 1af. lluosog	Inflected verb, subjunctive, 1st pers. Plural	ariannom
	Bdibdyf2ll	Berf dibynnol dyfodol, 2il pers. lluosog	Inflected verb, subjunctive, 2nd pers. Plural	ariannoch
	Bdibdyf3ll	Berf dibynnol dyfodol, 3ydd pers. lluosog	Inflected verb, subjunctive, 3rd pers. Plural	ariannont
	Bdibdyfamhers	Berf dibynnol dyfodol, amhersonol	Inflected verb, subjunctive, impersonal	arianner
	Bamod1u	Berf amodol, pers. 1af. unigol	Verb, conditional, 1st pers. Singular	byddwn

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Bamod2u	Berf amodol, 2il pers. unigol	Verb, conditional, 2nd pers. Singular	byddit, byd-det
	Bamod3u	Berf amodol, 3ydd pers. unigol	Verb, conditional, 3rd pers. Singular	byddai
	Bamod1ll	Berf amodol, pers. 1af. lluosog	Verb, conditional, 1st pers. Plural	byddem
	Bamod2ll	Berf amodol, 2il pers. lluosog	Verb, conditional, 2nd pers. Plural	byddech
	Bamod3ll	Berf amodol, 3ydd pers. lluosog	Verb, conditional, 3rd pers. Plural	byddent, bydden
	Bamodamhers	Berf amodol, amhersonol	Verb, conditional, imper-sonal	byddid
Rhagenw (Pronoun)	Rhapers1u	Rhagenw personol pers. 1af. unigol	Pronoun, personal, 1st pers. Singular	mi, myfi

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Rhapers2u	Rhagenw personol, 2il pers. unigol	Pronoun, personal, 2nd pers. Singular	ti, tydi
	Rhapers3gu	Rhagenw personol, 3ydd pers. gwrywaidd unigol	Pronoun, personal, 3rd pers. Masc. Singular	ef, efe
	Rhapers3bu	Rhagenw personol, 3ydd pers. benywaidd unigol	Pronoun, personal, 3rd pers. Fem. Singular	hi, hyhi
	Rhapers1ll	Rhagenw personol, pers. 1af. lluosog	Pronoun, personal, 1st pers. Plural	ni, nyni
	Rhapers2ll	Rhagenw personol, 2il pers. lluosog	Pronoun, personal, 2nd pers. Plural	chi, chwychwi
	Rhapers3ll	Rhagenw personol, 3ydd pers. lluosog	Pronoun, personal, 3rd pers. Plural	hwy, hwynt-hwy
	Rhadib1u	Rhagenw dibynnol, pers. 1af. unigol	Pronoun, dependent, 1st pers. Singular	fy, 'm

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Rhadib2u	Rhagenw dibynnol, 2il pers. unigol	Pronoun, dependent, 2nd pers. Singular	dy, 'th
	Rhadib3gu	Rhagenw dibynnol, 3ydd pers. gwrywaidd unigol	Pronoun, dependent, 3rd pers. Masc. Singular	ei, 'i
	Rhadib3bu	Rhagenw dibynnol, 3ydd pers. benywaidd unigol	Pronoun, dependent, 3rd pers. Fem. Singular	ei, 'i
	Rhadib1ll	Rhagenw dibynnol, pers. 1af. lluosog	Pronoun, dependent, 1st pers. Plural	ein, 'n
	Rhadib2ll	Rhagenw dibynnol, 2il pers. lluosog	Pronoun, dependent, 2nd pers. Plural	eich, 'ch
	Rhadib3ll	Rhagenw dibynnol, 3ydd pers. lluosog	Pronoun, dependent, 3rd pers. Plural	eu, 'u
	Rhamedd1u	Rhagenw meddiannol, pers. 1af. unigol	Pronoun, possessive, 1st pers. Singular	eiddof

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Rhamedd2u	Rhagenw meddiannol, 2il pers. unigol	Pronoun, possessive, 2nd pers. Singular	eiddot
	Rhamedd3gu	Rhagenw meddiannol, 3ydd pers. gwrywaidd unigol	Pronoun, possessive, 3rd pers. Masc. Singular	eiddo
	Rhamedd3bu	Rhagenw meddiannol, 3ydd pers. benywaidd unigol	Pronoun, possessive, 3rd pers. Fem. Singular	eiddi
	Rhamedd1ll	Rhagenw meddiannol, pers. 1af. lluosog	Pronoun, possessive, 1st pers. Plural	eiddom
	Rhamedd2ll	Rhagenw meddiannol, 2il pers. lluosog	Pronoun, possessive, 2nd pers. Plural	eiddoch
	Rhamedd3ll	Rhagenw meddiannol, 3ydd pers. lluosog	Pronoun, possessive, 3rd pers. Plural	eiddynt
	Rhacys1u	Rhagenw cysylltiol pers. 1af. unigol	Pronoun, conjunctive, 1st pers. Singular	minnau

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Rhacys2u	Rhagenw cysylltiol, 2il pers. unigol	Pronoun, conjunctive, 2nd pers. Singular	tithau
	Rhacys3gu	Rhagenw cysylltiol, 3ydd pers. gwrywaidd unigol	Pronoun, conjunctive, 3rd pers. Masc. Singular	yntau
	Rhacys3bu	Rhagenw cysylltiol, 3ydd pers. benywaidd unigol	Pronoun, conjunctive, 3rd pers. Fem. Singular	hithau
	Rhacys1ll	Rhagenw cysylltiol, pers. 1af. lluosog	Pronoun, conjunctive, 1st pers. Plural	ninnau
	Rhacys2ll	Rhagenw cysylltiol, 2il pers. lluosog	Pronoun, conjunctive, 2nd pers. Plural	chithau
	Rhacys3ll	Rhagenw cysylltiol, 3ydd pers. lluosog	Pronoun, conjunctive, 3rd pers. Plural	hwythau
	Rhagof	Rhagenw gofynnol	Pronoun, interrogative	pa, pwy, ble, beth

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Rhadangg	Rhagenw dangosol gwry-waidd	Pronoun, demonstrative, masc.	hwn
	Rhadangb	Rhagenw dangosol benywaidd	Pronoun, demonstrative, fem.	hon
	Rhadangd	Rhagenw dangosol diryw	Pronoun, demonstrative, neutral.	hyn
	Rhadangll	Rhagenw dangosol lluosog	Pronoun, demonstrative, pl.	rhain
	Rhaperth	Rhagenw perthynol	Pronoun, relative	a, y
	Rhaatb	Rhagenw atblygol	Pronoun, reflexive	hun, hunan, hunain
	Rhacil	Rhagenw cilyddol	Pronoun, reciprocal	gilydd
	Rhaamh	Rhagenw amhendant	Pronoun, indefinite	rhai
Banodol (Determiner)	Bancynnarnar	Banodol Syml (math cynnar)	Simple Determiner (early type)	holl

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
Bandang	Bandang	Banodol Dangosol	Demonstrative Determiner	yma
	Bandangb	Banodol Dangosol Beny-waidd	Demonstrative Determiner (feminine)	hon
	Bandangd	Banodol Dangosol Diryw	Demonstrative Determiner (neuter)	yna
	Bandangg	Banodol Dangosol Gwry-waidd	Demonstrative Determiner (masculine)	hwn
Bangof	Bangof	Banodol Gofynnol	Interrogative Determiner	pa
	Banhwyr	Banodol Syml (math hwyr)	Simple Determiner (late type)	amryw
Banmedd1ll	Banmedd1ll	Banodol Meddiannol Pers 1af. Lluosog	Possessive Determiner (pl., 1st Per)	ein
	Banmedd1u	Banodol Meddiannol Pers 1af. Unigol	Possessive Determiner (sing., 1st Per)	fy

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
	Banmedd2ll	Banodol Meddiannol 2il. Pers Lluosog	Possessive Determiner (pl., 2nd Per)	eich
	Banmedd2u	Banodol Meddiannol 2il. Pers Unigol	Possessive Determiner (sing., 2nd Per)	dy
	Banmedd3bu	Banodol Meddiannol 3ydd. Pers Benywaidd Unigol	Possessive Determiner (sing., 3rd Per, Fem.)	ei
	Banmedd3gu	Banodol Meddiannol 3ydd. Pers Gwrywaidd Unigol	Possessive Determiner (sing., 3rd Per, Masc.)	ei
	Banmedd3ll	Banodol Meddiannol 3ydd. Pers Lluosog	Possessive Determiner (pl., 3rd Per)	eu
	Banmeint	Banodol Meintiol	Quantifier	ambell
Unigryw (Unique)	Uneg	Unigryw - geirynnau ne- gyddol	Unique - negative particles	dim, ni
	Ucad	Unigryw - geirynnau cadarnhaol	Unique - affirmative parti- cles	mi, fe

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
	Ugof	Unigryw – geirynnau go-fynnol	Unique – interrogative particles	a
	Utra	Unigryw – geiryn traethi-adol	Unique – predicative particle	yn
	Ublaen	Unigryw – blaenddodiad	Unique – prefix	ôl-, gwrth-
Ebychiad (Interjection)	Ebych	Ebychiad	Interjection	oo, aa, hei
Gweddilliol (Other)	Gwest	Gweddilliol – estron	Foreign word	anyway
	Gwfform	Gweddilliol – fformiwla	Formula	9+x=y, E=mc ²
	Gwsym	Gweddilliol – symbol	Symbol	€
	Gwacr	Gweddilliol – acronym	Acronym	GIG
	Gwtalf	Gweddilliol – talfyriad	Abbreviation	Cyf.
	Gwdig	Gweddilliol – digid	Digit	1, 99, 2000

Rh. Ym. POS	Tag	Disg.	Desc.	Engl. Ex.
Aralnodïad (Punctuation)	Gwlyth	Gweddilliol - llythyren	Letter	A, B, C, Ch, a, b, c, ch
	Gwann	Gweddilliol - annos- barthedig	Unclassified	
	Ymadr	Ymadrodd	Phrase	dwnim (= nid wn i ddim)
	Anon	Gair/geiriau hanonymeiddio	Anonymized word(s)	
	Ardt	Marc Atalnodi terfynol	Punctuation mark, final	. ? !
	Atdcan	Marc Atalnodi canolig	Punctuation mark, medial	, ; : - (dash/linell doriad)
	Atdchw	Marc Atalnodi, chwith	Punctuation mark, left	([
	Atdde	Marc Atalnodi, de	Punctuation mark, right])

Rh. Ym. POS	Tag	Disg.	Desc.	Engh. Ex.
	Arcys	Marc Atalnod, cysylltnod	Punctuation mark, hyphen	- (as in/fel yn 'gor-rifo')
	Ardyf	Marc Atalnod, dyfynnod	Punctuation mark, quotation	“ ” ’ ’ ’

Atodiad | Appendix B Trefn yr wyddor Gymraeg

Welsh alphabetical ordering

Yr wyddor Gymraeg a ddefnyddir er mwyn trefnu'r rhestrï geiriau yn Atodiad C. Rhestrir llafariad a chanddynt acenion ar ôl bôn y lythyren, felly mae <Â> yn dilyn <A> ac <Ŵ> yn dilyn <W>.

Welsh alphabetical ordering is used for the alphabetic wordlists in Appendix C. Accented vowels are listed directly after their base letter form, so <Â> follows <A> and <Ŵ> follows <W>. The Welsh alphabet is ordered as follows:

A	Ll
B	M
C	N
Ch	O
D	P
Dd	Ph
E	R
F	Rh
Ff	S
G	T
Ng	Th
H	U
I	W
J	Y
L	

Atodiad | Appendix C Rhestri amllder

Frequency lists

Cyflwyniad

Mae'r rhestri yn yr Atodiad hwn yn darparu samplau cyfleus o ddata amllder CorCenCC. Mae yna hefyd set o daenlenni'n gysylltiedig â'r ddogfen hon sy'n cynnwys yr holl ddata yn y ffurfiau a ganlyn:

- Rhestr amllder o'r corpws llawn wedi'i lemateiddio, gyda thaenlenni ar wahân wedi'u trefnu yn ôl safle a'r wyddor fesul lema. (Enw ffeil *corcenclemmas.xlsx*.)
- Rhestr o'r holl eirffurfiau yn y corpws llawn, gyda thaenlenni ar wahân wedi'i threfnu yn ôl safle a'r wyddor. (Enw ffeil *corcencwords.xlsx*.) Mae'r rhestri lema a geiriau yn rhoi gwybodaeth am amllder crai eitem ac amllder y filiwn yn y corpws llawn, yn ogsytal â'i safle ac amllder crai yn yr isgorpora ysgrifenedig, llafar ac iaith electronig.
- Rhestr o eiriau yn ôl bandiau, sy'n rhestru 5,000 o'r geiriau dosbarth agored mwyaf cyffredin yn y corpws llawn, wedi'u rhannu yn daenlenni yn cynnwys bandiau o 500 gair yr un. Rhagwelwn y bydd y rhain o ddiddordeb arbennig o ran lefelu deunyddiau pedagogaid ac asesiad addysgiadol. (Enw ffeil *corcenc_banded.xlsx*.)

Gellir dod o hyd i ragor o fanylion am y tagiau rhan ymadrodd a ddefnyddir yn y rhestri wedi'u lemateiddio yn Atodiad A.

Cyflwynir pedair rhestr ar ddeg yn Atodiad C:

- **C.1 – 100 gair uchaf yn CorCenCC:** yn rhestru'r 100 gair uchaf yn y corpws llawn yn nhrefn eu hamlder, gan roi amllder crai pob gair yn y

corpws cyfan, amllder y filiwn yn y corpws cyfan, a threfn safle (yn ôl amllder) yn yr is-gorpora ysgrifenedig, llafar ac iaith electronig.

- **C.2 – 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor):** yn rhestru'r 1000 gair uchaf yn y corpws llawn yn nhrefn yr wyddor.
- **C.3 – 100 lema uchaf yn CorCenCC:** yn rhestru'r 100 lema uchaf yn y corpws llawn yn nhrefn eu hamlder, gan roi rhan ymadrodd pob lema, amllder crai yn y corpws cyfan, amllder y filiwn yn y corpws cyfan, a threfn safle (yn ôl amllder) yn yr is-gorpora ysgrifenedig, llafar ac iaith electronig.
- **C.4 – 1000 lema uchaf yn CorCenCC (rhestr yn ôl yr wyddor):** yn rhestru'r 1000 lema uchaf yn y corpws llawn yn nhrefn yr wyddor, gan gynnwys tag rhan ymadrodd pob lema.
- **C.5 – 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig):** yn rhestru'r 100 lema uchaf yn y corpws, gan gyfrif enwau, berfau, ansoddeiriau ac adferfau yn unig. Rhestrir lemâu yn nhrefn eu hamlder, gan roi rhan ymadrodd pob lema, amllder crai yn y corpws cyfan ac amllder y filiwn (wedi'u cyfrifo ar sail cyfanswm y geiriau dosbarth agored yn unig, nid cyfanswm y corpws llawn).
- **C.6 – 1000 gair uchaf yn CorCenCC (geiriau dosbarth agored yn unig; rhestr yn ôl yr wyddor):** yn rhestru'r 1000 gair dosbarth agored uchaf yn y corpws llawn yn nhrefn yr wyddor. (Fel yn yr holl restr geiriau, gall cofnodion gynrychioli mwy nag un categori rhan ymadrodd; fodd bynnag, categorïau dosbarth agored yn unig a gyfrifwyd. Golyga hyn, er enghraifft, fod y tocyn dim yn cynrychioli cyfrif cronus yr

achosion lle cafodd y tocyn ei dagio yn enw neu'n adferf yn unig, gan ddisytyru'r rhai lle cafodd ei dagio fel banodolyn.)

- **C.7 – 500 enw uchaf yn CorCenCC**
- **C.8 – 500 berf uchaf yn CorCenCC**
- **C.9 – 500 ansoddair uchaf yn CorCenCC**
- **C.10 – 50 adferf uchaf yn CorCenCC:** mae'r cyfrif geiriau llai am hyn a'r rhestr nesaf yn adlewyrchu'r ffaith fod llai o fathau gwahanol yn y categorïau hyn.
- **C.11 – 50 ebychiad uchaf yn CorCenCC**
- **C.12 – 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC:** mae'r rhestr hon a'r ddwy nesaf yn cynnwys y 100 gair uchaf yn is-gorpora CorCenCC. Rhestrï y geiriau yn nhrefn eu safle, ac wedyn eu hamlder crai yn yr is-gorpws penodol, amlder y fliwn ymhlith yr holl docynnau dosbarth agored yn y corpws llawn, ac amlder y fliwn ymhlith yr holl docynnau yn y corpws llawn.
- **C.13 – 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC**
- **C.14 – 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC**

Defnyddir y drefn Gymraeg (gw. Atodiad B.) ar gyfer pob rhestr yn ôl yr wyddor yn Atodiad C.

Introduction

The lists in this Appendix provide convenient samples of the CorCenCC frequency data; an accompanying set of spreadsheets contains the full data in the following forms:

- A lemmatized list of the full corpus, with separate worksheets sorted by rank and alphabetically by lemma. (Filename *corcencclmmas.xlsx*.)
- A word list of all word forms in the full corpus, with separate worksheets sorted by rank and alphabetically. (Filename *corcencc_words.xlsx*.) Both lemma and word lists provide information about an item's raw frequency and frequency per million in the full corpus, as well as its rank and raw frequency in the written, spoken, and e-language subcorpora.
- A banded wordlist, listing 5,000 of the most common open-class words in the full corpus, split into worksheets containing bands of 500 words each. We anticipate that these will be of particular interest for leveling pedagogical materials and for educational assessment. (Filename *corcencc_words.xlsx*.)

More detail about the part-of-speech tags used in the lemmatized lists may be found in Appendix A.

Fourteen lists are presented in Appendix C:

- **C.1 – Top 100 words in CorCenCC:** lists the top 100 words in the full corpus in order of frequency, giving each word's raw frequency in the entire corpus, frequency per million in the entire corpus, and rank order (by frequency) in the written, spoken, and e-language sub-corpora.

- **C.2 – Top 1000 words in CorCenCC (alphabetic list):** lists the top 1000 words in the full corpus in alphabetical order.
- **C.3 – Top 100 lemmas in CorCenCC:** lists the top 100 lemmas in the full corpus in order of frequency, giving each lemma’s part of speech, raw frequency in the entire corpus, frequency per million in the entire corpus, and rank order (by frequency) in the written, spoken, and e-language sub-corpora.
- **C.4 – Top 1000 lemmas in CorCenCC (alphabetic list):** lists the top 1000 lemmas in the full corpus in alphabetical order, including each lemma’s part-of-speech tag.
- **C.5 – Top 100 lemmas in CorCenCC (open-class words only):** lists the top 100 lemmas in the corpus, counting only nouns, verbs, adjectives, and adverbs. Lemmas are listed in order of frequency, giving each lemma’s part of speech, raw frequency in the entire corpus, and frequency per million (calculated based on the total of open-class words only, not the full corpus total).
- **C.6 – Top 1000 words in CorCenCC (open-class words only; alphabetic list):** lists the top 1000 open-class words in the full corpus in alphabetical order. (As in all the word lists, entries may represent more than one part-of-speech category; however, only open-class categories were counted. This means that, for example, the token *dim* represents only the cumulative count of cases where the token was tagged as a noun or an adverb, disregarding those where it was tagged as a determiner.)
- **C.7 – Top 500 nouns in CorCenCC**

- **C.8 – Top 500 verbs in CorCenCC**
- **C.9 – Top 500 adjectives in CorCenCC**
- **C.10 – Top 50 adverbs in CorCenCC:** the smaller word count for this and the next list reflects the fact that there are fewer distinct types in these categories.
- **C.11 – Top 50 interjections in CorCenCC**
- **C.12 – Top 100 open-class words in the written component of CorCenCC:** this and the following two lists contain the top 100 words in the CorCenCC sub-corpora. The words are listed in rank order, followed by their raw frequency in the given subcorpus, frequency per million amongst all open-class tokens in the subcorpus, frequency per million amongst all open-class tokens in the full corpus, and frequency per million amongst all tokens in the full corpus.
- **C.13 – Top 100 open-class words in the spoken component of CorCenCC**
- **C.14 – Top 100 open-class words in the electronic component of CorCenCC**

Welsh ordering is used for the alphabetic lists in Appendix C – see Appendix B for details.

C.1 100 gair uchaf yn CorCenCC
 Top 100 words in CorCenCC

Safle (rank)	Gair (word)	Cyfanswm Crai (raw total)	Cyfanswm y filiwn (per-million total)	Safle (ysg) (wr. rank)	Safle (lla) (sp. rank)	Safle (ele) (el. rank)
1	yn	520106	48785.51	1	1	1
2	i	352014	33018.62	3	3	3
3	y	347839	32627.01	2	5	2
4	a	274199	25719.64	5	6	5
5	'r	267688	25108.91	4	7	4
6	o	251858	23624.07	6	4	6
7	'n	234203	21968.05	7	2	8
8	ar	152685	14321.73	8	11	7
9	mae	146071	13701.34	10	8	9

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
10	yr	90978	8533.66	11	22	11
11	ei	88739	8323.64	9	54	12
12	ac	84568	7932.41	12	39	10
13	wedi	69108	6482.27	13	30	14
14	am	68381	6414.08	14	35	13
15	ni	59541	5584.90	30	9	24
16	ond	58196	5458.74	16	17	17
17	bod	55720	5226.49	19	19	16
18	gan	46282	4341.21	17	107	15
19	oedd	45471	4265.14	15	37	27
20	un	45302	4249.29	21	27	22
21	fel	43678	4096.96	23	28	23
22	â	40316	3781.61	18	97	19
23	mewn	40288	3778.98	24	53	20

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
24	chi	39222	3678.99	52	12	44
25	nhw	37755	3541.39	41	15	60
26	sy	36505	3424.14	29	45	21
27	eu	35375	3318.15	25	139	18
28	yw	35022	3285.03	28	48	25
29	'i	34659	3250.99	22	46	37
30	fod	33838	3173.98	26	64	28
31	ma	32475	3046.13	215	10	264
32	ti	31310	2936.85	56	16	102
33	hyn	31145	2921.37	32	56	29
34	gyda	31020	2909.65	31	66	26
35	na	30490	2859.94	43	25	49
36	neu	29971	2811.25	33	61	30
37	ddim	29898	2804.41	40	29	48

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
38	roedd	29321	2750.29	20	197	36
39	wrth	28461	2669.62	27	81	38
40	iawn	27922	2619.06	44	33	51
41	cael	27614	2590.17	39	55	31
42	mynd	25493	2391.22	51	26	81
43	hi	25230	2366.55	34	42	78
44	fi	24084	2259.06	78	23	99
45	sydd	24041	2255.03	47	68	32
46	ie	23918	2243.49	835	14	1016
47	e	23742	2226.98	53	24	140
48	ymm	23178	2174.08	22237	13	20532
49	ôl	22886	2146.69	35	78	45
50	beth	22158	2078.40	57	43	58
51	at	21206	1989.11	36	106	42

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
52	dw	20972	1967.16	184	21	155
53	'na	20888	1959.28	151	20	483
54	fe	20592	1931.51	48	49	75
55	hefyd	20552	1927.76	50	73	47
56	fy	20304	1904.50	38	163	40
57	os	20212	1895.87	81	51	55
58	yng	19894	1866.04	45	184	33
59	hynny	19706	1848.41	42	82	57
60	bydd	19688	1846.72	46	109	39
61	dim	19586	1837.15	58	52	64
62	'di	18867	1769.71	474	18	2110
63	da	18584	1743.16	118	36	107
64	newydd	18343	1720.56	64	129	34
65	ym	18341	1720.37	55	70	56

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
66	er	18178	1705.08	37	2890	43
67	yma	17981	1686.60	90	63	53
68	ein	17617	1652.46	66	167	35
69	felly	17021	1596.56	75	59	71
70	o'n	16391	1537.46	135	34	226
71	gymraeg	16116	1511.67	69	180	41
72	lle	15886	1490.09	59	79	74
73	gael	15651	1468.05	67	154	50
74	i'w	15201	1425.84	49	254	59
75	cymraeg	14819	1390.01	80	164	52
76	mi	14808	1388.98	62	100	65
77	arall	14675	1376.50	63	90	76
78	gyfer	14547	1364.50	68	234	54
79	iaith	14378	1348.64	54	194	63

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
80	dod	14226	1334.39	88	62	115
81	dros	13979	1311.22	70	128	62
82	pan	13739	1288.71	60	95	96
83	hwn	13703	1285.33	79	102	68
84	meddwl	13449	1261.50	140	57	111
85	dwi	13349	1252.13	89	120	61
86	so	13250	1242.84	758	32	1138
87	yna	13066	1225.58	86	75	105
88	hyd	12758	1196.69	65	190	67
89	nid	12619	1183.65	61	251	69
90	eich	12576	1179.62	154	245	46
91	neud	12439	1166.77	590	38	685
92	yyy	12277	1151.57	23003	31	19533
93	be	12208	1145.10	208	44	288

<i>safle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
94	wedyn	11890	1115.27	169	50	296
95	bob	11707	1098.11	74	121	98
96	rhaid	11584	1086.57	76	112	104
97	mwy	11487	1077.47	123	123	66
98	fydd	11465	1075.41	87	126	86
99	bach	11430	1072.12	98	80	135
100	gweld	11102	1041.36	114	83	125

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)

Top 1000 words in CorCenCC (alphabetic list)

a	angen	anodd
â	angenrheidiol	ansawdd
ââ	ah	ar
ac	ai	arall
achos	ail	arbennig
adael	alw	archifau
adeg	all	ardal
adeilad	allai	ardaloedd
adeiladu	allan	arfer
adnoddau	allweddol	arglwydd
adran	am	arian
adroddiad	ambell	ariannol
addas	amdano	ar-lein
addysg	aml	arni
aelod	amlwg	arno
aelodau	amrywiaeth	aros
aeth	amser	arwain
afon	aneglur	arweinydd
ag	anfon	astudio
agor	anffodus	at
agored	anifeiliaid	ateb
agos	annog	ati

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

athro	blynyddol	byddwn
awdur	bo	bynnag
awgrymu	bob	byth
awr	bobl	byw
awyr	bobol	bywyd
babi	bod	cadw
bach	bore	cael
bai	bosib	cafodd
barn	braf	ca'l
barod	brif	cam
bawb	broblem	camau
be	bron	canol
beidio	broses	canu
bell	bryd	capel
bellach	brysur	car
ben	bu	cartref
beth	busnes	casgliad
bethau	bwrdd	casglu
blaen	bwyd	cau
blaenorol	bwysig	cawn
blaid	bwyta	cefn
blant	byd	cefnidir
ble	bydd	cefnogaeth
blog	byddai	cefnogi
blwyddyn	bydde	ceir
blynyddoedd	byddwch	ceisio
		enedlaethol
		cerdded

cerddoriaeth	cychwyn	cyn
cig	cyfan	cynhyrchu
clwb	cyfarfod	cynllun
clywed	cyfeiriad	cynnal
coch	cyfieithu	cynnig
codi	cyfle	cynnwys
coed	cyflwyno	cynta
cofo	cyfnod	cyntaf
cofnod	cyfrannu	cynulliad
cofnodion	cyfres	cyrraedd
coleg	cyfrwng	cysylltiadau
colli	cyfryngau	cysylltu
comisiynydd	cyffredin	cysylltwch
côr	cyffredinol	cytuno
corff	cyngor	cyw
credu	cyhoeddi	'ch
creu	cyhoeddus	chael
criw	cylch	chdi
croeso	cymaint	chi
cwcis	cymdeithas	chwaith
cwestiwn	cymdeithasol	chwarae
cwestiynau	cymorth	chwaraeon
cwm	cymraeg	chwe
cwmni	cymreig	chwech
cwrdd	cymry	chwerthin
cwrs	cymryd	chwilio
		da
		dach

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

dad	dewch	do
daeth	dewis	dod
dangos	'di	doedd
daith	di	does
dal	di-	dosbarth
dan	diddordeb	drafod
darllen	diddorol	draw
darn	digon	draws
darparu	digwydd	dref
data	digwyddiad	dro
datblygu	digwyddiadau	droi
dathlu	dilyn	dros
dau	dim	drws
daw	diogelu	drwy
dawel	diogelwch	du
de	diolch	duw
deall	dipyn	dw
debyg	disgwyl	dweud
dechrau	disgyblion	dwi
defnydd	diwedd	dŵr
defnyddio	diweddar	dwy
deg	diweddaraf	dy
derbyn	diwethaf	dydy
destun	diwrnod	dydd
d'eud	diwydiant	dyddiau
deud	diwylliant	dyfodol
		dylai
		dylid

dyma	ddwy	enwau
dyn	ddydd	enwedig
dyna	ddysgu	enwog
dynnu	e	er
dysgu	eang	eraill
dyw	e-bost	erbyn
dywedodd	edrych	erioed
dda	ef	ers
ddaeth	efallai	erthygl
ddarllen	efo	es
ddatblygu	effaith	eto
ddau	eglwys	eu
dde	eng	ewch
ddechrau	enghraifft	ewropeaidd
ddefnyddio	ei	fach
ddeg	eich	faint
dderbyn	eidion	falch
ddewis	ein	falle
ddigon	eisiau	fam
ddim	eisoës	fan
ddinas	eistedd	farchnad
ddiwedd	eisteddfod	farw
ddiweddar	eitha	fath
ddiwrnod	eleni	fawr
ddod	ennill	fe
ddweud	enw	feddwl
		fel
		felly

ferch	ffilm	ganrif
fewn	ffôn	gartref
fi	ffordd	geiriau
fideo	ffrind	geisio
flaen	ffrindiau	gellir
flwyddyn	ffurf	gêm
flynyddoedd	ffurflen	gen
fo	ffwrdd	genedlaethol
fod	ffyrdd	gennyich
fodd	ga	gennym
for	gadael	ger
fu	gadw	ges
fwy	gael	gilydd
fwyaf	gafodd	glir
fy	gair	glywed
fyd	ga'l	go
fydd	galw	gobeithio
fyddai	gall	godi
fyddwn	gallai	gofal
fynd	gallu	gofio
fyny	gallwch	gofyn
fyw	gallwn	gogledd
fywyd	gan	golwg
ffaith	ganddo	golygu
fferm	ganlyniad	gorau
ffermwyr	ganolfan	gorfod
		gorffen
		gorllewin

gosod	gwrando	gynnal
greu	gwreiddiol	gynnig
grŵp	gwrs	gynnwys
gwahanol	gwybod	gynta
gwaith	gwybodaeth	gyntaf
gwasanaeth	gwybodaeth	gynulleidfa
gwasanaethau	gwybodaeth	gyrraedd
gwbl	gŵyl	haf
gw'bod	gwylio	hanes
gwbod	gwyn	hanesyddol
gwefan	gyd	hanner
gweinidog	gyda	hapus
gweithio	gydag	haul
gweithredu	gyfan	hawdd
gweld	gyfer	hawl
gwell	gyfle	hawliau
gwerthu	gyflwyno	heb
gweud	gyfnod	heddiw
gwir	gyfres	heddlu
gwirionedd	gyfrifol	hefo
gwlad	gyfrwng	hefyd
gwleidyddol	gyffredinol	heibio
gwmpas	gymaint	helo
gwneud	gymdeithas	help
gwobrau	gymraeg	helpu
gŵr	gymryd	hen
	gymuned	heno
		her

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

hi	hynod	lawer
hir	hyrwyddo	lawr
hmm	hytrach	le
hoff	'i	lefel
hoffi	i	leol
holl	ia	lot
hollol	iaith	lyfrau
hon	iau	lywodraeth
honno	iawn	llafur
hun	iddi	llai
hunain	iddo	llais
hunan	iddyn	llall
hwn	iddynt	llaw
hwanna	ie	llawer
hwynnw	iechyd	llawn
hwnt	ieithyddol	lle
hwyl	ifanc	lleol
hwyr	is	llun
hybu	isdeitlau	lluniau
hyd	isio	llwybr
hyfryd	isod	llwyddo
hyfforddiant	i'w	llwyr
hyn	jest	llyfr
hŷn	jyst	llyfrau
hynna	lan	llyfrgell
hynny	law	llywodraeth
		'm
		'ma

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

ma	mi	nag
mae	mis	naill
maen	mlaen	natur
maent	mlynedd	naturiol
maes	mmm	nath
mai	mod	naw
mam	modd	nawr
man	mor	ne
manylion	môr	neb
map	mudiad	neges
mas	munud	neis
materion	mwy	nes
math	mwyaf	nesa
mawr	mwyn	nesaf
mawrth	mwynhau	neu
medru	myfyrwyr	neud
meddai	mynd	newid
meddwl	mynediad	newydd
megis	mynydd	newyddion
meic	'n	nhw
meithrin	'na	ni
merch	na	nid
merched	nabod	nifer
methu	nac	nod
mewn	nad	nodi
m-hm	nadolig	nofel
		nôl
		nos

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

noson	pam	popeth
o	pan	prif
ochr	papur	prifysgol
o'dd	paratoi	profiad
oddi	parhau	prosiect
oed	pawb	pryd
oedolion	pe	prynu
oedd	pecyn	pum
oeddwn	pedwar	pump
oen	pen	pwylgor
oes	penderynu	pwynt
ofn	penodol	pwysig
ofyn	person	phobl
ogystal	personol	'r
oh	perthynas	radio
oherwydd	petai	rai
ohonyn	peth	raid
ohonynt	pethau	ran
ol	pethe	reit
ôl	plaid	ro
olaf	plant	roedd
o'n	plentyn	roeddwn
ond	pob	roi
os	pobl	rôl
pa	pobol	rownd
paid	polisi	rwan
		rwy

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

rydych	rhyfel	siarad
rydym	rhyw	siaradwyr
rygbi	rhywbeth	sicr
ryw	rhywun	sicrhau
rywbeth	sa	sioe
rywun	sadwrn	siop
rhag	saesneg	sir
rhaglen	safbwynt	siŵr
rhaglenni	safle	so
rhagor	safon	sôn
rhai	safonau	sori
rhaid	sail	staff
rhain	sain	stori
rhan	saith	straeon
rhannu	sawl	stryd
rhedeg	sector	sul
rheoli	sef	sut
rhestr	sefydliad	swn
rheswm	sefydlu	sŵn
rhieni	sefyll	swydd
rhif	sefyllfa	swyddfa
rhifyn	senedd	swyddog
rhoi	sesiwn	swyddogol
rhowch	sgiliau	sy
rhwng	sgwennu	sydd
rhy	sgwrs	sylw
		sylwadau
		sylweddol

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

sylweddoli	to	unig
syml	torri	unigolion
symud	tra	unigryw
syniad	trafod	union
syniadau	trefnu	unrhyw
system	tri	unwaith
syth	trio	uwch
ta	tro	w
tair	troi	wahanol
taith	trwy	waith
taliad	tu	wan
talun	tua	wastad
tan	tuag	we
tân	ty	wedi
te	tŷ	wedyn
tebyg	tybed	wefan
teg	tymor	weinidog
teimlo	tynnu	weithiau
teithio	tywydd	weithio
teledu	'th	wel
testun	the	weld
teulu	'u	well
ti	uchel	werth
tîm	uchod	wir
tipyn	un	wlad
tir	undeb	wnaeth
		wnes
		wneud

C.2 1000 gair uchaf yn CorCenCC (rhestr yn ôl yr wyddor)
Top 1000 words in CorCenCC (alphabetic list)

wraig	ydy	ymweld
wreiddiol	ydych	ymwneud
wrth	ydyn	yn
wy	yng	yna
wybod	ynghyd	yndi
wybodaeth	ynghylch	yno
wych	ynglŷn	yr
wyneb	ym	ysbyty
wynebu	yma	ysgol
wyt	ymarfer	ysgolion
wyth	ymateb	ysgrifennu
wythnos	ymchwil	ysgrifennydd
y	ymddangos	ystafell
ychwanegol	ymgyrch	ystod
ychwanegu	ymhlith	ystyr
ychydig	ymlaen	ystyried
ydi	ymm	yw
ydw	ymuno	yyy

C.3 100 lema uchaf yn CorCenCC
Top 100 lemmas in CorCenCC

Safle (rank)	Lemma	Rhan Ymadrodd (part of speech)	Cyfanswm Crai (raw total)	Cyfanswm y filiwn (per-million total)	Safle (ysg) (wr. rank)	Safle (lla) (sp. rank)	Safle (ele) (el. rank)
1	y	YFB (def. art.)	667289	62591.15	1	3	1
2	bod	B (v.)	649744	60945.44	2	1	2
3	yn	Ar (prep.)	575333	53965.75	3	2	3
4	a	Cys (conj.)	306927	28789.50	4	6	4
5	i	Ar (prep.)	263411	24707.73	5	8	5
6	o	Ar (prep.)	210676	19761.23	6	9	6
7	yn	U (unique)	205094	19237.65	7	4	8
8	ar	Ar (prep.)	161132	15114.05	8	13	7
9	mi	Rha (pron.)	152335	14288.90	10	5	10

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/sum</i>	<i>c/sum yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
10	ei	Rha (pron.)	118243	11091.09	9	23	9
11	fo	Rha (pron.)	90167	8457.59	16	7	19
12	â	Ar (prep.)	77015	7223.94	11	26	12
13	wedi	U (unique)	76537	7179.11	13	16	14
14	am	Ar (prep.)	74454	6983.72	12	28	11
15	cael	B (v.)	64101	6012.62	15	29	15
16	gan	Ar (prep.)	58589	5495.60	14	55	13
17	ond	Cys (conj.)	58198	5458.92	17	20	18
18	ni	Rha (pron.)	54043	5069.19	36	10	27
19	hwn	Rha (pron.)	52799	4952.50	18	34	16
20	gwneud	B (v.)	50508	4737.61	23	19	21
21	mewn	Ar (prep.)	45005	4221.43	19	41	17
22	fel	Cys (conj.)	43712	4100.15	21	27	23
23	mynd	B (v.)	39718	3725.52	28	21	37

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/sum</i>	<i>c/sum yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
24	chi	Rha (pron.)	39547	3709.48	51	15	42
25	nhw	Rha (pron.)	38060	3570.00	41	17	59
26	fy	Rha (pron.)	35953	3372.36	22	57	24
27	gyda	Ar (prep.)	35855	3363.17	24	48	22
28	da	Ans (adj.)	34127	3201.08	42	24	32
29	ie	Adf (adv.)	32876	3083.74	612	11	602
30	dod	B (v.)	32755	3072.39	25	35	36
31	peth	Rha (pron.)	32051	3006.36	46	22	51
32	ti	Rha (pron.)	31830	2985.63	55	18	101
33	cymraeg	E (n.)	31784	2981.31	37	72	20
34	hwnnw	Rha (pron.)	31666	2970.24	27	36	39
35	wrth	Ar (prep.)	30882	2896.71	20	64	33
36	ymm	Ebych (interj.)	30122	2825.42	573	12	282
37	neu	Cys (conj.)	29971	2811.25	31	53	25

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/sum</i>	<i>c/sum yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
38	mae	B (v.)	27681	2596.45	445	14	689
39	mawr	Ans (adj.)	27552	2584.35	32	52	30
40	un	Rha (pron.)	26551	2490.46	40	42	40
41	hi	Rha (pron.)	26242	2461.48	30	38	85
42	gweld	B (v.)	26230	2460.35	35	47	44
43	at	Ar (prep.)	26070	2445.34	29	81	29
44	dim	E (n.)	24999	2344.89	45	39	47
45	a	Rha (pron.)	24892	2334.85	26	208	28
46	dweud	B (v.)	24611	2308.49	53	32	62
47	gallu	B (v.)	24167	2266.84	43	59	34
48	eu	Rha (pron.)	24121	2262.53	39	142	26
49	dau	Rhi (num.)	23854	2237.49	34	49	57
50	dim	Adf (adv.)	23016	2158.88	57	30	97
51	yma	Adf (adv.)	22188	2081.22	83	37	54

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
52	eu	Ban (det.)	21753	2040.41	33	172	35
53	rhoi	B (v.)	20705	1942.11	54	58	46
54	hefyd	Adf (adv.)	20653	1937.23	48	65	45
55	arall	Ans (adj.)	20629	1934.98	44	69	48
56	yna	Adf (adv.)	20513	1924.10	104	25	200
57	na	Rha (pron.)	20262	1900.56	60	43	69
58	er	Ar (prep.)	18846	1767.74	38	1262	38
59	trwy	Ar (prep.)	18753	1759.02	49	104	41
60	peth	E (n.)	18554	1740.35	78	46	73
61	newydd	Ans (adj.)	18343	1720.56	69	123	31
62	pobl	E (n.)	17378	1630.04	88	75	43
63	felly	Adf (adv.)	17021	1596.56	77	51	78
64	un	Ban (det.)	16954	1590.27	50	74	70
65	rhan	E (n.)	16260	1525.17	59	115	52

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/swm</i>	<i>c/swm yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
66	gwaith	E (n.)	16187	1518.33	66	99	53
67	blwyddyn	E (n.)	16058	1506.23	64	88	61
68	iaith	E (n.)	15946	1495.72	47	173	56
69	iawn	Adf (adv.)	15939	1495.06	76	62	80
70	bach	Ans (adj.)	15342	1439.07	61	71	88
71	pob	Ban (det.)	15079	1414.40	56	94	75
72	tros	Ar (prep.)	14951	1402.39	71	110	60
73	cyfer	E (n.)	14876	1395.36	70	196	49
74	mwy	Rha (pron.)	14685	1377.44	87	97	55
75	wedi	Ar (prep.)	14584	1367.97	117	40	198
76	plentyn	E (n.)	14485	1358.68	63	126	65
77	ysgol	E (n.)	13474	1263.85	58	141	81
78	byd	E (n.)	13417	1258.50	72	90	95
79	fy	Ban (det.)	13254	1243.21	62	215	68

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/sum</i>	<i>c/sum yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
80	ôl	Ans (adj.)	13146	1233.08	68	95	99
81	ein	Rha (pron.)	13133	1231.86	84	70	100
82	dydd	E (n.)	12934	1213.20	92	112	64
83	defnyddio	B (v.)	12890	1209.07	81	161	63
84	rhaid	E (n.)	12849	1205.23	75	84	102
85	ein	Ban (det.)	12824	1202.88	99	211	50
86	nid	U (unique)	12595	1181.40	65	213	74
87	y	Rha (pron.)	12513	1173.71	52	336	82
88	pan	Cys (conj.)	12421	1165.08	67	101	107
89	yyy	Ebych (interj.)	12277	1151.57	11050	31	9838
90	meddwl	B (v.)	12196	1143.97	133	56	141
91	so	Adf (adv.)	12103	1135.25	909	33	1103
92	wedyn	Adf (adv.)	11890	1115.27	187	45	346
93	ôl	E (n.)	11595	1087.60	86	143	83

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/sum</i>	<i>c/sum yf.</i>	<i>saf. ysg.</i>	<i>saf. lla.</i>	<i>saf. ele.</i>
94	yn	Adf (adv.)	11450	1074.00	85	82	129
95	sut	Rha (pron.)	10895	1021.94	126	127	77
96	fe	U (unique)	10886	1021.10	74	154	110
97	eich	Rha (pron.)	10881	1020.63	178	102	76
98	rŷ	E (n.)	10834	1016.22	93	151	89
99	dyna	Adf (adv.)	10724	1005.90	127	61	228
100	efo	Ar (prep.)	10683	1002.06	169	63	160

C.4 1000 lema uchaf yn CorCenCC (rhestr yn ôl yr wyddor)

Top 1000 lemmas in CorCenCC (alphabetic list)

a CYS (CONJ.)	aeth B (V.)	annibynnol ANS (ADJ.)
a RHA (PRON.)	afon E (N.)	anodd ANS (ADJ.)
â AR (PREP.)	agor B (V.)	ansawdd E (N.)
â CYS (CONJ.)	agored ANS (ADJ.)	ar AR (PREP.)
ââ EBYCH (INTERJ.)	agos ANS (ADJ.)	arall ANS (ADJ.)
achos CYS (CONJ.)	agwedd E (N.)	arall RHA (PRON.)
achos E (N.)	angen E (N.)	arbennig ANS (ADJ.)
achub B (V.)	angenhaidiol ANS (ADJ.)	archif E (N.)
adeg E (N.)	anghofio B (V.)	ardal E (N.)
adeilad E (N.)	allan ADF (ADV.)	arfer B (V.)
adeiladu B (V.)	allweddol ANS (ADJ.)	arfer E (N.)
aderyn E (N.)	am AR (PREP.)	arglwydd E (N.)
adnabod B (V.)	amddiffyn B (V.)	arian E (N.)
adnodd E (N.)	aml ANS (ADJ.)	ariannol ANS (ADJ.)
adran E (N.)	amlwg ANS (ADJ.)	ar-lein ANS (ADJ.)
adref ADF (ADV.)	amrywiaeth E (N.)	aros B (V.)
adrodd B (V.)	amser E (N.)	arwain B (V.)
adroddiad E (N.)	aneglur ANS (ADJ.)	arweinydd E (N.)
addas ANS (ADJ.)	anfon B (V.)	arwydd E (N.)
addysg E (N.)	anffodus ANS (ADJ.)	as E (N.)
aelod E (N.)	anifail E (N.)	astudiaeth E (N.)
		astudio B (V.)
		at AR (PREP.)

ateb E (N.)	bod B (V.)	bywyd E (N.)
athro E (N.)	bod E (N.)	cacen E (N.)
awdur E (N.)	bodoli B (V.)	cadair E (N.)
awdurdod E (N.)	bodd E (N.)	cadw B (V.)
awgrymu B (V.)	boi E (N.)	cae E (N.)
awr E (N.)	bore E (N.)	cael B (V.)
awyr E (N.)	braf ANS (ADJ.)	cais E (N.)
babi E (N.)	brawd E (N.)	cal E (N.)
bach ANS (ADJ.)	brenin E (N.)	caled ANS (ADJ.)
bach E (N.)	bro E (N.)	calon E (N.)
bach RHA (PRON.)	bron ADF (ADV.)	cam E (N.)
bachgen E (N.)	brwydr E (N.)	cân E (N.)
bai E (N.)	bryd E (N.)	canfod B (V.)
banc E (N.)	bryn E (N.)	caniatáu B (V.)
band E (N.)	buan ANS (ADJ.)	canlyniad E (N.)
bardd E (N.)	busnes E (N.)	canllaw E (N.)
barn E (N.)	bwrdd E (N.)	canol E (N.)
bath E (N.)	bwriad E (N.)	canolbwyntio B (V.)
bathu B (V.)	bwyd E (N.)	canolfan E (N.)
ben E (N.)	bwyta B (V.)	canrif E (N.)
blaen E (N.)	bychan ANS (ADJ.)	canu B (V.)
blaenorol ANS (ADJ.)	byd E (N.)	canu E (N.)
ble RHA (PRON.)	bynnag ANS (ADJ.)	capel E (N.)
blog E (N.)	byth ADF (ADV.)	car E (N.)
blwyddyn E (N.)	byw ANS (ADJ.)	cariad E (N.)
blynyddol ANS (ADJ.)	byw B (V.)	cario B (V.)
		carreg E (N.)
		cartref E (N.)

caru B (v.)	clwb E (N.)	cwci E (N.)
casgliad E (N.)	clywed B (v.)	cwestiwn E (N.)
casglu B (v.)	coch ANS (ADJ.)	cwm E (N.)
castell E (N.)	codi B (v.)	cwmni E (N.)
categori E (N.)	coed E (N.)	cwmpas E (N.)
cau B (v.)	cof E (N.)	cwpan E (N.)
cefn E (N.)	cofio B (v.)	cwrdd E (N.)
cefnidir E (N.)	cofnod E (N.)	cwrs E (N.)
cefnogaeth E (N.)	cofnodi B (v.)	cychwyn B (v.)
cefnogi B (v.)	cofrestru B (v.)	cychwyn E (N.)
ceffyl E (N.)	coginio B (v.)	cyd ANS (ADJ.)
ceisio B (v.)	coleg E (N.)	cyd E (N.)
cenedl E (N.)	colli B (v.)	cyd- U (UNIQUE)
cenedlaethol ANS (ADJ.)	comisiynydd E (N.)	cydnabod B (v.)
cerdyn E (N.)	copi E (N.)	cyfaill E (N.)
cerdd E (N.)	côr E (N.)	cyfan ANS (ADJ.)
cerdded B (v.)	corff E (N.)	cyfan E (N.)
cerddoriaeth E (N.)	cost E (N.)	cyfarfod E (N.)
ci E (N.)	credu B (v.)	cyfarwydd ANS (ADJ.)
cig E (N.)	crefydd E (N.)	cyfeiriad E (N.)
cilydd E (N.)	creu B (v.)	cyfeirio B (v.)
cinio E (N.)	criw E (N.)	cyfer E (N.)
claf E (N.)	croesawu B (v.)	cyfieithu B (v.)
clicio B (v.)	croeso E (N.)	cyflawni B (v.)
clir ANS (ADJ.)	cryf ANS (ADJ.)	cyfle E (N.)
cloch E (N.)	cwbl E (N.)	cyflwyno B (v.)
		cyflym ANS (ADJ.)
		cyfnod E (N.)

cyfraith E (N.)	cyn AR (PREP.)	cytuno B (V.)
cyfraniad E (N.)	cynhadledd E (N.)	cyw E (N.)
cyfrannu B (V.)	cynhyrchu B (V.)	cywir ANS (ADJ.)
cyfres E (N.)	cynhyrchydd E (N.)	chdi RHA (PRON.)
cyfrifol ANS (ADJ.)	cynllun E (N.)	chi RHA (PRON.)
cyfrifoldeb E (N.)	cynllunio B (V.)	chwaer E (N.)
cyfrol E (N.)	cynnal B (V.)	chwaith ADF (ADV.)
cyfrwng E (N.)	cynnar ANS (ADJ.)	chwarae B (V.)
cyffredin ANS (ADJ.)	cynnig B (V.)	chwarae E (N.)
cyffredinol ANS (ADJ.)	cynnig E (N.)	chwech RHI (NUM.)
cyngor E (N.)	cynnwys B (V.)	chwerthin B (V.)
cyhoeddi B (V.)	cynnwys E (N.)	chwilio B (V.)
cyhoeddus ANS (ADJ.)	cynnyrch E (N.)	da ANS (ADJ.)
cylch E (N.)	cynradd ANS (ADJ.)	dad E (N.)
cylchgrawn E (N.)	cyntaf ANS (ADJ.)	dadl E (N.)
cylchoedd E (N.)	cyntaf RHI (NUM.)	dafad E (N.)
cymaint ANS (ADJ.)	cynulleidfa E (N.)	dangos B (V.)
cymdeithas E (N.)	cynulliad E (N.)	dal B (V.)
cymdeithasol ANS (ADJ.)	cynyddu B (V.)	darganfod B (V.)
cymharu B (V.)	cyrraedd B (V.)	darlledu B (V.)
cymorth E (N.)	cysgu B (V.)	darllen B (V.)
cymraeg E (N.)	cyson ANS (ADJ.)	darn E (N.)
cymreig ANS (ADJ.)	cystadleuaeth E (N.)	darparu B (V.)
cymro E (N.)	cysylltiad E (N.)	data E (N.)
cymryd B (V.)	cysylltu B (V.)	datblygiad E (N.)
cymuned E (N.)	cytundeb E (N.)	datblygu B (V.)
		datganiad E (N.)
		dathlu B (V.)

dau RHI (NUM.)	dim BAN (DET.)	dweud B (V.)
de ANS (ADJ.)	dim E (N.)	dŵr E (N.)
de E (N.)	dinas E (N.)	dwieithog ANS (ADJ.)
deall B (V.)	diogelu B (V.)	dwyrain E (N.)
dechrau B (V.)	diolch E (N.)	dy BAN (DET.)
dechrau E (N.)	disgrifio B (V.)	dy RHA (PRON.)
deddf E (N.)	disgwyl B (V.)	dychwelyd B (V.)
defnydd E (N.)	disgwyl E (N.)	dydd E (N.)
defnyddio B (V.)	disgybl E (N.)	dyddiad E (N.)
defnyddiol ANS (ADJ.)	diwedd E (N.)	dyfodol E (N.)
deg RHI (NUM.)	diweddar ANS (ADJ.)	dylu B (V.)
derbyn B (V.)	diweddaraf ANS (ADJ.)	dyma ADF (ADV.)
deunydd E (N.)	diwethaf ANS (ADJ.)	dyn E (N.)
dewis B (V.)	diwrnod E (N.)	dyna ADF (ADV.)
dewis E (N.)	diwydiant E (N.)	dyna U (UNIQUE)
di- U (UNIQUE)	diwylliant E (N.)	dysgu B (V.)
diddordeb E (N.)	dod B (V.)	dysgwyr E (N.)
diddorol ANS (ADJ.)	dogfen E (N.)	eang ANS (ADJ.)
diffyg E (N.)	dosbarth E (N.)	e-bost E (N.)
digon E (N.)	drama E (N.)	economaidd ANS (ADJ.)
digon RHA (PRON.)	draw ADF (ADV.)	economi E (N.)
digwydd B (V.)	drwg ANS (ADJ.)	edrych B (V.)
digwyddiad E (N.)	drws E (N.)	efallai ADF (ADV.)
dilyn B (V.)	du ANS (ADJ.)	efo AR (PREP.)
dilledyn E (N.)	dull E (N.)	effaith E (N.)
dim ADF (ADV.)	duw E (N.)	eglwys E (N.)
		eng ANS (ADJ.)
		enghraifft E (N.)

ei BAN (DET.)	ers AR (PREP.)	ffonio B (v.)
ei RHA (PRON.)	erthygl E (N.)	ffordd E (N.)
eich BAN (DET.)	esbonio B (v.)	ffrind E (N.)
eich RHA (PRON.)	eto ADF (ADV.)	ffurf E (N.)
eidion E (N.)	etholiad E (N.)	ffurflen E (N.)
eiliad E (N.)	eu BAN (DET.)	ffwrdd E (N.)
ein BAN (DET.)	eu RHA (PRON.)	gadael B (v.)
ein RHA (PRON.)	ewropeaidd ANS (ADJ.)	gafael E (N.)
eisiau B (v.)	faint RHA (PRON.)	gair E (N.)
eisiau E (N.)	fan YNA ADF (ADV.)	galw B (v.)
eisoës ADF (ADV.)	fath BAN (DET.)	galw E (N.)
eistedd B (v.)	fe U (UNIQUE)	gallu B (v.)
eisteddfod E (N.)	fel CYS (CONJ.)	gallu E (N.)
eitem E (N.)	felly ADF (ADV.)	gan AR (PREP.)
eithaf ADF (ADV.)	fersiwn E (N.)	gardd E (N.)
eleni ADF (ADV.)	fideo E (N.)	gêm E (N.)
elfen E (N.)	fo AR (PREP.)	geni B (v.)
ennill B (v.)	fo RHA (PRON.)	ger AR (PREP.)
ennill E (N.)	fy BAN (DET.)	glas ANS (ADJ.)
enw E (N.)	fy RHA (PRON.)	go ADF (ADV.)
enwedig ANS (ADJ.)	fy YFB (DEF. ART.)	gobaith E (N.)
enwog ANS (ADJ.)	fyny ADF (ADV.)	gobeithio B (v.)
er AR (PREP.)	ffaith E (N.)	gofal E (N.)
erbyn AR (PREP.)	ffarmwr E (N.)	gofyn B (v.)
erbyn B (v.)	fferm E (N.)	gogledd E (N.)
erioed ADF (ADV.)	ffilm E (N.)	golwg E (N.)
		golygu B (v.)

gorfod B (v.)	gwener E (N.)	gyrru B (v.)
gorfod E (N.)	gwera E (N.)	haf E (N.)
gorffen B (v.)	gwerth E (N.)	hanes E (N.)
gorllewin E (N.)	gwerthu B (v.)	hanesyddol ANS (ADJ.)
gosod B (v.)	gwir E (N.)	hanner E (N.)
gradd E (N.)	gwir EBYCH (INTERJ.)	hapus ANS (ADJ.)
grŵp E (N.)	gwirionedd E (N.)	haul E (N.)
gwae E (N.)	gwisgo B (v.)	hawdd ANS (ADJ.)
gwahaniaeth E (N.)	gwlad E (N.)	hawl E (N.)
gwahanol ANS (ADJ.)	gwleidyddol ANS (ADJ.)	heb AR (PREP.)
gwaith E (N.)	gwneud B (v.)	heddiw ADF (ADV.)
gwasanaeth E (N.)	gwobr E (N.)	heddlu E (N.)
gwasanaethu B (v.)	gŵr E (N.)	hefyd ADF (ADV.)
gwasg E (N.)	gwraig E (N.)	heibio AR (PREP.)
gwastad ADF (ADV.)	gwrando B (v.)	helo EBYCH (INTERJ.)
gwau E (N.)	gwreiddiol ANS (ADJ.)	help E (N.)
gweddill E (N.)	gwybod B (v.)	helpu B (v.)
gwefan E (N.)	gwybod E (N.)	hen ANS (ADJ.)
gweinidog E (N.)	gwybod EBYCH (INTERJ.)	heno ADF (ADV.)
gweithgaredd E (N.)	gwybodaeth E (N.)	her E (N.)
gweithio B (v.)	gwybodaeth E (N.)	hi RHA (PRON.)
gweithiwr E (N.)	gwybodaeth E (N.)	hir ANS (ADJ.)
gweithredu B (v.)	gwybodaeth E (N.)	hmm EBYCH (INTERJ.)
gweld B (v.)	gwybodaeth E (N.)	hoff ANS (ADJ.)
gwely E (N.)	gwybodaeth E (N.)	hoffi B (v.)
gwella B (v.)	gwybodaeth E (N.)	holi B (v.)
	gwybodaeth E (N.)	holl ANS (ADJ.)
	gwybodaeth E (N.)	holl BAN (DET.)

hollol ANS (ADJ.)	ie ADF (ADV.)	llenwi B (V.)
hun RHA (PRON.)	iechyd E (N.)	lleol ANS (ADJ.)
hunan RHA (PRON.)	ieithyddol ANS (ADJ.)	lleoliad E (N.)
hwn RHA (PRON.)	ifanc ANS (ADJ.)	llinell E (N.)
hwinnw ANS (ADJ.)	is ANS (ADJ.)	lliw E (N.)
hwinnw BAN (DET.)	is-deitl E (N.)	llu E (N.)
hwinnw RHA (PRON.)	isod ADF (ADV.)	llun E (N.)
hwnt ANS (ADJ.)	jest ADF (ADV.)	llunio B (V.)
hwyl E (N.)	jyst ADF (ADV.)	llwybr E (N.)
hybu B (V.)	lan ADF (ADV.)	llwyddiannus ANS (ADJ.)
hyd AR (PREP.)	lefel E (N.)	llwyddiant E (N.)
hyd E (N.)	licio B (V.)	llwyddo B (V.)
hyfryd ANS (ADJ.)	lot E (N.)	llwyfan E (N.)
hyfforddiant E (N.)	lot RHA (PRON.)	llwyr ANS (ADJ.)
hŷn ANS (ADJ.)	lladd B (V.)	llyfr E (N.)
hynod ANS (ADJ.)	llafur E (N.)	llyfrgell E (N.)
hyrwyddo B (V.)	llais E (N.)	llygad E (N.)
hytrach ADF (ADV.)	llall RHA (PRON.)	llys E (N.)
i AR (PREP.)	llaw E (N.)	llythyr E (N.)
iâ E (N.)	llawer RHA (PRON.)	llywodraeth E (N.)
iaith E (N.)	llawn ANS (ADJ.)	mab E (N.)
iau E (N.)	llawr ADF (ADV.)	mae B (V.)
iawn ADF (ADV.)	llawr E (N.)	maes E (N.)
iawn ANS (ADJ.)	lle ANS (ADJ.)	magu B (V.)
iawn E (N.)	lle E (N.)	mai CYS (CONJ.)
iawn EBYCH (INTERJ.)	lle RHA (PRON.)	mai E (N.)
		mam E (N.)
		man E (N.)

mân ANS (ADJ.)	milltir E (N.)	neis ANS (ADJ.)
manylyn E (N.)	mis E (N.)	neu CYS (CONJ.)
map E (N.)	mmm EBYCH (INTERJ.)	newid B (V.)
marchnad E (N.)	mor ADF (ADV.)	newydd ANS (ADJ.)
marw ANS (ADJ.)	môr E (N.)	newyddion E (N.)
marwolaeth E (N.)	moyn E (N.)	nhw RHA (PRON.)
màs E (N.)	mudiad E (N.)	ni RHA (PRON.)
masnach E (N.)	munud E (N.)	ni U (UNIQUE)
mater E (N.)	mwy RHA (PRON.)	nid U (UNIQUE)
math E (N.)	mwyn E (N.)	nifer E (N.)
mawr ANS (ADJ.)	mwynhau B (V.)	nod E (N.)
mawrth E (N.)	myfyriwr E (N.)	nodi B (V.)
medru B (V.)	mynd B (V.)	nofel E (N.)
medd B (V.)	mynediad E (N.)	nôl B (V.)
meddwl B (V.)	mynnu B (V.)	nos E (N.)
meddwl E (N.)	mynydd E (N.)	noson E (N.)
megis CYS (CONJ.)	na CYS (CONJ.)	o AR (PREP.)
meic E (N.)	na RHA (PRON.)	oce EBYCH (INTERJ.)
meithrin B (V.)	na U (UNIQUE)	ochr E (N.)
menyw E (N.)	nadolig E (N.)	oddi AR (PREP.)
merch E (N.)	natur E (N.)	oed E (N.)
methu B (V.)	naturiol ANS (ADJ.)	oedfa E (N.)
mewn AR (PREP.)	naw RHI (NUM.)	oedolyn E (N.)
mhm EBYCH (INTERJ.)	nawr ADF (ADV.)	oen E (N.)
mi RHA (PRON.)	neb RHA (PRON.)	oes E (N.)
mi U (UNIQUE)	neges E (N.)	ofn E (N.)
		ogystal ANS (ADJ.)
		oh EBYCH (INTERJ.)

oherwydd CYS (CONJ.)	penodol ANS (ADJ.)	presennol ANS (ADJ.)
ôl ANS (ADJ.)	pentref E (N.)	prif ANS (ADJ.)
ôl E (N.)	perffaith ANS (ADJ.)	prifysgol E (N.)
olaf ANS (ADJ.)	perfformiad E (N.)	prin ANS (ADJ.)
ond CYS (CONJ.)	perfformio B (V.)	priodas E (N.)
OS CYS (CONJ.)	person E (N.)	pris E (N.)
pa ADF (ADV.)	personol ANS (ADJ.)	problem E (N.)
pa BAN (DET.)	perthnasol ANS (ADJ.)	profi B (V.)
pam ADF (ADV.)	perthynas E (N.)	profiad E (N.)
pan CYS (CONJ.)	petai CYS (CONJ.)	proses E (N.)
papur E (N.)	peth E (N.)	prosiect E (N.)
paratoi B (V.)	peth RHA (PRON.)	pryd E (N.)
parod ANS (ADJ.)	plaid E (N.)	pryd RHA (PRON.)
parti E (N.)	plentyn E (N.)	prynhawn E (N.)
parhau B (V.)	pob ANS (ADJ.)	prynu B (V.)
pawb RHA (PRON.)	pob BAN (DET.)	prysur ANS (ADJ.)
pe CYS (CONJ.)	pobl E (N.)	pump RHI (NUM.)
pecyn E (N.)	poblogaidd ANS (ADJ.)	punt E (N.)
pedwar RHI (NUM.)	pobman E (N.)	pwll E (N.)
peidio B (V.)	pobol E (N.)	pwnc E (N.)
pêl E (N.)	poeni B (V.)	pwyl E (N.)
pell ADF (ADV.)	polisi E (N.)	pwyl RHA (PRON.)
pell ANS (ADJ.)	popeth E (N.)	pwylgor E (N.)
pen E (N.)	posib ANS (ADJ.)	pwylt E (N.)
penderfyniad E (N.)	posibl ANS (ADJ.)	pwys E (N.)
penderfynu B (V.)	post E (N.)	pwysig ANS (ADJ.)
		radio E (N.)
		reit EBYCH (INTERJ.)

rôl E (N.)	rhyfedd ANS (ADJ.)	seren E (N.)
rownd E (N.)	rhyfel E (N.)	sesiwn E (N.)
rŵan ADF (ADV.)	rhyngwladol ANS (ADJ.)	sgîl E (N.)
rygbi E (N.)	rhyw BAN (DET.)	sgwennu B (V.)
rhag AR (PREP.)	rhyw E (N.)	sgwrs E (N.)
rhaglen E (N.)	rhywbeth E (N.)	siarad B (V.)
rhagor RHA (PRON.)	rhywbeth RHA (PRON.)	siaradwr E (N.)
rhai BAN (DET.)	rhywun E (N.)	sicr ANS (ADJ.)
rhai RHA (PRON.)	rhywun RHA (PRON.)	sicrhau B (V.)
rhaid E (N.)	sadwrn E (N.)	sioe E (N.)
rhain RHA (PRON.)	saesneg E (N.)	siop E (N.)
rhan E (N.)	safbwynt E (N.)	sir E (N.)
rhannu B (V.)	safle E (N.)	siŵr ANS (ADJ.)
rhedeg B (V.)	safon E (N.)	so ADF (ADV.)
rheini RHA (PRON.)	sail E (N.)	sôn B (V.)
rheol E (N.)	sain E (N.)	sôn E (N.)
rheoli B (V.)	saith RHI (NUM.)	sofi EBYCH (INTERJ.)
rhestr E (N.)	sant E (N.)	staff E (N.)
rheswm E (N.)	sawl BAN (DET.)	stori E (N.)
rhiant E (N.)	sector E (N.)	stryd E (N.)
rhifyn E (N.)	sef CYS (CONJ.)	sul E (N.)
rhoddi B (V.)	sefydliad E (N.)	sut RHA (PRON.)
rhoi B (V.)	sefydlu B (V.)	sŵn E (N.)
rhwng AR (PREP.)	sefyll B (V.)	swydd E (N.)
rhy ADF (ADV.)	sefyllfa E (N.)	swyddfa E (N.)
rhyddhau B (V.)	senedd E (N.)	swyddog E (N.)
		swyddogol ANS (ADJ.)
		sydyn ANS (ADJ.)

sylw E (N.)	term E (N.)	troi B (v.)
sylwad E (N.)	testun E (N.)	tros AR (PREP.)
sylweddol ANS (ADJ.)	teulu E (N.)	trwy AR (PREP.)
sylweddoli B (v.)	tewi B (v.)	tu E (N.)
sylwi B (v.)	teyrnas E (N.)	tua AR (PREP.)
syml ANS (ADJ.)	ti RHA (PRON.)	tudalen E (N.)
symud B (v.)	tîm E (N.)	tŷ E (N.)
syniad E (N.)	tipyn E (N.)	tybed ADF (ADV.)
system E (N.)	tir E (N.)	tyfu B (v.)
syth ANS (ADJ.)	tiwtor E (N.)	tymor E (N.)
tad E (N.)	to E (N.)	tynnu B (v.)
tag E (N.)	torri B (v.)	tystiolaeth E (N.)
taith E (N.)	tra CYS (CONJ.)	tywydd E (N.)
taliad E (N.)	traed E (N.)	theatr E (N.)
talw B (v.)	trafod B (v.)	thema E (N.)
tan AR (PREP.)	trafod E (N.)	uchel ANS (ADJ.)
tân E (N.)	trafodaeth E (N.)	uchod ADF (ADV.)
tasg E (N.)	traaws AR (PREP.)	ugain RHI (NUM.)
tawel ANS (ADJ.)	tref E (N.)	un BAN (DET.)
te E (N.)	trefn E (N.)	un RHA (PRON.)
tebyg ANS (ADJ.)	trefnu B (v.)	un RHI (NUM.)
technoleg E (N.)	treulio B (v.)	undeb E (N.)
teg ANS (ADJ.)	tri RHI (NUM.)	unig ANS (ADJ.)
teimlo B (v.)	trin B (v.)	unig BAN (DET.)
teithio B (v.)	trio B (v.)	unigolyn E (N.)
teledu E (N.)	tro E (N.)	unigryw ANS (ADJ.)
		union ANS (ADJ.)
		unrhyw ANS (ADJ.)

unrhyw BAN (DET.)	ychwanegu B (v.)	ymweliad E (N.)
unwaith ADF (ADV.)	ychedig BAN (DET.)	ymwneud B (v.)
uwch ANS (ADJ.)	ychedig RHA (PRON.)	yn ADF (ADV.)
wal E (N.)	ynghyd ADF (ADV.)	yn AR (PREP.)
wedi ADF (ADV.)	ynghylch AR (PREP.)	yn U (UNIQUE)
wedi AR (PREP.)	ynglŷn AR (PREP.)	yna ADF (ADV.)
wedi U (UNIQUE)	yma ADF (ADV.)	yna BAN (DET.)
wedyn ADF (ADV.)	yma BAN (DET.)	ynys E (N.)
weithiau ADF (ADV.)	ymarfer E (N.)	ysbyty E (N.)
wel EBYCH (INTERJ.)	ymateb E (N.)	ysgol E (N.)
wrth AR (PREP.)	ymchwil E (N.)	ysgrifennu B (v.)
wyneb E (N.)	ymdrech E (N.)	ysgrifennydd E (N.)
wynebu B (v.)	ymddangos B (v.)	ystafell E (N.)
ŵyr E (N.)	ymgyrch E (N.)	ystod E (N.)
wyth RHI (NUM.)	ymhlith AR (PREP.)	ystyr E (N.)
wythnos E (N.)	ymlaen ADF (ADV.)	ystyried B (v.)
y RHA (PRON.)	ymm EBYCH (INTERJ.)	yyy EBYCH (INTERJ.)
y YFB (DEF. ART.)	ymuno B (v.)	
ychwanegol ANS (ADJ.)	ymweld B (v.)	

C.5 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig)

Top 100 lemmas in CorCenCC (open-class words only)

Safle (rank)	Lemma	Rhan Ymadrodd (part of speech)	Cyfanswm Crai (raw total)	Cyfanswm y fliwn (per-million total)
1	bod	B (v.)	649744	122343.49
2	cael	B (v.)	64101	12069.89
3	gwneud	B (v.)	50508	9510.40
4	mynd	B (v.)	39718	7478.70
5	da	Ans (adj.)	34127	6425.94
6	ie	Adf (adv.)	32876	6190.38
7	dod	B (v.)	32755	6167.60
8	cymraeg	E (n.)	31784	5984.77
9	mae	B (v.)	27681	5212.19
10	mawr	Ans (adj.)	27552	5187.90
11	gweld	B (v.)	26230	4938.98
12	dim	E (n.)	24999	4707.18
13	dweud	B (v.)	24611	4634.13
14	gallu	B (v.)	24167	4550.52
15	dim	Adf (adv.)	23016	4333.80
16	yma	Adf (adv.)	22188	4177.89
17	rhoi	B (v.)	20705	3898.65

C.5 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig)
 Top 100 lemmas in CorCenCC (open-class words only)

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/swm</i>	<i>c/swm y.f.</i>
18	hefyd	Adf (adv.)	20653	3888.85
19	arall	Ans (adj.)	20629	3884.34
20	yna	Adf (adv.)	20513	3862.49
21	peth	E (n.)	18554	3493.62
22	newydd	Ans (adj.)	18343	3453.89
23	pobl	E (n.)	17378	3272.19
24	felly	Adf (adv.)	17021	3204.97
25	rhan	E (n.)	16260	3061.68
26	gwaith	E (n.)	16187	3047.93
27	blwyddyn	E (n.)	16058	3023.64
28	iaith	E (n.)	15946	3002.55
29	iawn	Adf (adv.)	15939	3001.23
30	bach	Ans (adj.)	15342	2888.82
31	cyfer	E (n.)	14876	2801.08
32	plentyn	E (n.)	14485	2727.45
33	ysgol	E (n.)	13474	2537.09
34	byd	E (n.)	13417	2526.35
35	ôl	Ans (adj.)	13146	2475.32
36	dydd	E (n.)	12934	2435.41
37	defnyddio	B (v.)	12890	2427.12
38	rhaid	E (n.)	12849	2419.40
39	meddwl	B (v.)	12196	2296.44
40	so	Adf (adv.)	12103	2278.93
41	wedyn	Adf (adv.)	11890	2238.83

C.5 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig)
 Top 100 lemmas in CorCenCC (open-class words only)

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/swm</i>	<i>c/swm y.f.</i>
42	ôl	E (n.)	11595	2183.28
43	yn	Adf (adv.)	11450	2155.98
44	tŷ	E (n.)	10834	2039.99
45	dyna	Adf (adv.)	10724	2019.27
46	mor	Adf (adv.)	10578	1991.78
47	angen	E (n.)	10425	1962.97
48	gwybod	B (v.)	10249	1929.83
49	amser	E (n.)	10109	1903.47
50	ffordd	E (n.)	9992	1881.44
51	enw	E (n.)	9855	1855.65
52	dyma	Adf (adv.)	9640	1815.16
53	siarad	B (v.)	9577	1803.30
54	pen	E (n.)	9501	1788.99
55	edrych	B (v.)	9358	1762.06
56	lle	E (n.)	9240	1739.84
57	agos	Ans (adj.)	9236	1739.09
58	plaid	E (n.)	8946	1684.49
59	mis	E (n.)	8740	1645.70
60	mwyn	E (n.)	8561	1611.99
61	gweithio	B (v.)	8545	1608.98
62	tro	E (n.)	8499	1600.32
63	llyfr	E (n.)	8400	1581.68
64	cynnwys	E (n.)	8318	1566.24
65	codi	B (v.)	8286	1560.21

C.5 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig)
 Top 100 lemmas in CorCenCC (open-class words only)

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/swm</i>	<i>c/swm y.f.</i>
66	cwrs	E (n.)	8091	1523.49
67	rhaglen	E (n.)	8086	1522.55
68	dysgu	B (v.)	8022	1510.50
69	diolch	E (n.)	8015	1509.18
70	aelod	E (n.)	7889	1485.46
71	nawr	Adf (adv.)	7722	1454.01
72	eto	Adf (adv.)	7721	1453.83
73	gair	E (n.)	7677	1445.54
74	de	E (n.)	7640	1438.57
75	hen	Ans (adj.)	7607	1432.36
76	cyntaf	Ans (adj.)	7512	1414.47
77	oed	E (n.)	7435	1399.97
78	llywodraeth	E (n.)	7430	1399.03
79	gofyn	B (v.)	7429	1398.84
80	gwybodaeth	E (n.)	7420	1397.15
81	cyfle	E (n.)	7379	1389.43
82	byw	B (v.)	7371	1387.92
83	darllen	B (v.)	7344	1382.84
84	hanes	E (n.)	7333	1380.77
85	gwahanol	Ans (adj.)	7303	1375.12
86	bodd	E (n.)	7288	1372.29
87	cofio	B (v.)	7278	1370.41
88	heddiw	Adf (adv.)	7277	1370.22
89	hyd	E (n.)	7249	1364.95

C.5 100 lema uchaf yn CorCenCC (geiriau dosbarth agored yn unig)
 Top 100 lemmas in CorCenCC (open-class words only)

<i>safle</i>	<i>lemma</i>	<i>rh. ym.</i>	<i>c/swm</i>	<i>c/swm y.f.</i>
90	allan	Adf (adv.)	7246	1364.38
91	dal	B (v.)	7215	1358.55
92	llun	E (n.)	7200	1355.72
93	man	E (n.)	7174	1350.83
94	diwrnod	E (n.)	7047	1326.91
95	rhyw	E (n.)	6991	1316.37
96	gwasanaeth	E (n.)	6961	1310.72
97	gwlad	E (n.)	6935	1305.83
98	clywed	B (v.)	6908	1300.74
99	saesneg	E (n.)	6890	1297.35
100	cyfnod	E (n.)	6882	1295.85

C.6 1000 gair uchaf yn CorCenCC (geiriau dosbarth agored yn unig; rhestr yn ôl yr wyddor) Top 1000 lemmas in CorCenCC (open-class words only; alphabetic list)

achos E (N.)	agored ANS (ADJ.)	anifail E (N.)
achub B (V.)	agos ANS (ADJ.)	annibynnol ANS (ADJ.)
adeg E (N.)	agwedd E (N.)	anodd ANS (ADJ.)
adeilad E (N.)	angen E (N.)	ansawdd E (N.)
adeiladu B (V.)	angenrheidiol ANS (ADJ.)	arall ANS (ADJ.)
aderyn E (N.)	anghofio B (V.)	arbennig ANS (ADJ.)
adnabod B (V.)	allan ADF (ADV.)	archebu B (V.)
adnodd E (N.)	allweddol ANS (ADJ.)	archif E (N.)
adolygiad E (N.)	amddiffyn B (V.)	archwilio B (V.)
adran E (N.)	amgueddfa E (N.)	ardal E (N.)
adref ADF (ADV.)	amgylch E (N.)	arfer B (V.)
adrodd B (V.)	aml ANS (ADJ.)	arfer E (N.)
adroddiad E (N.)	amlwg ANS (ADJ.)	arglwydd E (N.)
addas ANS (ADJ.)	amod E (N.)	arian E (N.)
addysg E (N.)	amrywiaeth E (N.)	ariannol ANS (ADJ.)
addysgu B (V.)	amser E (N.)	ar-lein ANS (ADJ.)
aelod E (N.)	an E (N.)	aros B (V.)
aeth B (V.)	aneglur ANS (ADJ.)	arwain B (V.)
afon E (N.)	anfon B (V.)	arweiniad E (N.)
agor B (V.)	anffodus ANS (ADJ.)	arweinydd E (N.)
		arwydd E (N.)

as E (N.)	blas E (N.)	bwrdd E (N.)
astudiaeth E (N.)	blodyn E (N.)	bwriad E (N.)
astudio B (v.)	blog E (N.)	bwrw B (v.)
atal B (v.)	blwydd E (N.)	bws E (N.)
ateb B (v.)	blwyddyn E (N.)	bwyd E (N.)
ateb E (N.)	blynyddol ANS (ADJ.)	bwyta B (v.)
athro E (N.)	bocs E (N.)	bychan ANS (ADJ.)
awdur E (N.)	bod B (v.)	byd E (N.)
awdurdod E (N.)	bod E (N.)	bynnag ANS (ADJ.)
awgrymu B (v.)	bodlon ANS (ADJ.)	byr ANS (ADJ.)
awr E (N.)	bodoli B (v.)	byth ADF (ADV.)
awyr E (N.)	bodd E (N.)	byw ANS (ADJ.)
babi E (N.)	boi E (N.)	byw B (v.)
bach ANS (ADJ.)	bore E (N.)	bywyd E (N.)
bach E (N.)	braf ANS (ADJ.)	cacen E (N.)
bachgen E (N.)	brawd E (N.)	cadair E (N.)
bai E (N.)	brawddeg E (N.)	cadeirydd E (N.)
banc E (N.)	brenin E (N.)	cadw B (v.)
band E (N.)	bro E (N.)	cae E (N.)
bardd E (N.)	bron ADF (ADV.)	cael B (v.)
barn E (N.)	brwydr E (N.)	cael E (N.)
bath E (N.)	bryd E (N.)	cais E (N.)
bathu B (v.)	bryn E (N.)	cal E (N.)
ben E (N.)	buan ANS (ADJ.)	caled ANS (ADJ.)
blaen E (N.)	budd E (N.)	calon E (N.)
blaenorol ANS (ADJ.)	busnes E (N.)	cam E (N.)
		cân E (N.)
		canfod B (v.)

caniatáu B (v.)	cefnogwr E (N.)	cof E (N.)
canlyniad E (N.)	ceffyl E (N.)	cofio B (v.)
canlynol ANS (ADJ.)	cegin E (N.)	cofnod E (N.)
canllaw E (N.)	ceisio B (v.)	cofnodi B (v.)
canol E (N.)	celf E (N.)	cofrestru B (v.)
canolbwyntio B (v.)	cenedl E (N.)	coffi E (N.)
canolfan E (N.)	cenedlaethol ANS (ADJ.)	coginio B (v.)
canrif E (N.)	cerdyn E (N.)	coleg E (N.)
canu B (v.)	cerdd E (N.)	colli B (v.)
canu E (N.)	cerdded B (v.)	comisiwn E (N.)
capel E (N.)	cerddoriaeth E (N.)	comisiynydd E (N.)
car E (N.)	ci E (N.)	copi E (N.)
cariad E (N.)	cig E (N.)	côr E (N.)
cario B (v.)	cilydd E (N.)	corff E (N.)
carreg E (N.)	cinio E (N.)	cost E (N.)
cartref E (N.)	claf E (N.)	credu B (v.)
caru B (v.)	clicio B (v.)	crefydd E (N.)
casgliad E (N.)	clip E (N.)	creu B (v.)
casglu B (v.)	clir ANS (ADJ.)	criw E (N.)
castell E (N.)	cloch E (N.)	croes E (N.)
categori E (N.)	clwb E (N.)	croesawu B (v.)
cau B (v.)	clywed B (v.)	croeso E (N.)
cefn E (N.)	coch ANS (ADJ.)	cronfa E (N.)
cefnidir E (N.)	codi B (v.)	cryf ANS (ADJ.)
cefnogaeth E (N.)	coed E (N.)	cwbl E (N.)
cefnogi B (v.)	coes E (N.)	cwci E (N.)
		cwestiwn E (N.)
		cwm E (N.)

cwmni E (N.)	cyflwr E (N.)	cymdeithasol ANS (ADJ.)
cwmpas E (N.)	cyflwyno B (V.)	cymeriad E (N.)
cwpan E (N.)	cyflym ANS (ADJ.)	cymharu B (V.)
cwr E (N.)	cyfnod E (N.)	cymorth E (N.)
cwrdd E (N.)	cyfraith E (N.)	cymraeg E (N.)
cwrs E (N.)	cyfraniad E (N.)	cymreig ANS (ADJ.)
cwsmer E (N.)	cyfrannu B (V.)	cymro E (N.)
cychwyn B (V.)	cyfres E (N.)	cymryd B (V.)
cychwyn E (N.)	cyfrif B (V.)	cymuned E (N.)
cyd ANS (ADJ.)	cyfrif E (N.)	cynhadledd E (N.)
cyd E (N.)	cyfrifiad E (N.)	cynhyrchu B (V.)
cydnabod B (V.)	cyfrifol ANS (ADJ.)	cynhyrchydd E (N.)
cydymffurfio B (V.)	cyfrifoldeb E (N.)	cynllun E (N.)
cyfaill E (N.)	cyfrol E (N.)	cynllunio B (V.)
cyfan ANS (ADJ.)	cyfrwng E (N.)	cynnal B (V.)
cyfan E (N.)	cyffredin ANS (ADJ.)	cynnar ANS (ADJ.)
cyfarfod E (N.)	cyffredinol ANS (ADJ.)	cynnig B (V.)
cyfarwydd ANS (ADJ.)	cyffrous ANS (ADJ.)	cynnig E (N.)
cyfeiriad E (N.)	cyngor E (N.)	cynnwys B (V.)
cyfeirio B (V.)	cyhoeddi B (V.)	cynnwys E (N.)
cyfer E (N.)	cyhoeddus ANS (ADJ.)	cynnydd E (N.)
cyfeithiad E (N.)	cylch E (N.)	cynnyrch E (N.)
cyfeithu B (V.)	cylchgrawn E (N.)	cynradd ANS (ADJ.)
cyflawni B (V.)	cylchoedd E (N.)	cynrychioli B (V.)
cyfle E (N.)	cymaint ANS (ADJ.)	cyntaf ANS (ADJ.)
cyflog E (N.)	cymdeithas E (N.)	cynulleidfa E (N.)
		cynulliad E (N.)
		cynyddu B (V.)

cyrraedd B (v.)	darganfod B (v.)	dewis E (N.)
cysgu B (v.)	darlledu B (v.)	dibynnu B (v.)
cyson ANS (ADJ.)	darllen B (v.)	diddordeb E (N.)
cystadleuaeth E (N.)	darn E (N.)	diddorol ANS (ADJ.)
cystadlu B (v.)	darpariaeth E (N.)	diddymu B (v.)
cyswllt E (N.)	darparu B (v.)	diflannu B (v.)
cysylltiad E (N.)	data E (N.)	difyr ANS (ADJ.)
cysylltu B (v.)	datblygiad E (N.)	diffyg E (N.)
cytundeb E (N.)	datblygu B (v.)	digon E (N.)
cytuno B (v.)	datganiad E (N.)	digwydd B (v.)
cyw E (N.)	dathlu B (v.)	digwyddiad E (N.)
cywir ANS (ADJ.)	de ANS (ADJ.)	dilyn B (v.)
chwaer E (N.)	de E (N.)	dilledyn E (N.)
chwaith ADF (ADV.)	deall B (v.)	dim ADF (ADV.)
chwarae B (v.)	dechrau B (v.)	dim E (N.)
chwarae E (N.)	dechrau E (N.)	dinas E (N.)
chwerthin B (v.)	deddf E (N.)	diogelu B (v.)
chwilio B (v.)	defnydd E (N.)	diolch E (N.)
chwith ANS (ADJ.)	defnyddio B (v.)	disgrifio B (v.)
da ANS (ADJ.)	defnyddiol ANS (ADJ.)	disgwyl B (v.)
dad E (N.)	defnyddiwr E (N.)	disgwyl E (N.)
dadl E (N.)	denu B (v.)	disgybl E (N.)
daear E (N.)	derbyn B (v.)	diwedd E (N.)
dafad E (N.)	deunydd E (N.)	diweddar ANS (ADJ.)
dangos B (v.)	dewin E (N.)	diweddaraf ANS (ADJ.)
dal B (v.)	dewis B (v.)	diwethaf ANS (ADJ.)
		diwrnod E (N.)
		diwydiant E (N.)

diwylliant E (N.)	dymuno B (V.)	eitem E (N.)
dod B (V.)	dyn E (N.)	eithaf ADF (ADV.)
dogfen E (N.)	dyna ADF (ADV.)	eleni ADF (ADV.)
dolen E (N.)	dysgu B (V.)	elfen E (N.)
dosbarth E (N.)	dysgwr E (N.)	elusen E (N.)
drama E (N.)	ddoe ADF (ADV.)	enfawr ANS (ADJ.)
draw ADF (ADV.)	eang ANS (ADJ.)	ennill B (V.)
drwg ANS (ADJ.)	e-bost E (N.)	ennill E (N.)
drws E (N.)	economaidd ANS (ADJ.)	enw E (N.)
du ANS (ADJ.)	economi E (N.)	enwedig ANS (ADJ.)
dull E (N.)	edrych B (V.)	enwog ANS (ADJ.)
duw E (N.)	efallai ADF (ADV.)	erbyn B (V.)
dweud B (V.)	effaith E (N.)	erioed ADF (ADV.)
dŵr E (N.)	effeithio B (V.)	erthygl E (N.)
dwyieithog ANS (ADJ.)	effeithiol ANS (ADJ.)	esbonio B (V.)
dwyn B (V.)	eglwys E (N.)	eto ADF (ADV.)
dwyrain E (N.)	egwyddor E (N.)	etholiad E (N.)
dychwelyd B (V.)	eng ANS (ADJ.)	ewropeaidd ANS (ADJ.)
dydd E (N.)	enghraifft E (N.)	fan YNA ADF (ADV.)
dyddiad E (N.)	eidion E (N.)	felly ADF (ADV.)
dyfodol E (N.)	eiliad E (N.)	fersiwn E (N.)
dyffryn E (N.)	eisiau B (V.)	fideo E (N.)
dylanwad E (N.)	eisiau E (N.)	fory ADF (ADV.)
dyletswydd E (N.)	eisoes ADF (ADV.)	fyny ADF (ADV.)
dylu B (V.)	eistedd B (V.)	ffaith E (N.)
dyma ADF (ADV.)	eisteddfod E (N.)	ffarmwr E (N.)
		ffeindio B (V.)
		fferm E (N.)

ffi E (N.)	go ADF (ADV.)	gwasanaeth E (N.)
ffilm E (N.)	gobaith E (N.)	gwasanaethu B (V.)
ffin E (N.)	gobeithio B (V.)	gwasg E (N.)
ffôn E (N.)	gofal E (N.)	gwastad ADF (ADV.)
ffonio B (V.)	gofalus ANS (ADJ.)	gwau E (N.)
ffordd E (N.)	gofyn B (V.)	gwe E (N.)
ffrind E (N.)	gogledd E (N.)	gweddill E (N.)
ffurf E (N.)	golwg E (N.)	gwefan E (N.)
ffurfio B (V.)	golygfa E (N.)	gweinidog E (N.)
ffurflen E (N.)	golygu B (V.)	gweithgaredd E (N.)
ffwrdd E (N.)	gorfod B (V.)	gweithio B (V.)
ffydd E (N.)	gorfod E (N.)	gweithiwr E (N.)
gadael B (V.)	gorfodi B (V.)	gweithredu B (V.)
gaeaf E (N.)	gorffen B (V.)	gweld B (V.)
gafael E (N.)	gorllewin E (N.)	gweledigaeth E (N.)
gair E (N.)	gormod E (N.)	gwely E (N.)
galw B (V.)	gosod B (V.)	gwella B (V.)
galw E (N.)	gradd E (N.)	gwener E (N.)
gall E (N.)	grŵp E (N.)	gwenu B (V.)
gallu B (V.)	gwae E (N.)	gwerin E (N.)
gallu E (N.)	gwaelod E (N.)	gwerau E (N.)
gardd E (N.)	gwahaniaeth E (N.)	gwerth E (N.)
gêm E (N.)	gwahanol ANS (ADJ.)	gwerthu B (V.)
geni B (V.)	gwaith E (N.)	gwir E (N.)
glas ANS (ADJ.)	gwallt E (N.)	gwirionedd E (N.)
glo E (N.)	gwas E (N.)	gwisgo B (V.)
		gwlad E (N.)
		gwladwriaeth E (N.)

gwleidyddol ANS (ADJ.)	hawl E (N.)	hyrwyddo B (V.)
gwneud B (V.)	heddiw ADF (ADV.)	hytrach ADF (ADV.)
gwobr E (N.)	heddlu E (N.)	iâ E (N.)
gŵr E (N.)	hefyd ADF (ADV.)	iaith E (N.)
gwraig E (N.)	help E (N.)	iau E (N.)
gwrando B (V.)	helpu B (V.)	iawn ADF (ADV.)
gwreiddiol ANS (ADJ.)	hen ANS (ADJ.)	iawn ANS (ADJ.)
gwybod B (V.)	heno ADF (ADV.)	iawn E (N.)
gwybod E (N.)	heol E (N.)	ie ADF (ADV.)
gwybodaeth E (N.)	her E (N.)	iechyd E (N.)
gwych ANS (ADJ.)	hir ANS (ADJ.)	ieithyddol ANS (ADJ.)
gŵyl E (N.)	hoff ANS (ADJ.)	ifanc ANS (ADJ.)
gwylio B (V.)	hoffi B (V.)	is ANS (ADJ.)
gwyllt ANS (ADJ.)	hogyn E (N.)	is-deitl E (N.)
gwyn ANS (ADJ.)	holi B (V.)	isod ADF (ADV.)
gwynt E (N.)	holl ANS (ADJ.)	jest ADF (ADV.)
gyrru B (V.)	hollol ANS (ADJ.)	jyst ADF (ADV.)
hacio B (V.)	hwinnw ANS (ADJ.)	lan ADF (ADV.)
haf E (N.)	hwnt ANS (ADJ.)	ledled ADF (ADV.)
hanes E (N.)	hwyl E (N.)	lefel E (N.)
hanesyddol ANS (ADJ.)	hybu B (V.)	licio B (V.)
hanfodol ANS (ADJ.)	hyd E (N.)	lot E (N.)
hanner E (N.)	hyfryd ANS (ADJ.)	lladd B (V.)
hapus ANS (ADJ.)	hyfforddiant E (N.)	llafur E (N.)
haul E (N.)	hŷn ANS (ADJ.)	llais E (N.)
hawdd ANS (ADJ.)	hynod ANS (ADJ.)	llaw E (N.)
		llawn ANS (ADJ.)
		llawr ADF (ADV.)

llawr E (N.)	llywodraeth E (N.)	meic E (N.)
lle ANS (ADJ.)	mab E (N.)	meithrin B (V.)
lle E (N.)	mae B (V.)	menter E (N.)
llenwi B (V.)	maes E (N.)	menyw E (N.)
llynyddiaeth E (N.)	magu B (V.)	merch E (N.)
lleol ANS (ADJ.)	mai E (N.)	mercher E (N.)
lleoliad E (N.)	maint E (N.)	mesur E (N.)
lles E (N.)	mam E (N.)	methu B (V.)
llinell E (N.)	man E (N.)	milwr E (N.)
lliw E (N.)	mân ANS (ADJ.)	milltir E (N.)
llong E (N.)	manwl ANS (ADJ.)	mis E (N.)
llu E (N.)	manylyn E (N.)	MOR ADF (ADV.)
llun E (N.)	map E (N.)	môr E (N.)
llunio B (V.)	marchnad E (N.)	moyn E (N.)
llwybr E (N.)	marw ANS (ADJ.)	mudiad E (N.)
llwyddiannus ANS (ADJ.)	marwolaeth E (N.)	munud E (N.)
llwyddiant E (N.)	màs E (N.)	mwyn E (N.)
llwyddo B (V.)	masnach E (N.)	mwynhau B (V.)
llwyfan E (N.)	mater E (N.)	myfyriwr E (N.)
llwyr ANS (ADJ.)	math E (N.)	mynd B (V.)
llyfr E (N.)	mawr ANS (ADJ.)	mynediad E (N.)
llyfrgell E (N.)	mawrth E (N.)	mynnu B (V.)
llyfu B (V.)	medru B (V.)	mynychu B (V.)
llygad E (N.)	medd B (V.)	mynydd E (N.)
llys E (N.)	meddwl B (V.)	nadolig E (N.)
llythyr E (N.)	meddwl E (N.)	natur E (N.)
		naturiol ANS (ADJ.)
		nawr ADF (ADV.)

neges E (N.)	ôl E (N.)	perffaith ANS (ADJ.)
neidio B (V.)	olaf ANS (ADJ.)	perfformiad E (N.)
neis ANS (ADJ.)	osgoi B (V.)	perfformio B (V.)
neuadd E (N.)	pa ADF (ADV.)	person E (N.)
newid B (V.)	pam ADF (ADV.)	personol ANS (ADJ.)
newidiad E (N.)	papur E (N.)	perthnasol ANS (ADJ.)
newydd ANS (ADJ.)	paratoi B (V.)	perthynas E (N.)
newyddion E (N.)	parod ANS (ADJ.)	perygl E (N.)
nifer E (N.)	parti E (N.)	peth E (N.)
nod E (N.)	partner E (N.)	plaid E (N.)
nodi B (V.)	parhau B (V.)	pleidlais E (N.)
nodwedd E (N.)	pasg E (N.)	pleidleisio B (V.)
nofel E (N.)	patrwm E (N.)	plentyn E (N.)
nôl B (V.)	pecyn E (N.)	pob ANS (ADJ.)
nos E (N.)	peidio B (V.)	pobl E (N.)
noson E (N.)	pêl E (N.)	poblogaidd ANS (ADJ.)
ochr E (N.)	pell ADF (ADV.)	pobman E (N.)
oed E (N.)	pell ANS (ADJ.)	pobol E (N.)
oedfa E (N.)	pen E (N.)	poeni B (V.)
oedolyn E (N.)	pen-blwydd E (N.)	polisi E (N.)
oen E (N.)	penderfyniad E (N.)	popeth E (N.)
oer ANS (ADJ.)	penderfynu B (V.)	posib ANS (ADJ.)
oes E (N.)	pennod E (N.)	posibl ANS (ADJ.)
ofn E (N.)	penodol ANS (ADJ.)	post E (N.)
ogystal ANS (ADJ.)	pentref E (N.)	prawf E (N.)
ôl ANS (ADJ.)	penwythnos E (N.)	presennol ANS (ADJ.)
		prif ANS (ADJ.)
		prifysgol E (N.)

prin ANS (ADJ.)	rŵan ADF (ADV.)	saesneg E (N.)
priodas E (N.)	rygbi E (N.)	safbwynt E (N.)
pris E (N.)	rhaglen E (N.)	saffle E (N.)
problem E (N.)	rhaid E (N.)	safon E (N.)
profi B (V.)	rhan E (N.)	sail E (N.)
profiad E (N.)	rhannu B (V.)	sain E (N.)
proses E (N.)	rhedeg B (V.)	sant E (N.)
prosiect E (N.)	rheol E (N.)	sector E (N.)
pryd E (N.)	rheoli B (V.)	sefydliad E (N.)
prynhawn E (N.)	rhestr E (N.)	sefydlu B (V.)
prynu B (V.)	rheswm E (N.)	sefyll B (V.)
prysur ANS (ADJ.)	rhiant E (N.)	sefyllfa E (N.)
punt E (N.)	rhif E (N.)	senedd E (N.)
pŵer E (N.)	rhifyn E (N.)	seren E (N.)
pwll E (N.)	rhoddi B (V.)	sesiwn E (N.)
pwnc E (N.)	rhoi B (V.)	sgil E (N.)
pwyl E (N.)	rhy ADF (ADV.)	sgil E (N.)
pwyllog E (N.)	rhyddhau B (V.)	sgwennu B (V.)
pwylt E (N.)	rhyddid E (N.)	sgwrs E (N.)
pwys E (N.)	rhyfedd ANS (ADJ.)	sgwrsio B (V.)
pwysig ANS (ADJ.)	rhyfel E (N.)	siarad B (V.)
radio E (N.)	rhyngwladol ANS (ADJ.)	siaradwr E (N.)
ras E (N.)	rhyw E (N.)	sicr ANS (ADJ.)
recordio B (V.)	rhywbeth E (N.)	sicrhau B (V.)
rôl E (N.)	rhywun E (N.)	sioe E (N.)
rownd E (N.)	sadwrn E (N.)	siop E (N.)
		siopa B (V.)
		sir E (N.)

siŵr ANS (ADJ.)	syniad E (N.)	tewi B (v.)
siŵr ANS (ADJ.)	system E (N.)	teyrnas E (N.)
so ADF (ADV.)	syth ANS (ADJ.)	tîm E (N.)
sôn B (v.)	tad E (N.)	tipyn E (N.)
sôn E (N.)	tafarn E (N.)	tir E (N.)
stafell E (N.)	taflu B (v.)	tiwtor E (N.)
staff E (N.)	tag E (N.)	to E (N.)
stopio B (v.)	taith E (N.)	torri B (v.)
stori E (N.)	taliad E (N.)	traddodiad E (N.)
strategaeth E (N.)	talw B (v.)	traddodiadol ANS (ADJ.)
stryd E (N.)	tân E (N.)	traed E (N.)
sul E (N.)	tasg E (N.)	traeth E (N.)
sŵn E (N.)	tawel ANS (ADJ.)	trafod B (v.)
swydd E (N.)	te E (N.)	trafod E (N.)
swyddfa E (N.)	tebyg ANS (ADJ.)	trafodaeth E (N.)
swyddog E (N.)	tebygol ANS (ADJ.)	tramor ANS (ADJ.)
swyddogol ANS (ADJ.)	technoleg E (N.)	tre E (N.)
sydyn ANS (ADJ.)	teg ANS (ADJ.)	tref E (N.)
sylfaen E (N.)	teimlad E (N.)	trefn E (N.)
sylw E (N.)	teimlo B (v.)	trefnu B (v.)
sylwad E (N.)	teitl E (N.)	trên E (N.)
sylweddol ANS (ADJ.)	teithio B (v.)	treulio B (v.)
sylweddoli B (v.)	teledu E (N.)	trin B (v.)
sylwi B (v.)	term E (N.)	trio B (v.)
syml ANS (ADJ.)	testun E (N.)	tro E (N.)
symud B (v.)	teulu E (N.)	troi B (v.)
		trosglwyddo B (v.)
		trwydded E (N.)

tu E (N.)	unwaith ADF (ADV.)	ymsgyrch E (N.)
tudalen E (N.)	urdd E (N.)	ymlaen ADF (ADV.)
tŷ E (N.)	uwch ANS (ADJ.)	ymuno B (V.)
tybed ADF (ADV.)	wal E (N.)	ymweld B (V.)
tyfu B (V.)	wedi ADF (ADV.)	ymweliad E (N.)
tymor E (N.)	wedyn ADF (ADV.)	ymwneud B (V.)
tynnu B (V.)	weithiau ADF (ADV.)	ymwybodol ANS (ADJ.)
tystiolaeth E (N.)	wyneb E (N.)	yn ADF (ADV.)
tywydd E (N.)	wynebu B (V.)	yna ADF (ADV.)
theatr E (N.)	ŵyr E (N.)	ynys E (N.)
thema E (N.)	wythnos E (N.)	ysbryd E (N.)
uchel ANS (ADJ.)	ychwanegol ANS (ADJ.)	ysbyty E (N.)
uchod ADF (ADV.)	ychwanegu B (V.)	ysgol E (N.)
undeb E (N.)	ynghyd ADF (ADV.)	ysgrifennu B (V.)
unig ANS (ADJ.)	yma ADF (ADV.)	ysgrifennydd E (N.)
unigol ANS (ADJ.)	ymarfer E (N.)	ystafell E (N.)
unigolyn E (N.)	ymateb E (N.)	ystod E (N.)
unigryw ANS (ADJ.)	ymchwil E (N.)	ystyr E (N.)
union ANS (ADJ.)	ymdrech E (N.)	ystyried B (V.)
unrhyw ANS (ADJ.)	ymddangos B (V.)	

C.7 500 enw uchaf yn CorCenCC

Top 500 noun lemmas in CorCenCC

1. cymraeg	23. plaid	45. rhyw
2. dim	24. mis	46. gwasanaeth
3. peth	25. mwyn	47. gwlad
4. pobl	26. tro	48. saesneg
5. rhan	27. llyfr	49. cyfnod
6. gwaith	28. cynnwys	50. bywyd
7. blwyddyn	29. cwrs	51. cig
8. iaith	30. rhaglen	52. nifer
9. cyfer	31. diolch	53. cal
10. plentyn	32. aelod	54. gwefan
11. ysgol	33. gair	55. ystod
12. byd	34. de	56. newyddion
13. dydd	35. oed	57. wythnos
14. rhaid	36. llywodraeth	58. cyd
15. ôl	37. gwybodaeth	59. teulu
16. tŷ	38. cyfle	60. cwmni
17. angen	39. hanes	61. meddwl
18. amser	40. bodd	62. pryd
19. ffordd	41. hyd	63. tu
20. enw	42. llun	64. arian
21. pen	43. man	65. diwedd
22. lle	44. diwrnod	66. dechrau

67. ardal	93. sŵn	119. canol
68. cynulliad	94. sylw	120. dewis
69. cilydd	95. maes	121. sir
70. cynllun	96. cyfrwng	122. prosiect
71. stori	97. oen	123. profiad
72. gallu	98. eisiau	124. cartref
73. mam	99. enghraifft	125. gogledd
74. syniad	100. cerddoriaeth	126. problem
75. taith	101. addysg	127. grŵp
76. pobol	102. hanner	128. ymchwil
77. merch	103. data	129. sgwrs
78. cyngor	104. hawl	130. ymateb
79. blaen	105. cwestiwn	131. gêm
80. rhywbeth	106. cyfres	132. papur
81. tân	107. bore	133. cymorth
82. nos	108. rhywun	134. clwb
83. bwyd	109. gwybod	135. prifysgol
84. ffrind	110. cynnig	136. dosbarth
85. llaw	111. eisteddfod	137. llawr
86. dyn	112. iechyd	138. hwyl
87. mudiad	113. cofnod	139. cymuned
88. tîm	114. cyfarfod	140. canlyniad
89. polisi	115. digon	141. popeth
90. eglwys	116. digwyddiad	142. ateb
91. busnes	117. cymdeithas	143. gofal
92. awr	118. duw	144. llais
		145. swydd
		146. rhifyn

147. canolfan	173. cwbl	199. trafod
148. mater	174. testun	200. athro
149. ochr	175. e-bost	201. erthygl
150. rhestr	176. mai	202. disgybl
151. cefndir	177. barn	203. myfyriwr
152. dyfodol	178. pwyllgor	204. proses
153. cefn	179. nadolig	205. cysylltiad
154. sul	180. cam	206. system
155. tad	181. cyfeiriad	207. corff
156. munud	182. neges	208. safle
157. adroddiad	183. diddordeb	209. llafur
158. gŵyl	184. darn	210. effaith
159. rheswm	185. swyddog	211. sefyllfa
160. tipyn	186. dŵr	212. ffwrdd
161. rhyfel	187. car	213. adeilad
162. siop	188. rhiant	214. cân
163. person	189. ffaith	215. mynydd
164. bwrdd	190. gwobr	216. staff
165. drws	191. swyddfa	217. llyfrgell
166. tir	192. golwg	218. llygad
167. blog	193. chwarae	219. côr
168. arfer	194. defnydd	220. canu
169. gweinidog	195. môr	221. rownd
170. math	196. capel	222. safon
171. tywydd	197. pwynt	223. archif
172. canrif	198. cwmpas	224. lefel
		225. cymro
		226. gwir

227. adran	253. cwci	279. cynulleidfa
228. manylyn	254. adeg	280. brawd
229. criw	255. ysbyty	281. galw
230. coleg	256. sesiwn	282. ennill
231. perthynas	257. sioe	283. cenedl
232. fferm	258. awdur	284. is-deitl
233. comisiynydd	259. diwydiant	285. gradd
234. unigolyn	260. dull	286. pwys
235. lot	261. marchnad	287. bai
236. undeb	262. gwraig	288. pwnc
237. cwm	263. noson	289. tymor
238. tref	264. trafodaeth	290. coed
239. gweithgaredd	265. cefnogaeth	291. lleoliad
240. iawn	266. radio	292. llwybr
241. cerdd	267. sant	293. stryd
242. siaradwr	268. sadwrn	294. cais
243. cof	269. croeso	295. fideo
244. sefydliad	270. gwerth	296. sôn
245. ffilm	271. dinas	297. sylwad
246. tudalen	272. llythyr	298. cylch
247. dadl	273. cystadleuaeth	299. oedolyn
248. map	274. gardd	300. ymgyrch
249. achos	275. nod	301. pentref
250. cyfan	276. penderfyniad	302. ffurflen
251. cost	277. îa	303. sgil
252. casgliad	278. arweinydd	304. pecyn
		305. bardd
		306. ystafell

307. adnodd	333. sail	359. trefn
308. moyn	334. agwedd	360. gall
309. gŵr	335. datblygiad	361. help
310. ffurf	336. ymdrech	362. astudiaeth
311. rygbi	337. deunydd	363. milltir
312. ofn	338. mawrth	364. rheol
313. bod	339. safbwynt	365. technoleg
314. calon	340. cyfraniad	366. gwahaniaeth
315. cerdyn	341. dad	367. heddlu
316. gweddill	342. lliw	368. bachgen
317. categori	343. pwy	369. llwyddiant
318. anifail	344. dogfen	370. arglwydd
319. nofel	345. datganiad	371. etholiad
320. cariad	346. ci	372. cychwyn
321. afon	347. elfen	373. canllaw
322. senedd	348. awdurdod	374. oes
323. diwylliant	349. llwyfan	375. cinio
324. gwely	350. hyfforddiant	376. drama
325. cyw	351. cynnyrch	377. cyfrol
326. cwpan	352. bro	378. cylchgrawn
327. taliad	353. ffarmwr	379. carreg
328. natur	354. gwers	380. ŵyr
329. dysgwr	355. tystiolaeth	381. ysgrifennydd
330. tiwtor	356. cwrdd	382. punt
331. wyneb	357. cytundeb	383. disgwyl
332. ben	358. haf	384. llinell
		385. cyfrifoldeb
		386. gorllewin

387. gwasg	413. castell	439. post
388. term	414. bach	440. gafael
389. sain	415. haul	441. seren
390. brenin	416. diffyg	442. cynhyrchydd
391. babi	417. gweithiwr	443. arwydd
392. pobman	418. bath	444. fersiwn
393. sector	419. masnach	445. eidion
394. eitem	420. mynediad	446. dafad
395. dyddiad	421. as	447. theatr
396. parti	422. amrywiaeth	448. claf
397. menyw	423. wal	449. ymarfer
398. iau	424. te	450. pêl
399. gwirionedd	425. ystyr	451. llys
400. gwae	426. pris	452. cae
401. boi	427. meic	453. traed
402. crefydd	428. tasg	454. her
403. ynys	429. cacen	455. cadair
404. awyr	430. ymweliad	456. màs
405. deddf	431. mab	457. bryn
406. prynhawn	432. band	458. gwynt
407. rôl	433. tag	459. teyrnas
408. cyfraith	434. copi	460. ansawdd
409. cloch	435. dwyrain	461. chwaer
410. perfformiad	436. bwriad	462. thema
411. gobaith	437. dilledyn	463. ceffyl
412. cynhadledd	438. economi	464. cylchoedd
		465. eiliad
		466. marwolaeth

467. to	479. banc	491. partner
468. gwener	480. gorfod	492. stafell
469. llu	481. teimlad	493. cyfrifiad
470. teledu	482. heol	494. sgil
471. gwau	483. pennod	495. bocs
472. aderyn	484. gwas	496. lles
473. bryd	485. egwyddor	497. llenyddiaeth
474. oedfa	486. elusen	498. sylfaen
475. cyfaill	487. maint	499. amod
476. priodas	488. tafarn	500. adolygiad
477. brwydr	489. nodwedd	
478. pwell	490. comisiwn	

C.8 500 berf uchaf yn CorCenCC

Top 500 verb lemmas in CorCenCC

1. bod	23. dal	45. dechrau
2. cael	24. clywed	46. chwilio
3. gwneud	25. cymryd	47. cyflwyno
4. mynd	26. digwydd	48. cysylltu
5. dod	27. creu	49. helpu
6. mae	28. newid	50. sicrhau
7. gweld	29. gadael	51. symud
8. dweud	30. cadw	52. erbyn
9. gallu	31. dangos	53. meithrin
10. rhoi	32. cynnal	54. hoffi
11. defnyddio	33. medd	55. datblygu
12. meddwl	34. credu	56. aros
13. gwybod	35. dilyn	57. deall
14. siarad	36. derbyn	58. peidio
15. edrych	37. eisiau	59. rhannu
16. gweithio	38. teimlo	60. golygu
17. codi	39. tynnu	61. cefnogi
18. dysgu	40. troi	62. colli
19. gofyn	41. cyrraedd	63. ystyried
20. byw	42. dylu	64. ysgrifennu
21. darllen	43. cyhoeddi	65. mwynhau
22. cofio	44. ceisio	66. medru

67. nodi	93. rhedeg	119. gorfod
68. talu	94. darparu	120. cofnodi
69. parhau	95. cau	121. darganfod
70. torri	96. rheoli	122. nôl
71. trio	97. gweithredu	123. cynllunio
72. arwain	98. casglu	124. gyrru
73. cerdded	99. sefyll	125. wynebu
74. trefnu	100. methu	126. tewi
75. agor	101. adeiladu	127. diogelu
76. gobeithio	102. llwyddo	128. hyrwyddo
77. prynu	103. gwylio	129. geni
78. penderfynu	104. teithio	130. profi
79. ymuno	105. tyfu	131. licio
80. sefydlu	106. paratoi	132. dewis
81. ymweld	107. ychwanegu	133. coginio
82. gosod	108. eistedd	134. mynnu
83. adnabod	109. rhoddi	135. cyflawni
84. galw	110. cyfrannu	136. sylweddoli
85. cynhyrchu	111. cyfeirio	137. treulio
86. dathlu	112. gwella	138. disgrifio
87. cytuno	113. awgrymu	139. sylwi
88. cyfeithu	114. sôn	140. gwisgo
89. anfon	115. chwerthin	141. disgwyl
90. ymddangos	116. gwerthu	142. caniatáu
91. bwyta	117. lladd	143. ymwneud
92. gwranddo	118. sgwennu	144. dychwelyd
		145. astudio
		146. gorffen

147. poeni	173. rhyddhau	199. archwilio
148. aeth	174. cysgu	200. cynrychioli
149. hybu	175. darlledu	201. dwyn
150. holi	176. croesawu	202. denu
151. adrodd	177. perfformio	203. ffurfio
152. esbonio	178. cychwyn	204. recordio
153. cymharu	179. cario	205. hacio
154. trin	180. magu	206. diflannu
155. cofrestru	181. achub	207. diddymu
156. cynnwys	182. bodoli	208. dymuno
157. ffonio	183. ennill	209. mynychu
158. cynyddu	184. amddiffyn	210. cydymffurfio
159. canfod	185. arfer	211. addysgu
160. anghofio	186. gorfodi	212. llyfu
161. llenwi	187. dibynnu	213. atal
162. cynnig	188. ateb	214. stopio
163. caru	189. pleidleisio	215. siopa
164. canolbwyntio	190. trosglwyddo	216. taflu
165. llunio	191. cyfrif	217. archebu
166. bathu	192. neidio	218. gwenu
167. trafod	193. effeithio	219. cwblhau
168. gwasanaethu	194. sgwrsio	220. lansio
169. clicio	195. cystadlu	221. achosi
170. canu	196. osgoi	222. hyfforddi
171. chwarae	197. bwrw	223. galluogi
172. cydnabod	198. ffeindio	224. cyfleu
		225. bwriadu
		226. cyfathrebu

227. gwarchod	253. mentro	279. para
228. pasio	254. datgelu	280. arddangos
229. cynghori	255. sïo	281. tasu
230. pwyso	256. mynegi	282. gall
231. delio	257. ymestyn	283. dioddef
232. croesi	258. pori	284. pesychu
233. dawnsio	259. cwyno	285. honni
234. gwario	260. dringo	286. anwybyddu
235. dychmygu	261. cydweithio	287. teledu
236. egluro	262. datganoli	288. cloi
237. dileu	263. gwthio	289. lleoli
238. addasu	264. dlid	290. bario
239. cuddio	265. camu	291. parcio
240. adlewyrchu	266. golchi	292. datgan
241. priodi	267. swnio	293. malu
242. cadarnhau	268. glynu	294. diweddarau
243. elwa	269. herio	295. enwi
244. cyflogi	270. ymchwilio	296. nofio
245. dadansoddi	271. prosesu	297. buddsoddi
246. adolygu	272. llifo	298. brwydro
247. lleihau	273. llwytho	299. cymysgu
248. gofalu	274. crwydro	300. cwympo
249. gweiddi	275. hedfan	301. beirniadu
250. marchnata	276. disgyn	302. dosbarthu
251. yfed	277. marw	303. gwerthfawrogi
252. hysbysebu	278. pwysleisio	304. curo
		305. tueddu
		306. diffinio

307. ysbrydoli	333. synnu	359. amlygu
308. claddu	334. postio	360. sgorio
309. llechu	335. haeddu	361. darfod
310. ffermio	336. llosgi	362. cipio
311. ariannu	337. uno	363. boddi
312. clirio	338. allforio	364. cymeradwyo
313. dianc	339. syrthio	365. addurno
314. anelu	340. syllu	366. seilio
315. dotio	341. gorwedd	367. maddau
316. gwasgu	342. blasu	368. ymosod
317. cymdeithasu	343. peri	369. amrywio
318. dylanwadu	344. ymlacio	370. comisiynu
319. datrys	345. apelio	371. bwydo
320. annog	346. ehangu	372. hawlio
321. blogio	347. asesu	373. danfon
322. ymddeol	348. sonio	374. saethu
323. noddi	349. rhegi	375. addo
324. penodi	350. gwahanu	376. trawsgrifio
325. mabwysiadu	351. rhwystro	377. pigo
326. manteisio	352. gwrthod	378. goroesi
327. chwalu	353. cyrchu	379. ymgyrchu
328. oedi	354. eithrio	380. diolch
329. cynorthwyo	355. deillio	381. hunanasesu
330. mewngofnodi	356. pallu	382. tanio
331. meddiannu	357. rhestru	383. sbio
332. dehongli	358. pregethu	384. dyfynnu
		385. atgoffa
		386. rhybuddio

387. ymdopi	413. menu	439. integreiddio
388. dylunio	414. betio	440. cywiro
389. glanhau	415. cyfuno	441. cydio
390. cyffwrdd	416. taro	442. pedoli
391. gwrthwynebu	417. garddio	443. plannu
392. ymrwymo	418. sgrifennu	444. begio
393. beicio	419. arwyddo	445. plygu
394. deffro	420. ofni	446. tywys
395. cyfansoddi	421. cyfaddef	447. toi
396. crio	422. perthyn	448. paid
397. cennu	423. bodloni	449. barnu
398. amlinellu	424. canslo	450. ysgwyd
399. olrhain	425. anghytuno	451. meddu
400. cryfhau	426. dyddio	452. dethol
401. cludo	427. ethol	453. lawrlwytho
402. gwirio	428. cyfyngu	454. gwahaniaethu
403. blino	429. gwirfoddoli	455. perswadio
404. dadlau	430. costio	456. addoli
405. argymell	431. argraffu	457. bachu
406. gwireddu	432. trochi	458. gostwng
407. ffilmio	433. plesio	459. parchu
408. hwylio	434. arbed	460. dyfalu
409. joio	435. pobi	461. graddio
410. cwrdd	436. udo	462. hwyluso
411. sychu	437. actio	463. ymateb
412. glanio	438. cyflenwi	464. teipio
		465. brifo
		466. crynhoi

467. gwahodd	479. cwestiynu	491. ymgysylltu
468. cyfnewid	480. cyfiawnhau	492. gweddïo
469. anadlu	481. crafu	493. clymu
470. llefaru	482. recriwtio	494. ailadrodd
471. chwythu	483. dinistrio	495. gwn
472. tybio	484. llusgo	496. estyn
473. mesur	485. llyncu	497. hysbysu
474. dynodi	486. rasio	498. ymgynghori
475. llywio	487. craffu	499. piau
476. trial	488. cynilo	500. cyfarfod
477. cigyddio	489. ildio	
478. blero	490. pennu	

C.9 500 ansoddair uchaf yn CorCenCC

Top 500 adjective lemmas in CorCenCC

1. da	23. unig	45. coch
2. mawr	24. bynnag	46. is
3. arall	25. amlwg	47. ariannol
4. newydd	26. hir	48. penodol
5. bach	27. drwg	49. diddorol
6. ôl	28. parod	50. cyhoeddus
7. agos	29. anodd	51. hawdd
8. hen	30. aneglur	52. cynnar
9. cyntaf	31. ogystal	53. gwreiddiol
10. gwahanol	32. pob	54. hapus
11. ifanc	33. diwethaf	55. uchel
12. pwysig	34. cymaint	56. union
13. llawn	35. cyffredinol	57. neis
14. arbennig	36. gwyach	58. du
15. prif	37. olaf	59. personol
16. lleol	38. lle	60. cyffredin
17. cenedlaethol	39. cymreig	61. blynyddol
18. iawn	40. cymdeithasol	62. sicr
19. siŵr	41. pell	63. uwch
20. unrhyw	42. eng	64. gwleidyddol
21. diweddar	43. aml	65. enwedig
22. tebyg	44. posib	66. hoff

67. cryf	93. perthnasol	119. cyfarwydd
68. cywir	94. ychwanegol	120. enwog
69. hollol	95. ewropeaidd	121. cynradd
70. eang	96. poblogaidd	122. unigryw
71. hwnnw	97. holl	123. swyddogol
72. blaenorol	98. allweddol	124. agored
73. hyfryd	99. glas	125. addas
74. naturiol	100. cyd	126. buan
75. braf	101. syml	127. sydyn
76. dwyieithog	102. teg	128. gwyn
77. prysur	103. hwnt	129. annibynnol
78. clir	104. syth	130. marw
79. posibl	105. sylweddol	131. economaidd
80. byw	106. perffaith	132. tramor
81. cyflym	107. prin	133. gwyllt
82. ar-lein	108. hanesyddol	134. ymwybodol
83. cyfrifol	109. de	135. byr
84. llwyr	110. ieithyddol	136. chwith
85. presennol	111. llwyddiannus	137. bodlon
86. caled	112. diweddaraf	138. siwr
87. tawel	113. bychan	139. canlynol
88. rhyngwladol	114. hŷn	140. manwl
89. mân	115. angenrheidiol	141. hanfodol
90. hynod	116. rhyfedd	142. tebygol
91. cyfan	117. cyson	143. unigol
92. defnyddiol	118. anffodus	144. enfawr
		145. difyr
		146. gofalus

147. oer	173. modern	199. rhyfeddol
148. effeithiol	174. crefyddol	200. eithriadol
149. cyffrous	175. unol	201. od
150. traddodiadol	176. cynnes	202. sawl
151. uniongyrchol	177. prydeinig	203. proffesiynol
152. helaeth	178. uwchradd	204. cadarn
153. ofnadwy	179. diwylliannol	205. lliwgar
154. digidol	180. glân	206. adnabyddus
155. cwbl	181. dynol	207. gweithredol
156. terfynol	182. cynt	208. gweddol
157. araf	183. uchaf	209. agor
158. iach	184. gwael	210. fer
159. awyddus	185. cymunedol	211. seiliedig
160. priodol	186. trwm	212. ysgafn
161. amrywiol	187. arferol	213. brenhinol
162. anhygoel	188. cysylltiedig	214. gonest
163. isel	189. gwledig	215. seneddol
164. diogel	190. ymarferol	216. ysgrifenedig
165. creadigol	191. difrifol	217. rhoddi
166. tywyll	192. cyfoes	218. byd-eang
167. unedig	193. sylfaenol	219. blasus
168. cyfreithiol	194. ardderchog	220. annwyl
169. pendant	195. sengl	221. fechan
170. cwawl	196. gwyrdd	222. anghywir
171. gwag	197. trist	223. gwahân
172. balch	198. canolog	224. gwerthfawr
		225. cam
		226. cymhleth

227. academaidd	253. gwirioneddol	279. rhad
228. llawen	254. penderfynol	280. dwfn
229. sâl	255. cymharol	281. cyfforddus
230. llwyd	256. rhesymol	282. integredig
231. poeth	257. annisgwyl	283. ffodus
232. piws	258. technegol	284. bendigedig
233. preifat	259. brwd	285. diwydiannol
234. rheolaidd	260. difrif	286. priod
235. hyderus	261. rhannol	287. diflas
236. pwysicaf	262. mewnl	288. sefydlog
237. dyddiol	263. cadarnhaol	289. garw
238. cefndir	264. blin	290. melys
239. blaen	265. anferth	291. teulu
240. melyn	266. parhaol	292. isaf
241. cerddorol	267. cynyddol	293. reit
242. cryn	268. hud	294. cyfartal
243. corfforol	269. symudol	295. amgen
244. llaw	270. gofynnol	296. hollbwysig
245. cymwys	271. gŵyl	297. trawiadol
246. strategol	272. meddygol	298. ysbrydol
247. lwcus	273. cyfoethog	299. misol
248. ffurfiol	274. rhwydd	300. profiadol
249. llenyddol	275. hwyr	301. saff
250. lleia	276. dwys	302. cyfredol
251. doniol	277. allanol	303. hardd
252. peryglus	278. rhydd	304. ffyddlon
		305. delfrydol
		306. dibynnol

307. rhugl	333. golygyddol	359. ffres
308. moesol	334. statudol	360. cyfyngedig
309. amaethyddol	335. pinc	361. ynghlwm
310. rhagorol	336. uniaith	362. dylanwadol
311. masnachol	337. twt	363. rhywiol
312. cyflawn	338. nosweithiol	364. pwrpasol
313. derbyniol	339. tlawd	365. dewr
314. swnllyd	340. cysgod	366. parhaus
315. call	341. dilynol	367. anarferol
316. diolchgar	342. parchedig	368. sirol
317. digonol	343. eithafol	369. euog
318. brown	344. cul	370. gwladol
319. safonol	345. ethnig	371. hwyl
320. gwirfoddol	346. lleiafrifol	372. cymysg
321. rhyddfrydol	347. maith	373. cwrs
322. moel	348. erchyll	374. cristnogol
323. gwybyddol	349. arloesol	375. cyfrinachol
324. meddal	350. teithiol	376. ystadegol
325. blaenllaw	351. corfforaethol	377. wythnosol
326. llonydd	352. hysbys	378. awtomatig
327. heriol	353. positif	379. cynhenid
328. negyddol	354. bywiog	380. anffurfiol
329. celtaidd	355. emosiynol	381. milwrol
330. arwyddocaol	356. sifil	382. domestig
331. tynn	357. heini	383. ceidwadol
332. amhosib	358. diduedd	384. electronig
		385. niwclear
		386. gwlyb

387. lluosog	413. main	439. americanaidd
388. tanddaearol	414. aruthrol	440. amlaf
389. nodedig	415. brodorol	441. tew
390. lyffi	416. bregus	442. gosod
391. gwyddonol	417. real	443. pwerus
392. ansicr	418. ifainc	444. cyfryw
393. rhif	419. di-ri	445. cynhwysol
394. benywaidd	420. agoriadol	446. cynaliadwy
395. amgylcheddol	421. eglur	447. eithaf
396. ardal	422. meddyliol	448. blydi
397. democrataidd	423. cynhwysfawr	449. tenau
398. llydan	424. tlws	450. dwl
399. bwriadol	425. canolig	451. gwan
400. etholiadol	426. chwech	452. pennaf
401. arbenigol	427. trwchus	453. cyfun
402. dilys	428. cofiadwy	454. nodweddiadol
403. cyfeillgar	429. gwrdd	455. addysgol
404. cu	430. daearyddol	456. crand
405. genedigol	431. gwastad	457. sensitif
406. uchelgeisiol	432. rhanbarthol	458. gweledol
407. hynafol	433. crwn	459. pur
408. cofrestredig	434. doeth	460. cychwynnol
409. brwdfrydig	435. hwy	461. cynifer
410. llinellol	436. llythrennol	462. cyfunol
411. clyfar	437. gweithgar	463. enwocaf
412. siomedig	438. distaw	464. hynaf
		465. dramatig
		466. sownd

467. ambell	479. amheus	491. gwir
468. amhosibl	480. llym	492. cryno
469. cudd	481. serth	493. clasurol
470. ffantastig	482. poen	494. gwerdd
471. sych	483. nerfus	495. effeithlon
472. cyfagos	484. cefnogol	496. ffyrnig
473. prydferth	485. disglair	497. anghyfreithlon
474. hyll	486. creulon	498. tueddol
475. allgymorth	487. normal	499. hoyw
476. hwylus	488. coll	500. niferus
477. rhufeinig	489. syfrdanol	
478. buddugol	490. brawychus	

C.10 50 adferf uchaf yn CorCenCC

Top 50 adverb lemmas in CorCenCC

- | | | |
|------------|-------------|--------------|
| 1. ie | 18. pam | 35. llawr |
| 2. dim | 19. ymlaen | 36. hytrach |
| 3. yma | 20. jyst | 37. byth |
| 4. hefyd | 21. rŵan | 38. lan |
| 5. yna | 22. efallai | 39. eleni |
| 6. felly | 23. pell | 40. ynghyd |
| 7. iawn | 24. wedi | 41. weithiau |
| 8. so | 25. jest | 42. adref |
| 9. wedyn | 26. unwaith | 43. eisoes |
| 10. yn | 27. pa | 44. eithaf |
| 11. dyna | 28. erioed | 45. isod |
| 12. mor | 29. fyny | 46. chwaith |
| 13. dyma | 30. bron | 47. tybed |
| 14. nawr | 31. rhy | 48. uchod |
| 15. eto | 32. draw | 49. fan yna |
| 16. heddiw | 33. go | 50. gwastad |
| 17. allan | 34. heno | |

C.11 50 ebychiad uchaf yn CorCenCC

Top 50 interjection lemmas in CorCenCC

1. ymm	18. yndê	35. diawl
2. yyy	19. waw	36. heia
3. wel	20. haha	37. damia
4. reit	21. eh	38. awê
5. iawn	22. hwrê	39. iww
6. gwir	23. duw	40. wff
7. ââ	24. bechod	41. sboso
8. oh	25. diar	42. joia
9. gwybod	26. ew	43. iym
10. mhm	27. helô	44. haleliwia
11. sori	28. gweld	45. aha
12. hmm	29. shh	46. shwsh
13. mmm	30. och	47. joiwch
14. helo	31. wps	48. ffyc
15. oce	32. iesu	49. doh
16. ôô	33. ffwc	50. dwnim
17. w	34. grr	

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

Safle (rank)	Gair (word)	Cyfanswm Crai (raw total)	C/swm y fliwn [d-a, ysg yn unig] (tot. per m. [o-c wr. only])	C/swm y fliwn [d-a cyfan] (tot. per m. [all o-c])	Cyfanswm y fliwn [corpws llawn] (tot. per m. [full cor- pus])
1	mae	45373	23708.54	27042.25	13701.34
2	oedd	20235	10573.30	8418.09	4265.14
3	bod	18157	9487.49	10315.49	5226.49
4	roedd	17437	9111.27	5428.22	2750.29
5	fod	14165	7401.57	6244.83	3173.98
6	yw	12086	6315.24	6483.65	3285.03
7	sy	12064	6303.74	6758.20	3424.14
8	ôl	9265	4841.20	4236.91	2146.69
9	cael	8539	4461.84	5112.20	2590.17

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
10	ddim	8280	4326.51	5416.19	2804.41
11	bydd	7396	3864.60	3644.86	1846.72
12	sydd	7388	3860.42	4450.73	2255.03
13	iawn	6917	3614.31	4343.54	2619.06
14	hefyd	6863	3586.09	3804.81	1927.76
15	mynd	6497	3394.85	4719.54	2391.22
16	iaith	6162	3219.80	2661.81	1348.64
17	arall	5484	2865.53	2647.37	1376.50
18	newydd	5361	2801.26	3395.86	1720.56
19	gael	5199	2716.61	2897.48	1468.05
20	gyfer	5174	2703.55	2693.10	1364.50
21	gymraeg	4962	2592.77	2983.57	1511.67
22	ysgol	4806	2511.26	1844.27	934.43
23	felly	4757	2485.65	3151.11	1596.56

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
24	rhaid	4755	2484.61	2144.56	1086.57
25	cymraeg	4494	2348.23	2743.45	1390.01
26	byd	4172	2179.98	2038.66	1032.92
27	fydd	4132	2159.07	2122.53	1075.41
28	dod	4106	2145.49	2633.67	1334.39
29	dim	4048	2115.18	2513.15	1837.15
30	dwi	3941	2059.27	2471.31	1252.13
31	gwneud	3865	2019.56	2000.89	1013.78
32	mwyn	3860	2016.95	1588.05	804.61
33	mor	3847	2010.15	1716.72	869.80
34	lle	3656	1910.35	1751.15	1490.09
35	bach	3622	1892.59	2116.05	1072.12
36	rhan	3609	1885.79	1839.46	931.99
37	hen	3576	1868.55	1398.85	708.75

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
38	gwaith	3574	1867.51	2039.22	1033.20
39	saesneg	3574	1867.51	1275.55	646.28
40	meddai	3571	1865.94	1015.26	514.39
41	yma	3564	1862.28	3014.86	1686.60
42	byddai	3564	1862.28	1296.66	656.97
43	wneud	3548	1853.92	1711.35	867.08
44	mawr	3399	1776.06	1478.08	748.89
45	oes	3357	1754.12	2031.63	1029.35
46	rhaglen	3356	1753.59	990.64	501.92
47	plant	3306	1727.47	1533.99	777.22
48	gweld	3258	1702.39	2055.32	1041.36
49	amser	3233	1689.32	1688.21	855.35
50	ffordd	3208	1676.26	1715.05	868.96
51	hyd	3206	1675.22	1338.68	1196.69

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
52	da	3190	1666.86	3440.47	1743.16
53	weld	3105	1622.44	1597.86	809.58
54	ben	3096	1617.74	1126.15	570.58
55	eto	3063	1600.49	1428.47	723.75
56	fynd	3031	1583.77	1439.39	729.29
57	yna	2985	1559.74	1949.06	1225.58
58	bu	2958	1545.63	1082.65	548.54
59	fawr	2847	1487.63	1411.25	715.03
60	rhoi	2816	1471.43	1843.16	933.86
61	angen	2794	1459.94	1675.06	848.69
62	o'n	2744	1433.81	3034.48	1537.46
63	oed	2719	1420.75	1376.45	697.40
64	doedd	2712	1417.09	758.85	384.48
65	fyddai	2704	1412.91	969.53	491.23

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
66	dyma	2678	1399.32	1784.66	904.22
67	meddwl	2653	1386.26	2489.83	1261.50
68	edrych	2631	1374.76	1546.95	783.79
69	arbennig	2628	1373.20	1061.73	537.94
70	tro	2611	1364.31	1202.80	609.41
71	gallu	2560	1337.66	1904.63	965.01
72	yno	2517	1315.20	872.71	494.98
73	ystod	2431	1270.26	1207.98	612.04
74	heddiw	2430	1269.74	1178.54	597.13
75	newyddion	2421	1265.03	1199.83	607.91
76	dweud	2407	1257.72	1366.27	692.24
77	nifer	2391	1249.36	1159.66	587.56
78	hanes	2388	1247.79	1303.69	660.53
79	allan	2374	1240.48	1370.15	694.21

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
80	enw	2365	1235.77	1437.17	728.16
81	llawn	2348	1226.89	948.98	480.81
82	eraill	2291	1197.11	1142.63	809.21
83	siarad	2274	1188.22	1706.54	864.64
84	dyna	2266	1184.04	1396.07	707.34
85	gilydd	2258	1179.86	1042.29	528.09
86	byw	2234	1167.32	1095.97	555.29
87	tu	2206	1152.69	977.68	495.35
88	wedyn	2186	1142.24	2201.21	1115.27
89	chwarae	2177	1137.54	876.59	444.14
90	ifanc	2152	1124.47	1028.03	520.87
91	dydd	2111	1103.05	1618.60	820.09
92	'na	2107	1100.96	3074.65	1959.28
93	dal	2095	1094.69	1051.73	532.87

C.12 100 gair dosbarth agored uchaf yng nghydran ysgrifenedig CorCenCC
 Top 100 open-class words in the written component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ysg.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
94	dysgu	2064	1078.49	928.06	470.22
95	cyntaf	2055	1073.79	939.91	579.96
96	ran	2043	1067.52	1042.66	528.28
97	defnyddio	2036	1063.86	1181.88	598.81
98	daeth	2028	1059.68	642.96	325.76
99	dw	1986	1037.74	3882.56	1967.16
100	bobl	1960	1024.15	1436.62	727.88

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

Safle (rank)	Gair (word)	Cyfanswm Crai (raw total)	C/swm y fliwn [d-a, lla yn unig] (tot. per m. [o-c sp. only])	C/swm y fliwn [d-a cyfan] (tot. per m. [all o-c])	Cyfanswm y fliwn [corpws llawn] (tot. per m. [full cor- pus])
1	mae	40162	33301.35	27042.25	13701.34
2	ma	26871	22280.78	5398.97	3046.13
3	ie	22917	19002.22	4427.96	2243.49
4	bod	17591	14586.03	10315.49	5226.49
5	dw	16129	13373.77	3882.56	1967.16
6	mynd	14085	11678.94	4719.54	2391.22
7	'na	13474	11172.31	3074.65	1959.28
8	ddim	12561	10415.28	5416.19	2804.41
9	so	12267	10171.50	2452.98	1242.84

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. IIa.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
10	o'n	11616	9631.71	3034.48	1537.46
11	da	11365	9423.58	3440.47	1743.16
12	oedd	11332	9396.22	8418.09	4265.14
13	neud	10981	9105.18	2302.84	1166.77
14	o'dd	10038	8323.27	2045.33	1036.29
15	ia	9841	8159.92	1949.24	987.61
16	iawn	9236	7658.27	4343.54	2619.06
17	sy	8579	7113.50	6758.20	3424.14
18	yw	8345	6919.47	6483.65	3285.03
19	wedyn	8083	6702.23	2201.21	1115.27
20	cael	7084	5873.88	5112.20	2590.17
21	meddwl	6886	5709.70	2489.83	1261.50
22	felly	6667	5528.11	3151.11	1596.56
23	dod	6336	5253.66	2633.67	1334.39

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. lla.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
24	fod	5875	4871.41	6244.83	3173.98
25	yma	5500	4560.47	3014.86	1686.60
26	dim	5414	4489.16	2513.15	1837.15
27	sydd	5402	4479.21	4450.73	2255.03
28	bo	5260	4361.46	1245.19	630.89
29	nawr	5019	4161.63	1449.20	734.26
30	hefyd	5009	4153.34	3804.81	1927.76
31	siarad	4776	3960.14	1706.54	864.64
32	gallu	4771	3956.00	1904.63	965.01
33	'ma	4694	3892.15	1077.09	603.13
34	ôl	4688	3887.18	4236.91	2146.69
35	bach	4497	3728.80	2116.05	1072.12
36	yna	4300	3565.46	1949.06	1225.58
37	gweld	4283	3551.36	2055.32	1041.36

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. IIa.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
38	ydy	4251	3524.83	1244.45	630.52
39	fan	4180	3465.95	1147.63	581.46
40	gwybod	4143	3435.27	1412.73	715.78
41	'm	3942	3268.61	1017.48	859.76
42	ca'i	3861	3201.45	768.85	389.55
43	de	3858	3198.96	1180.39	598.06
44	arall	3799	3150.04	2647.37	1376.50
45	diolch	3775	3130.14	1424.95	723.29
46	jyst	3673	3045.56	770.51	390.39
47	dyna	3532	2928.65	1396.07	707.34
48	aneglur	3519	2917.87	662.03	335.43
49	jest	3202	2655.02	662.21	335.52
50	edrych	2972	2464.31	1546.95	783.79
51	gyd	2829	2345.74	1196.50	606.22

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. lla.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
52	bydd	2828	2344.91	3644.86	1846.72
53	oes	2821	2339.10	2031.63	1029.35
54	eng	2820	2338.28	522.62	264.80
55	rhaid	2789	2312.57	2144.56	1086.57
56	do	2770	2296.82	766.44	388.33
57	lle	2763	2291.01	1751.15	1490.09
58	siŵr	2763	2291.01	897.70	454.83
59	rhoi	2718	2253.70	1843.16	933.86
60	byd	2706	2243.75	2038.66	1032.92
61	sa	2693	2232.97	552.25	279.80
62	dwi	2648	2195.66	2471.31	1252.13
63	fynd	2573	2133.47	1439.39	729.29
64	fydd	2546	2111.08	2122.53	1075.41
65	peth	2530	2097.81	1092.83	590.18

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. lla.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
66	newydd	2493	2067.14	3395.86	1720.56
67	pobl	2440	2023.19	1499.19	759.59
68	dyn	2397	1987.53	996.74	505.01
69	dan	2397	1987.53	916.21	906.85
70	dweud	2358	1955.20	1366.27	692.24
71	ydi	2356	1953.54	1022.11	517.87
72	lawr	2353	1951.05	965.83	489.35
73	allan	2297	1904.62	1370.15	694.21
74	falle	2280	1890.52	536.88	272.02
75	mor	2275	1886.37	1716.72	869.80
76	pam	2257	1871.45	1189.10	602.47
77	credu	2230	1849.06	775.88	393.11
78	gweud	2227	1846.57	455.05	230.56
79	dydd	2213	1834.97	1618.60	820.09

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. lla.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
80	amser	2207	1829.99	1688.21	855.35
81	weld	2201	1825.02	1597.86	809.58
82	dda	2189	1815.07	1069.13	541.69
83	fawr	2183	1810.09	1411.25	715.03
84	ffordd	2138	1772.78	1715.05	868.96
85	gael	2122	1759.51	2897.48	1468.05
86	gweithio	2119	1757.02	1178.17	596.94
87	dyma	2070	1716.39	1784.66	904.22
88	na	2022	1676.59	587.05	2859.94
89	dach	2008	1664.98	443.39	224.65
90	cerddoriaeth	2006	1663.33	594.27	301.10
91	cymraeg	2001	1659.18	2743.45	1390.01
92	gwrs	1977	1639.28	989.15	501.17
93	digwydd	1942	1610.26	924.17	468.25

C.13 100 gair dosbarth agored uchaf yng nghydran llafar CorCenCC
 Top 100 open-class words in the spoken component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. lla.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
94	gw'bod	1915	1587.87	363.97	184.41
95	gilydd	1899	1574.60	1042.29	528.09
96	gwaith	1895	1571.29	2039.22	1033.20
97	angen	1891	1567.97	1675.06	848.69
98	deud	1890	1567.14	486.34	246.41
99	wyt	1851	1534.80	650.36	329.52
100	bobl	1850	1533.98	1436.62	727.88

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

Safle (rank)	Gair (word)	Cyfanswm Crai (raw total)	C/swm y fliwn [d-a, ele yn unig] (tot. per m. [o-c el. only])	C/swm y fliwn [d-a cyfan] (tot. per m. [all o-c])	Cyfanswm y fliwn [corpws llawn] (tot. per m. [full cor- pus])
1	mae	60536	26530.12	27042.25	13701.34
2	bod	19972	8752.80	10315.49	5226.49
3	sy	15862	6951.58	6758.20	3424.14
4	yw	14591	6394.56	6483.65	3285.03
5	oedd	13904	6093.48	8418.09	4265.14
6	fod	13692	6000.57	6244.83	3173.98
7	cael	11991	5255.10	5112.20	2590.17
8	sydd	11251	4930.79	4450.73	2255.03
9	newydd	10489	4596.84	3395.86	1720.56

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
10	roedd	10312	4519.27	5428.22	2750.29
11	bydd	9464	4147.63	3644.86	1846.72
12	gymraeg	9357	4100.74	2983.57	1511.67
13	ól	8933	3914.92	4236.91	2146.69
14	hefyd	8680	3804.04	3804.81	1927.76
15	ddim	8415	3687.90	5416.19	2804.41
16	gael	8330	3650.65	2897.48	1468.05
17	cymraeg	8324	3648.02	2743.45	1390.01
18	gyfer	8021	3515.23	2693.10	1364.50
19	iawn	7309	3203.20	4343.54	2619.06
20	yma	7221	3164.63	3014.86	1686.60
21	dwi	6760	2962.59	2471.31	1252.13
22	iaith	6621	2901.68	2661.81	1348.64
23	felly	5597	2452.91	3151.11	1596.56

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
24	gwneud	5565	2438.88	2000.89	1013.78
25	gwaith	5546	2430.55	2039.22	1033.20
26	arall	5017	2198.72	2647.37	1376.50
27	rhan	5004	2193.02	1839.46	931.99
28	cynulliad	4960	2173.74	1019.89	516.74
29	mynd	4911	2152.26	4719.54	2391.22
30	dyma	4892	2143.94	1784.66	904.22
31	oes	4796	2101.86	2031.63	1029.35
32	fydd	4787	2097.92	2122.53	1075.41
33	cig	4731	2073.38	925.28	468.81
34	llywodraeth	4632	2029.99	1118.56	566.73
35	plaid	4611	2020.79	940.28	476.41
36	wneud	4554	1995.81	1711.35	867.08
37	rhoi	4422	1937.96	1843.16	933.86

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
38	dydd	4419	1936.64	1618.60	820.09
39	angen	4363	1912.10	1675.06	848.69
40	pobl	4195	1838.47	1499.19	759.59
41	byd	4134	1811.74	2038.66	1032.92
42	dim	4113	1802.54	2513.15	1837.15
43	rhaid	4040	1770.54	2144.56	1086.57
44	da	4029	1765.72	3440.47	1743.16
45	bobl	3950	1731.10	1436.62	727.88
46	mis	3939	1726.28	1270.74	643.84
47	ffordd	3918	1717.08	1715.05	868.96
48	meddwl	3910	1713.57	2489.83	1261.50
49	arian	3882	1701.30	1076.54	545.44
50	mwyn	3880	1700.42	1588.05	804.61
51	newyddion	3794	1662.73	1199.83	607.91

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
52	dod	3784	1658.35	2633.67	1334.39
53	enw	3735	1636.88	1437.17	728.16
54	hanes	3729	1634.25	1303.69	660.53
55	amser	3679	1612.33	1688.21	855.35
56	meithrin	3662	1604.88	769.03	389.64
57	mudiad	3649	1599.19	756.82	383.45
58	ystod	3611	1582.53	1207.98	612.04
59	cynnwys	3591	1573.77	1055.80	534.94
60	oed	3571	1565.00	1376.45	697.40
61	gweld	3561	1560.62	2055.32	1041.36
62	wefan	3502	1534.76	752.74	381.39
63	ysgol	3445	1509.78	1844.27	934.43
64	plant	3431	1503.65	1533.99	777.22
65	gwasanaeth	3424	1500.58	809.95	410.37

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
66	nifer	3341	1464.21	1159.66	587.56
67	weld	3325	1457.19	1597.86	809.58
68	bach	3311	1451.06	2116.05	1072.12
69	eraill	3306	1448.87	1142.63	809.21
70	yna	3243	1421.26	1949.06	1225.58
71	newid	3216	1409.42	1166.51	591.03
72	mor	3151	1380.94	1716.72	869.80
73	hyd	3145	1378.31	1338.68	1196.69
74	aelodau	3119	1366.91	842.53	426.88
75	iechyd	3059	1340.62	741.45	375.67
76	lle	3040	1332.29	1751.15	1490.09
77	mawr	3004	1316.51	1478.08	748.89
78	defnyddio	2999	1314.32	1181.88	598.81
79	chwilio	2996	1313.01	836.42	423.78

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
80	byddai	2980	1306.00	1296.66	656.97
81	gallu	2957	1295.92	1904.63	965.01
82	hen	2865	1255.60	1398.85	708.75
83	dw	2857	1252.09	3882.56	1967.16
84	eto	2813	1232.81	1428.47	723.75
85	gwybodaeth	2791	1223.17	760.15	385.14
86	lleol	2765	1211.77	767.74	388.99
87	edrych	2753	1206.51	1546.95	783.79
88	allan	2730	1196.43	1370.15	694.21
89	bu	2711	1188.11	1082.65	548.54
90	ardal	2695	1181.09	918.43	465.34
91	oen	2658	1164.88	515.77	261.32
92	ifanc	2657	1164.44	1028.03	520.87
93	cyntaf	2629	1152.17	939.91	579.96

C.14 100 gair dosbarth agored uchaf yng nghydran electronig CorCenCC
 Top 100 open-class words in the electronic component of CorCenCC

<i>sgfle</i>	<i>lemma</i>	<i>c/swm</i>	<i>c/s yf. dosb.-ag. ele.</i>	<i>c/s yf. dosb.-ag.</i>	<i>c/s yf. llawn</i>
94	wythnos	2617	1146.91	987.49	500.32
95	dweud	2615	1146.03	1366.27	692.24
96	heddiw	2608	1142.97	1178.54	597.13
97	gweithio	2605	1141.65	1178.17	596.94
98	diolch	2596	1137.71	1424.95	723.29
99	fawr	2593	1136.39	1411.25	715.03
100	nesaf	2561	1122.37	722.01	365.82