

# **The *Geirfan* wordlist: A Vocabulary list for adult learners of Welsh**

*Version 2*

*February 2023*

**Knight, D., Tovey-Walsh, B., Davies, E., Morris S. & Prosser, H.**



Corpws Cenedlaethol Cymraeg Cyfoes  
National Corpus of Contemporary Welsh

**Cardiff University, 2022**

**To cite this version of the *Geirfan* wordlist:**

Knight, D., Tovey-Walsh, B., Davies, E., Morris, S. and Prosser, H. (2022). The *Geirfan* wordlist: A vocabulary list for adult learners of Welsh (Version 2), Cardiff University, DOI: 10.17035/d.2022.0234583226

**Version statement:**

Version 2 of the *Geirfan* wordlists includes new Appendices, which incorporate minor corrections to errors in the original data. No changes have been made to the main document.

## Contents

1. <a href="#">Introduction</a>	2
2. <a href="#">Context</a>	3
3. <a href="#">Developing the <i>Geirfan</i> vocabulary list</a>	5
4. <a href="#">Interpreting the list</a>	9
5. <a href="#">Acknowledgements</a>	11
6. <a href="#">References</a>	12
7. <a href="#">Appendices</a>	13
<a href="#">Appendix A</a> : the most frequent 750 words from CorCenCC	13
<a href="#">Appendix B</a> : the basic 500-word list, without additions	48
<a href="#">Appendix C</a> : the working list of additions, as an alphabetical list	62

### 1. Introduction

This following document presents the *Geirfan* wordlist: a curated list of 500 of the most frequent words in the Welsh language, designed for use by learners at A1/A2<sup>1</sup> levels of proficiency (Council of Europe, 2021). This vocabulary list was developed using an innovative symbiosis of corpus-based methods (using data from the CorCenCC corpus) and expert-led introspection and reflection; an approach which can be replicated and adapted for use in any other language context. This document gives details of the approach used to compile this vocabulary list. The list itself is provided in the Appendices.

### 2. Context

In November 2020, CorCenCC (*Corpws Cenedlaethol Cymraeg Cyfoes* - The National Corpus of Contemporary Welsh) was launched (see Knight et al., 2020 and [www.corcenc.org](http://www.corcenc.org)). CorCenCC is the first large-scale corpus of Welsh designed to capture language use across communication types (spoken, written and e-language), genres, language varieties (regional

---

<sup>1</sup> A2 refers to the Common European Framework of Reference for Languages (CEFR) basic user, waystage level. All references to levels in this paper are as defined by CEFR and in the context of Welsh. See [https://www.wjec.co.uk/qualifications/welsh-for-adults-qualification-suite/#tab\\_overview](https://www.wjec.co.uk/qualifications/welsh-for-adults-qualification-suite/#tab_overview) [Accessed 26.08.22]

and social) and contexts, with its contributors representative of the 538,300 Welsh speakers in the UK. CorCenCC extends to over 11-million-words of contemporary Welsh language usage.

In addition to the release of the CorCenCC v1.0.0 dataset, another key output of the CorCenCC project was the production of *Yr Amliadur*, word frequency lists informing us which words and lemmas are most often used in the Welsh language. *Yr Amliadur* frequency lists (available [here](#)) provide detail on:

- the top 1,000 lemmas in CorCenCC, sorted both by rank and alphabetically by lemma,
- the top 1,000-word forms in CorCenCC, sorted both by rank and alphabetically by word,
- banded wordlists of the 5,000 of the most common open-class words in CorCenCC, split into lists of bands containing 500 words each (i.e. including lists for the top 500 nouns, verbs, adjectives),
- the top 50 adverbs and interjections in CorCenCC, and,
- the top 100 open-class words in CorCenCC's written, spoken and e-language subcorpora.

As seen above, *Yr Amliadur* includes information about the grammatical class of each item in the list, alongside a list of the most frequent lemmas that feature in the CorCenCC corpus. The grammatical information was ascribed using CorCenCC's rule-based POS tagging and tokeniser software (i.e. a tool that breaks down a text into the smallest units of a language), *CyTag* (pronounced 'kuh-tag', see Neale et al., 2018), which was built specifically to support the tagging of Welsh language data. Early versions of *CyTag* worked primarily by using the information found in *Eurfa*, the largest open-source, freely available dictionary for Welsh (Donnelly, 2013), to produce a list of possible tags for each word in a Welsh text. This dictionary was supplemented by specific lists of place names, first names and surnames extracted from Wikipedia in *CyTag*, to provide reference materials for grammatically classifying Welsh-language tokens. *CyTag*'s tagset includes major syntactic categories (e.g. noun, article, preposition, verb) as well as one category representing unique particles to Welsh and one representing other forms such as abbreviations, acronyms, symbols, digits etc. The full set of tags covers Welsh morphology based on gender (masculine or feminine), number (singular or plural), person (first, second, third) and tense (past, present, future, etc.).

*Yr Amliadur* was created automatically, using frequency-list building software built by the project team. While corpus-based frequency lists are useful in giving us evidence-based

lists of those lexical items most often occurring in a language, it is nonetheless inevitable that these lists will require further adaptation before being utilised by the end-user. For *Yr Amliadur* to be a more effective and targeted resource, it was acknowledged that its usability would need further enhancing to bridge the gap between lists generated purely by means of a corpus and lists that include input from end-users themselves. This is what the *Geirfan* wordlist aims to do.

### 3. Developing the *Geirfan* wordlist

To achieve our aims, an innovative user-driven iterative approach was utilised in the development of the *Geirfan* wordlist, which focused on four main stages of development:

- enhancing *CyTag* and the *Geirfan* wordlist,
- expert review and surveying end-users,
- finalising the wordlist, and
- developing targeted reference materials for pedagogic purposes.

#### *Enhancing CyTag and the Geirfan wordlist*

The original version of *CyTag* had a high level of tagging precision, with recall scores noted at well over 95% in the initial releases of the tagger (Neale et al., 2018). As with any other tagger, however, some items were more difficult for the tagger to disambiguate than others, leading to the potential for mis-tagging of data. This limitation was noted in the release of the frequency-based *Yr Amliadur* wordlists, with the caveat that ‘this should be borne in mind when interpreting frequencies of words or lemmas which are susceptible to mis-tagging’ (Knight et al., 2020: 20). An acknowledgement of this limitation motivated the team to enhance the tagger’s accuracy in tagging high-frequency words in the first instance (as these were of primary concern to the external project partners), as well as making *CyTag*’s code easier to debug and maintain.

To enhance *CyTag*’s accuracy it was, first, necessary to define what was meant by ‘high frequency words’. Extensive manual inspections of *Yr Amliadur* (by members of the project team), indicated that some of the highest-frequency lemmas were actually fossilized forms preserved only in multi-word units (MWUs), such as the noun *cyfer* [direction] which is obsolete except in the compound preposition *ar gyfer* [in respect of]. In order to provide more accurate POS tagging, and thereby generate more accurate frequency lists, *CyTag*’s original

code needed to be adapted to better support these MWUs. By generating new lists as stages of the code re-write were completed, we were able to assess the impact of including the MWUs, and to identify other high-frequency lemmas that appeared to be out of place/suspicious to our team of Welsh language experts. The later steps of the project's first stage involved making fixes to the code and subsequently checking the top 1,000 lemmas.

Our team of Welsh language experts manually reviewed frequency lists at numerous stages as *CyTag* was being improved, and their feedback was used to isolate words which were being mis-tagged (e.g., the noun *rhyw* [sex] was over-represented, because the tagger was inaccurately tagging instances of the determiner *rhyw* [some] as the noun). Subsequent changes to the tagger's lexicon and grammar were made because of this feedback and to produce more accurate results for such lemmas. The lists of high-frequency words in the final versions of the *Geirfan* vocabulary list provide higher accuracy of individual homographic forms, and also contain high-frequency multi-word units instead of the individual fossilized words that make up those units.

The following changes to the *CyTag* code were required to produce more accurate and useful frequency lists. The revised tagger with additional lexicon terms, *CyTag2* (Tovey-Walsh et al., 2022), is available at <https://github.com/CorCenCC/CyTag2>:

- *Rewriting the code base*: Although it produced reliable, high accuracy tagged output, the original *CyTag* code was arguably difficult to read and maintain. This has been mitigated by reproducing *CyTag*'s functionality, but in an object-oriented (OOP, rather than imperative) style using Python. OOP is a programming style which focuses on the creation of 'objects' with interrelated variables and functions, which include both data and code.
- *Extending the lexicon*: As *CyTag* relies on a lexicon to identify the possibilities for tagging a word, it is important that the lexicon has very broad coverage. At the beginning of the project, the project team identified that the *CyTag* lexicon was missing words in a number of key areas which were added to the lexicon, including: some irregular verb paradigms; terminology for more recent technology and science (e.g. the verbs *e-bostio* [to email] and *ail-drydar* [to retweet]); variant spellings of words with prefixes (e.g. the verb meaning 'to cooperate', which can be written *cyd-weithredu*, *cydweithredu*, or (exceptionally) *cyd weithredu*); numerals (both ordinal and cardinal); some common colloquialisms/regionalisms.

- *Improving the grammar*: Improving the grammar’s ability to resolve ambiguity and recognise words without readings from the lexicon was an important step to take, to increase the tagger accuracy. In addition, a set of rules that merge words which have been separated during tokenization because they contain apostrophes was introduced. Further was the introduction of a grammar rule which replaces all remaining unknown parts of speech with the tag ‘E p’ (= proper noun) if the word is capitalized, or ‘E gb u’ (= masculine/feminine singular noun) if not capitalized. Finally, a rule that removes all English readings where a Welsh reading is possible was added.
- *Adding additional support for multi-word units (MWUs)*: The original version of *CyTag* was not able to apply contextual rules effectively when MWUs were involved. The solution to this problem was to add MWUs (mostly prepositions and adverbs, with some pronouns and conjunctions) to the *CyTag* lexicon, and to modify the tagger code to take account of these additions and treat them as single units. MWUs were first identified in Welsh language dictionaries (*Geiriadur Prifysgol Cymru* [Dictionary of the Welsh Language] and in King, 2007) and added to the lexicon. The code was then reworked to better identify the MWUs and rules were added to the grammar to identify these units. Further development might therefore consider adding more of these ambiguous MWUs using grammar rules.

### Expert review and surveying end-users

The expert review and end-user consultation stage involved not only those members of the project team whose expertise was mainly in Welsh pedagogy and the wider field of *Learning Welsh* in the over 18 (i.e. adult) sector, but also experienced practitioners in the area who were not part of the core team. The most frequent 500 open class lemmas created from the CorCenCC corpus were initially checked with the most frequent items occurring in the Learning Welsh curriculum currently in use (2020a, 2020b, 2021). Through this initial process, experts<sup>2</sup> on the team were able to preclude any entries which would not typically be expected to be found in A2 course materials.

The next step was to distribute the amended list to a group of *Learning Welsh* experts with experience in curriculum and assessment to ensure (i) that items included reflect what experience would suggest should be in such a list and (ii) the list included items which an A2

---

<sup>2</sup> These consisted of project members leading in Welsh Language curriculum, assessment and pedagogy.

learner might be expected to recognise but not necessarily produce e.g. meta language. A revised iteration of the wordlist following this stage was then circulated to a wider group of practitioners in the field, who were all very experienced tutors working across all parts of Wales. Here the participants were asked to rate all the lexical items in the updated list according to where on the CEFR description spectrum they would place them as well as noting any entries that might require passive knowledge only at A2 level.

### Finalising the wordlist

To produce a finalised *Geirfan* wordlist, the output resulting from the stage outlined above was considered for one final time by the Welsh language experts within the project team. Agreed rationales were put in place to justify the removal (or addition) of items from the most recent iteration of the wordlist (see [Appendix A](#)). These include:

- Items which were mainly functional in nature e.g. the negative denominator *ddim*
- Items adjudged to be too complex, difficult or abstract for A2 e.g. *comisiynydd* (commissioner) *cyflwyno* (to present)
- Items recorded with more than one possible part of speech e.g. *achos* to be retained as conjunction (because) but not as a noun (case)
- Items which were only components of collocations and unlikely to be encountered at A2 level otherwise e.g. *ambell* in *ambell waith* (occasionally).
- Cross referenced dialect forms e.g. *allan / mas* (out).

Whenever an item was removed, the next one in frequency position 501 was automatically added, and so on, until a final wordlist of 500 items was agreed.

### Developing targeted reference materials for pedagogic purposes

To demonstrate the potential of the *Geirfan* wordlist as a basis for materials development, dictionary entries were created, and hosted on the *Geirfan* website (Tovey-Walsh, 2022). *Geirfan* (the name is a portmanteau of “gair” (*word*) and “man” (*site* or *place*)) provides users with dictionary entries designed to be accessible to learners at A1/A2 level, but also to be comprehensive enough for advanced learners. At this stage, around ten percent of the enhanced wordlist items (see [Appendix C](#)) have been completed as entries, and further development is planned after seeking feedback from early users.



## 4. Interpreting the list

[Appendix A](#) contains the most frequent 750 words from CorCenCC, the result of tagging the corpus with *CyTag2*. This was the list that was curated by the project team, resulting in the basic 500-word wordlist ([Appendix B](#)).

[Appendix A](#) provides the following information:

- Rank: the item's overall rank in the CorCenCC corpus, according to raw frequency.
- Lemma and part of speech: in this list, homographs have not yet been united, so (for example) there are entries for both “dechrau (verb)” and “dechrau (noun)”.
- Rank in each of the CorCenCC subcorpora: the item's rank, according to frequency, in the written, spoken, and electronic subcorpora of CorCenCC. This information shows that some items are heavily weighted to a particular type of language. For example, the word “ie” (*yes*) is the tenth most frequent word in the spoken data, but 640<sup>th</sup> and 632<sup>nd</sup> most frequent in the written and electronic data, respectively.
- *Geirfa Graidd?*: indicates whether or not this word is included in the *Geirfa Graidd* vocabulary list (Morris and Meara, 2019). This is the principal list used for all A1 and A2 assessment and therefore an important point of cross-reference and comparison.
- *Omit?*: indicates whether the team ultimately decided to omit this item from the wordlist. A brief explanation is given for each omitted item. Items highlighted in green are homographs; these will be subsumed under a single entry, as the basic wordlist does not list parts of speech. Items highlighted in yellow are those which will be included in the *Geirfan* (online) dictionary entries, usually as derivatives, but will not appear as entries in the *Geirfan* word list. Items highlighted in blue are those which were omitted from the list entirely. The explanation for omission may be:
  - “advanced” (the word is frequent, but too advanced for A1/A2 learners);
  - “spoken discourse marker”: these were generally omitted, as they are often non-standard words or words with no standard written form, like “ymm” (*mmm*). The data about relative rank in the spoken subcorpus, as compared to the other subcorpora, was one of the factors used when choosing to omit a word on this basis.
  - “non-standard”: these are items which are on the borderline between English and Welsh, like “jyst” (a loanword from English *just*) and are not always accepted as standard Welsh words.

- “tagging error”: the frequency of these items was determined to be artificially inflated because of an erroneous tag attribution by *CyTag2*. An example is the word “mân” (*fine* or *small*), which is often an error for the word “man” (*place* or *site*).
- “overrepresented”: these items were determined to be overrepresented in the data. Often, these items are nouns which are used as quasi-proper nouns in the names of organizations and institutions: “undeb” (*union*) and “mudiad” (*organization*) are two of the items omitted on this basis.
- “specialized”: these items belong to a specialized field or genre, and were determined not to be particularly useful to learners, in comparison to items with a wider application. An example is the word “archif” (*archive*); it is not really too advanced for A1/A2 learners, and its form is similar enough to its English cognate that it would probably be easy to learn. However, it has no great value to most learners in performing common communicative tasks.
- Additions?: this column lists items added to the enhanced vocabulary list ([Appendix C](#)). These extra words were determined to be necessary additions to complement the item to which they are attached. For example, since Christian places of worship (“capel” (*chapel*) and “eglwys” (*church*)) were on the list, we also added “mosg” (*mosque*), “synagog” (*synagogue*) and “templ” (*temple*). The list included some, but not all, basic colour terms, so words such as “oren” (*orange*) and “porffor” (*purple*) were added.
- Tutors’ survey average: this column summarizes the result of an informal survey of 27 tutors of Welsh for adults, who were asked to rate words according to their suitability for learners at level A1, A2, B1, and B2 or higher. The average response is given here: a response of 1 would indicate that respondents largely agreed that the word was suitable at level A1. A response of 2 indicates level A2, 3 indicates B1, and 4 indicates B2 or above. The score does not reflect how unanimous responses were; for example, if all respondents agree that a word is level A2, the average score will be 2; if half of respondents agree on level A1 and the other half agree on level B1, the average score will still be 2. Nonetheless, the score gives a rough idea of tutors’ views on how advanced these items are. This score was useful in determining whether a word fitted into the “advanced” category when deciding on omissions.
- Rank similarity: this is an approximate indication of how consistent the item’s rank is across the three subcorpora, and how well its overall rank reflects its frequency in all

three subcorpora. A high rank similarity means that the item's rank in the three subcorpora is near to consistent, and that its overall rank is a good reflection of its rank in the three subcorpora. A low rank similarity means that the item has very different ranks in the three subcorpora, and that its overall rank is not a good reflection of its actual frequency in each of the subcorpora. This number must be understood strictly as a relative score; it is only useful as an indication that some items have a more balanced representation across CorCenCC's subcorpora than others. The item "bod (verb)", for example, has an overall rank of 2, and ranks of 2, 1, and 2 in the written, spoken, and electronic subcorpora, respectively. It has a high rank similarity of 100.33, reflecting the fact that its overall rank is a good reflection of its frequency in the subcorpora: it is highly frequent in every subcorpus. The item "ie (adv.)", on the other hand, has an overall rank of 31, and subcorpus ranks of 640, 10, and 632. The overall rank is not a good reflection of its actual frequency in any of the subcorpora: in spoken Welsh, it is much more frequent than the rank 31 suggests; but in written and electronic Welsh, it is much less frequent. Rank similarity is therefore a shorthand for how evenly-distributed a word is. In many cases, words which show a very uneven distribution across subcorpora are also those which the team omitted as too advanced or specialized, or as spoken discourse markers.

[Appendix B](#) contains the basic 500-word list, without additions. These 500 words are those which are directly drawn from CorCenCC's frequency data.

[Appendix C](#) contains the working list of additions, as an alphabetical list. The 500-word basic list plus these additions is the initial batch of headwords for the dictionary on the *Geirfan* website.

## 5. Acknowledgements

The research on which the *Geirfan* vocabulary list was based was funded by the UK Economic and Social Research Council (ESRC) Impact Acceleration Account (IAA). This work also feeds more broadly into the work carried out as part of the ESRC and Arts and Humanities Research Council (AHRC) funded *Corpws Cenedlaethol Cymraeg Cyfoes* (The National Corpus of Contemporary Welsh): A community driven approach to linguistic corpus construction project (Grant Number ES/M011348/1).

## 6. References

- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment* [online]. Retrieved from: <https://rm.coe.int/1680459f97> [Accessed 04.10.22]
- Donnelly, K. (2013). *Eurfa v3.0 - Free (GPL) Dictionary (incorporating Konjugator and Rhymer)* [Online]. Retrieved from: <http://eurfa.org.uk> [Accessed 04.10.22]
- King, G. (2007). *Modern Welsh Dictionary: A guide to the living language*. Oxford: OUP.
- Knight, D., Morris, S., Fitzpatrick, T., Rayson, P., Spasić, I., Thomas, E.-M., Scannell, K. (2020). *CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh*. Retrieved from: <http://doi.org/10.17035/d.2020.0119878310> [Accessed 04.10.22]
- Knight, D., Morris, S., Fitzpatrick, T., Rayson, P., Spasić, I., Thomas, E.-M., Lovell, A., Morris, J., Evas, J., Stonelake, M., Arman, L., Davies, J., Ezeani, I., Neale, S., Needs, J., Piao, S., Rees, M., Watkins, G., Williams, L., Muralidaran, V., Tovey-Walsh, B., Anthony, L., Cobb, T., Deuchar, M., Donnelly, K., McCarthy, M. and Scannell, K. (2020). *CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh*. Retrieved from: <http://doi.org/10.17035/d.2020.0120164107> [Accessed 04.10.22]
- Morris, S. and Meara, P. (2019). *Geirfa Graidd Lefel Mynediad* [online]. Retrieved from: <https://wjecwebsitelive.blob.core.windows.net/media/wjbi4egs/mynediad-geirfa-graidd.pdf?sv=2019-07-07&sr=b&sig=Cwrl7sSSYRqJgpE3j6kqVvKgvxAM9h6Lirv4qAx5Eos%3D&se=2022-10-07T15%3A34%3A04Z&sp=r> [Accessed 04.10.22]
- NCLW. (2021a). *Mynediad y De (Fersiwn 2)*. Canolfan Dysgu Cymraeg Genedlaethol.
- NCLW. (2021b). *Sylfaen y De*. Canolfan Dysgu Cymraeg Genedlaethol.
- NCLW. (2020). *Canolradd y De*. Canolfan Dysgu Cymraeg Genedlaethol.
- Neale, S., Donnelly, K., Watkins, G., and Knight, D. (2018). *Leveraging Lexical Resources and Constraint Grammar for Rule-Based Part-of-Speech Tagging in Welsh*. Paper presented at the Language Resources Evaluation 2018 Conference, Miyazaki, Japan.
- Tovey-Walsh, B., Neale, S., and Donnelly, K. (2022). *CyTag2*. GitHub repository [online]. Retrieved from: <https://github.com/CorCenCC/CyTag2> [Accessed 04.10.22]
- Tovey-Walsh, B. (2022). *Geirfan* [online]. Retrieved from: <https://geirfan.cymru/> [Accessed 04.10.22]

## 7. Appendices

### Appendix A: the most frequent 750 words from CorCenCC

Rank	Lemma	Part of speech	Rank (written)	Rank (spoken)	Rank (electronic)	Geirfa Graidd?	Omit?	Additions?	Tutors' survey average	Rank similarity
1	y	YFB (def. art.)	1	3	1	YES			no data	99.34
2	bod	B (v.)	2	1	2	YES			no data	100.33
3	yn	U (unique)	3	2	3	YES	homograph		no data	100.32
4	a	Cys (conj.)	4	8	4	YES			1.64	98.73
5	i	Ar (prep.)	5	6	5	YES			no data	99.68
6	o	Ar (prep.)	6	4	6	YES			no data	100.63
7	yn	Ar (prep.)	7	7	7	YES			no data	100.00
8	fi	Rha (pron.)	10	5	11	YES			no data	99.39
9	ei	Ban (det.)	8	26	9	YES			no data	95.34
10	ar	Ar (prep.)	9	19	8	YES			no data	98.21
11	am	Ar (prep.)	11	25	10	YES			no data	96.24
12	wedi	U (unique)	15	15	15	YES	advanced		no data	97.39
13	cael	B (v.)	14	23	14	YES			no data	96.58
14	ni	Rha (pron.)	26	9	23	YES			no data	95.53
15	gan	Ar (prep.)	13	40	13	NO			no data	94.26

16	gwneud	B (v.)	21	16	20	YES			no data	97.48
17	ond	Ar (prep.)	16	21	21	YES			no data	98.04
18	â	Ar (prep.)	12	38	17	NO			no data	96.46
19	hwn	Rha (pron.)	17	39	16	YES	advanced		no data	95.97
20	gyda	Ar (prep.)	19	30	18	YES			no data	98.09
21	mynd	B (v.)	20	18	27	YES			no data	99.45
22	a	Rha (pron.)	18	37	19	YES	homograph		no data	97.86
23	fo	Rha (pron.)	33	12	60	YES			no data	91.11
24	fel	Ar (prep.)	22	29	24	YES			no data	99.20
25	chi	Rha (pron.)	52	13	41	YES			no data	92.36
26	cymru	E (n.)	38	107	12	YES		lloegr, yr alban, iwerddon, gogledd iwerddon, ewrop	no data	82.71
27	nhw	Rha (pron.)	37	17	56	YES			no data	92.93
28	ti	Rha (pron.)	45	14	81	YES			no data	87.27
29	mewn	Ar (prep.)	23	46	22	YES			no data	98.98
30	ddim	Adf (adv.)	39	22	50	YES	homograph		no data	94.89
31	ie	Adf (adv.)	640	10	632	NO	advanced		1.15	24.84

32	ymm	Ebych (interj.)	287	11	197	NO	spoken discourse marker		1.27	49.81
33	dod	B (v.)	27	34	37	YES			no data	100.25
34	gallu	B (v.)	31	44	28	YES			no data	99.75
35	hwnnw	Rha (pron.)	28	33	42	YES	advanced		no data	100.50
36	yna	Adf (adv.)	53	20	84	YES			no data	89.28
37	neu	Cys (conj.)	30	51	26	YES			no data	100.98
38	fy	Ban (det.)	24	88	25	YES			no data	94.74
39	mawr	Ans (adj.)	32	52	29	YES			no data	100.97
40	gweld	B (v.)	34	42	40	YES			no data	100.96
41	hi	Rha (pron.)	29	36	83	YES			no data	94.42
42	dweud	B (v.)	50	31	54	YES			no data	97.93
43	yma	Adf (adv.)	97	24	77	YES			no data	86.14
44	beth	Rha (pron.)	79	27	80	YES			no data	88.89
45	wrth	Ar (prep.)	25	91	47	YES			no data	93.95
46	da	Ans (adj.)	47	48	38	YES			no data	101.15
47	peth	E (n.)	56	32	58	NO			1.45	98.88
48	eu	Ban (det.)	35	115	32	YES			no data	92.12
49	rhoi	B (v.)	43	55	36	NO			1.11	103.00
50	pobl	E (n.)	63	50	34	NO			1.00	100.67
51	at	Ar (prep.)	36	84	45	YES			no data	97.42
52	hefyd	Adf (adv.)	49	59	44	YES			no data	100.88

53	na	Rha (pron.)	46	57	53	NO	advanced		no data	100.66
54	os	Cys (conj.)	84	45	49	NO			no data	96.65
55	wedi	Ar (prep.)	41	113	39	YES	advanced		no data	94.32
56	pob	Ban (det.)	40	72	51	YES			no data	101.08
57	eich	Ban (det.)	95	78	30	YES			no data	93.64
58	trwy	Ar (prep.)	51	93	43	YES			no data	97.33
59	cymraeg	Ans (adj.)	72	134	31	YES	homograph		no data	88.83
60	dau/dwy	Rhi (num.)	44	60	75	NO			1.00	100.21
61	newydd	Ans (adj.)	68	118	33	NO			1.00	93.06
62	ein	Ban (det.)	70	155	35	YES			no data	86.79
63	gwaith	E (n.)	59	89	48	YES			no data	98.59
64	felly	Adf (adv.)	80	53	82	NO			1.33	95.53
65	un	Rha (pron.)	58	62	72	NO	homograph		1.00	100.61
66	iawn	Adf (adv.)	71	54	79	YES	homograph		no data	98.81
67	blwyddyn	E (n.)	62	75	55	YES			no data	101.83
68	iaith	E (n.)	48	159	52	NO			1.15	90.16
69	arall	Ban (det.)	57	70	66	YES			no data	102.84
70	y	Rha (pron.)	42	200	59	YES	homograph		no data	84.86
71	ar gyfer	Ar (prep.)	65	188	46	NO			no data	85.64
72	plant	E (n.)	60	120	64	YES			no data	94.85
73	yn ôl	Adf (adv.)	82	68	76	YES			no data	98.67
74	ysgol	E (n.)	55	125	78	YES			no data	93.55



75	cymraeg	E (n.)	75	138	57	YES		saesneg	no data	92.11
76	pan	Cys (conj.)	61	81	97	YES			no data	97.96
77	rhai	Rha (pron.)	74	90	74	YES	advanced		no data	98.70
78	defnyddio	B (v.)	77	145	61	YES		defnyddiol	no data	91.60
79	cyntaf	Rhi (num.)	67	96	86	NO			1.52	97.81
80	tros	Ar (prep.)	76	139	62	NO			no data	93.59
81	gwybod	B (v.)	106	47	192	YES			no data	84.19
82	nid	Adf (adv.)	64	202	68	NO	advanced		3.05	86.12
83	dydd	E (n.)	94	102	65	YES		dyddiad	no data	97.86
84	rhaid	E (n.)	69	82	98	NO			1.27	100.55
85	yyy	Ebych (interj.)	11832	28	10533	NO	spoken discourse marker		1.41	2.45
86	wedyn	Adf (adv.)	178	41	333	YES			no data	65.49
87	rhan	E (n.)	83	167	71	NO			1.65	90.34
88	fe	U (unique)	66	117	100	YES	advanced		no data	96.74
89	byd	E (n.)	81	105	95	NO			1.38	97.59
90	a	U (unique)	73	163	94	YES	homograph		no data	90.48
91	tŷ	E (n.)	92	144	85	YES			no data	92.27
92	efo	Ar (prep.)	154	58	153	YES			no data	86.62
93	cyn	Ar (prep.)	78	124	107	YES			no data	95.07
94	so	Adf (adv.)	1730	35	1375	NO	spoken discourse marker		2.00	16.92

95	angen	E (n.)	105	121	87	YES			no data	95.43
96	yma	Ban (det.)	190	65	115	YES			no data	87.76
97	meddwl	B (v.)	158	56	233	YES			no data	79.12
98	bach	Ans (adj.)	85	99	136	YES			1.00	95.81
99	ffordd	E (n.)	104	94	104	YES			no data	99.17
100	dim	Rha (pron.)	114	66	204	YES			no data	87.72
101	enw	E (n.)	109	151	93	YES			no data	92.34
102	dyma	Adf (adv.)	129	130	91	YES			no data	93.23
103	wel	Ebych (interj.)	419	43	689	NO	spoken discourse marker		1.05	41.97
104	amser	E (n.)	111	104	103	YES			no data	99.03
105	rhaglen	E (n.)	54	366	203	YES			no data	66.63
106	lle	Rha (pron.)	110	87	159	YES			no data	94.21
107	gweithio	B (v.)	145	92	121	YES			no data	94.38
108	mis	E (n.)	148	186	90	YES			no data	86.19
109	hen	Ans (adj.)	93	222	110	YES			no data	86.48
110	mi	U (unique)	103	108	150	YES	advanced		no data	95.31
111	rhywbeth	Rha (pron.)	169	67	215	YES			no data	84.29
112	eisiau	E (n.)	227	64	225	YES		moyn	no data	77.94
113	unrhyw	Ban (det.)	90	278	116	NO			no data	81.51
114	er	Ar (prep.)	86	2792	99	NO	advanced		no data	19.59
115	llyfr	E (n.)	229	303	67	YES			no data	71.75

116	mai	Cys (conj.)	87	312	122	YES	advanced		no data	78.93
117	heb	Ar (prep.)	100	203	117	YES			no data	90.42
118	na	Cys (conj.)	102	141	154	NO	advanced		no data	93.83
119	pawb	Rha (pron.)	108	123	165	YES			no data	94.40
120	jyst	Adf (adv.)	1007	49	735	NO	non-standard		1.35	31.56
121	nawr	Adf (adv.)	306	61	265	YES			no data	71.14
122	eto	Adf (adv.)	107	129	161	YES			no data	95.55
123	dysgu	B (v.)	99	148	168	YES			no data	93.57
124	siarad	B (v.)	144	79	218	YES			no data	90.69
125	rhwng	Ar (prep.)	89	274	142	YES			no data	83.85
126	aelod	E (n.)	152	465	89	NO			2.19	67.40
127	lle	E (n.)	122	127	162	YES	homograph		no data	95.78
128	pen	E (n.)	88	247	173	YES			no data	84.65
129	er mwyn	Ar (prep.)	91	4490	111	NO	advanced		no data	13.76
130	meddwl	E (n.)	170	101	169	YES	homograph		no data	93.24
131	gwasanaeth	E (n.)	269	573	70	NO			1.81	57.18
132	ar ôl	Adf (adv.)	98	173	180	YES			no data	92.68
133	dyna	Adf (adv.)	171	73	299	YES			no data	82.92
134	gair	E (n.)	96	165	213	YES			no data	90.70
135	llywodraeth	E (n.)	368	398	69	NO	advanced		2.48	62.11
136	cyfle	E (n.)	131	323	101	NO			2.33	82.81
137	dechrau	E (n.)	139	146	148	YES	homograph		no data	97.00
138	darllen	B (v.)	234	204	96	YES			no data	85.61

139	hanes	E (n.)	137	250	108	NO			1.85	90.19
140	gwahanol	Ans (adj.)	119	152	172	NO		gwahaniaeth	1.85	96.90
141	gwybodaeth	E (n.)	189	480	92	NO			1.93	68.14
142	dy	Ban (det.)	162	76	358	YES			no data	81.03
143	tri/tair	Rhi (num.)	135	98	241	NO			1.00	94.19
144	cofio	B (v.)	118	110	248	YES			no data	94.33
145	efallai	Adf (adv.)	263	77	216	YES			1.60	85.86
146	heddiw	Adf (adv.)	140	128	178	YES		ddoe, yfory	no data	98.93
147	un	Rhi (num.)	136	112	210	NO		naw, wyth, cant	1.00	97.76
148	codi	B (v.)	101	208	186	YES			no data	93.58
149	diwrnod	E (n.)	116	233	141	YES			no data	94.56
150	arall	Rha (pron.)	126	402	112	YES	homograph		no data	79.79
151	llun	E (n.)	180	243	109	YES			no data	90.50
152	mwy	Rha (pron.)	205	157	130	YES			no data	95.45
153	gwlad	E (n.)	134	271	126	YES			no data	91.34
154	cyfnod	E (n.)	120	356	127	NO	advanced		2.78	84.39
155	cymryd	B (v.)	161	185	149	YES			no data	96.23
156	wythnos	E (n.)	210	142	152	YES			no data	95.52
157	digwydd	B (v.)	184	106	219	YES			no data	95.30
158	bywyd	E (n.)	115	219	171	NO			2.07	96.15
159	cig	E (n.)	1347	1392	63	YES		llyisiau, ffrwythau	no data	25.05
160	dylai	B (v.)	128	254	151	NO	advanced		2.17	93.64

161	gadael	B (v.)	133	258	144	YES			no data	93.77
162	nifer	E (n.)	141	404	118	NO			2.37	81.62
163	newid	B (v.)	208	182	132	YES			no data	95.99
164	llawer	Rha (pron.)	167	251	135	YES			no data	92.85
165	tro	E (n.)	121	190	205	YES			no data	97.43
166	creu	B (v.)	165	252	139	NO	advanced		2.37	93.22
167	gwefan	E (n.)	591	668	73	NO			1.44	49.08
168	clywed	B (v.)	112	156	291	YES		blasu, aroglu	no data	93.60
169	newyddion	E (n.)	149	698	113	YES			no data	64.05
170	cadw	B (v.)	142	207	181	YES			no data	97.59
171	math	E (n.)	175	135	235	NO			2.25	96.21
172	ifanc	Ans (adj.)	151	282	158	YES			no data	91.58
173	ymlaen	Adf (adv.)	197	149	200	YES			no data	96.81
174	teulu	E (n.)	123	304	184	YES			no data	90.23
175	nesaf	Adf (adv.)	295	143	160	YES			no data	91.87
176	bach	Adf (adv.)	219	97	338	YES	homograph		no data	86.79
177	dangos	B (v.)	192	266	143	NO			1.67	92.23
178	cwmni	E (n.)	124	368	174	NO			1.63	86.34
179	yn ystod	Ar (prep.)	172	460	131	NO			no data	78.74
180	pwysig	Ans (adj.)	212	184	170	YES			no data	97.00
181	llawn	Ans (adj.)	132	377	175	NO			1.44	85.67
182	stori	E (n.)	156	275	182	YES			no data	92.66
183	arbennig	Ans (adj.)	130	330	193	NO			1.56	89.09
184	cynnwys	E (n.)	216	712	114	NO	homograph		2.33	63.49
185	arian	E (n.)	359	310	105	YES			no data	79.61

186	cynnal	B (v.)	174	533	137	NO	advanced		2.48	75.00
187	rhywun	Rha (pron.)	214	114	316	YES			no data	91.21
188	mor	Adf (adv.)	147	175	277	NO			2.06	96.11
189	na	U (unique)	454	69	571	NO	advanced		no data	62.20
190	edrych	E (n.)	188	160	255	YES	tagging error		no data	96.35
191	dyn	E (n.)	113	226	311	YES		dynes	no data	91.89
192	prif	Ans (adj.)	211	395	134	NO			1.74	84.23
193	pryd	Rha (pron.)	173	180	257	YES			no data	96.59
194	rhyw	Ban (det.)	221	119	329	NO			2.24	91.02
195	pam	Adf (adv.)	252	137	242	YES			no data	95.06
196	mwy	Adf (adv.)	206	220	185	YES	homograph		no data	97.48
197	cynllun	E (n.)	328	509	106	NO	advanced		2.37	71.68
198	cynulliad	E (n.)	1188	779	88	NO	overrepresented		2.91	37.96
199	allan	Adf (adv.)	220	131	300	YES			no data	94.32
200	credu	B (v.)	271	109	319	YES			no data	90.09
201	dilyn	B (v.)	168	430	176	NO			2.15	84.08
202	ysgrifennu	B (v.)	281	235	156	YES			no data	93.21
203	cenetlaethol	Ans (adj.)	249	539	125	NO			2.37	74.94
204	derbyn	B (v.)	181	466	163	NO			1.96	82.16
205	lleol	Ans (adj.)	311	408	120	NO			1.48	80.33
206	dim	Ban (det.)	224	136	312	YES	homograph		no data	94.44
207	siŵr	Ans (adj.)	327	103	305	YES			no data	88.99
208	erbyn	Ar (prep.)	166	261	243	YES			no data	95.26

209	chwarae	B (v.)	150	196	336	YES			no data	94.40
210	mam	E (n.)	117	162	596	YES			no data	79.15
211	teimlo	B (v.)	146	191	351	YES			no data	94.43
212	ers	Ar (prep.)	191	228	227	YES			no data	98.94
213	car	E (n.)	127	262	315	YES			no data	93.53
214	syniad	E (n.)	242	183	226	YES			no data	99.05
215	oherwydd	Cys (conj.)	183	241	228	NO	advanced		no data	99.26
216	dim ond	Adf (adv.)	233	164	260	YES			no data	99.06
217	taith	E (n.)	203	598	155	YES			no data	75.72
218	sef	Cys (conj.)	164	353	232	NO	advanced		no data	90.94
219	cyngor	E (n.)	344	435	129	NO			1.63	79.22
220	merch	E (n.)	155	234	328	YES		menyw, mab	no data	94.40
221	yndê	Ebych (interj.)	1866	63	1641	NO	spoken discourse marker		3.01	24.88
222	byw	B (v.)	186	285	231	YES			no data	96.41
223	tynnu	B (v.)	202	218	258	NO			1.63	99.08
224	iawn	Ans (adj.)	335	95	522	YES			no data	77.64
225	iawn	Ebych (interj.)	528	85	463	YES	homograph		no data	70.86
226	cwestiwn	E (n.)	217	240	237	YES			no data	98.39
227	diwedd	E (n.)	177	236	303	NO			1.30	96.56
228	cyhoeddi	B (v.)	283	695	140	NO	advanced		2.63	69.39
229	cyrraedd	B (v.)	143	427	267	YES			no data	86.81
230	ceisio	B (v.)	138	451	271	NO			2.30	85.34

231	chwilio	B (v.)	394	421	133	NO			1.67	79.57
232	ail	Rhi (num.)	153	248	373	NO			1.67	92.74
233	troi	B (v.)	160	245	364	YES			no data	93.45
234	cysylltu	B (v.)	543	494	123	NO			2.30	68.63
235	bwyd	E (n.)	370	316	164	YES			no data	87.39
236	cyflwyno	B (v.)	176	493	239	NO	advanced		2.41	83.44
237	wrth gwrs	Adf (adv.)	309	147	360	YES			no data	90.59
238	helpu	B (v.)	253	255	240	YES			no data	96.76
239	polisi	E (n.)	433	786	128	NO	advanced		2.52	61.75
240	mudiad	E (n.)	964	1006	102	NO	overrepresented		2.93	43.00
241	sicrhau	B (v.)	218	484	202	NO	advanced		3.15	84.97
242	busnes	E (n.)	619	562	124	NO			1.59	63.93
243	cynnig	E (n.)	243	524	189	NO			2.33	81.93
244	ardal	E (n.)	254	386	201	YES			no data	90.45
245	eglwys	E (n.)	199	535	220	YES			no data	82.54
246	rŵan	Adf (adv.)	865	80	695	YES			no data	53.51
247	maes	E (n.)	182	392	280	YES			no data	90.21
248	diolch	E (n.)	457	122	393	YES	homograph		no data	82.08
249	un	Ans (adj.)	222	300	275	NO	homograph		1.00	95.44
250	symud	B (v.)	225	192	404	YES			no data	93.67
251	dechrau	B (v.)	201	210	406	YES			no data	94.27
252	pwy	Rha (pron.)	238	153	546	YES			no data	85.37
253	ffrind	E (n.)	125	413	464	YES			no data	81.34
254	sylw	E (n.)	259	584	194	NO	advanced		2.67	79.43



255	tad	E (n.)	157	199	700	YES			no data	78.54
256	hoffi	B (v.)	430	269	196	YES		lico, casáu	no data	89.37
257	ei hun	Rha (pron.)	163	496	332	NO	advanced		no data	82.96
258	cyfrwng	E (n.)	228	679	209	NO	advanced		3.44	75.85
259	diweddar	Ans (adj.)	275	557	190	NO			2.67	81.47
260	môr	E (n.)	179	292	395	YES			no data	92.62
261	ar gael	Adf (adv.)	494	827	138	YES			no data	61.57
262	cartref	E (n.)	241	477	223	YES		adref, gartref	no data	87.51
263	cwrs	E (n.)	323	302	222	YES			no data	94.94
264	gofyn	E (n.)	255	209	359	YES	homograph		no data	97.24
265	bodd	E (n.)	200	523	254	NO	include as "wrth fy modd"		2.32	85.75
266	cymdeithas	E (n.)	286	444	198	NO			2.33	89.41
267	dewis	E (n.)	290	412	207	NO			1.46	91.07
268	amlwg	Ans (adj.)	226	334	287	NO	advanced		2.81	96.25
269	addysg	E (n.)	266	572	206	NO			2.19	82.37
270	datblygu	B (v.)	282	464	208	NO			2.74	88.52
271	enghraifft	E (n.)	268	411	229	NO	advanced		1.85	92.14
272	aros	B (v.)	231	232	367	YES			no data	98.76
273	cyfarfod	E (n.)	296	362	221	YES			no data	94.91
274	cyfres	E (n.)	159	763	324	NO	overrepresented		2.33	72.57
275	hawl	E (n.)	196	821	251	NO	advanced		3.07	71.75
276	oed	E (n.)	260	385	249	YES			no data	94.47
277	tîm	E (n.)	207	283	368	YES			no data	97.67

278	te	Ebych (interj.)	985	74	1588	YES	spoken discourse marker		no data	38.48
279	uchel	Ans (adj.)	230	358	286	NO			1.89	96.85
280	nos	E (n.)	235	253	363	YES		noswaith	no data	99.04
281	na	Ar (prep.)	272	227	348	NO	advanced		no data	99.65
282	edrych	B (v.)	251	168	521	YES			no data	92.42
283	data	E (n.)	273	1326	187	NO	advanced		2.23	55.08
284	reit	Adf (adv.)	1645	71	2063	NO			1.34	28.24
285	achos	Cys (conj.)	840	83	981	YES			no data	52.40
286	eisteddfod	E (n.)	215	383	304	YES			no data	96.34
287	rhannu	B (v.)	414	423	188	NO			2.00	87.62
288	saesneg	Ans (adj.)	185	361	391	NO	homograph		1.00	94.10
289	hanner	E (n.)	274	194	432	YES			no data	97.25
290	deall	B (v.)	246	211	427	YES			no data	98.82
291	tebyg	Ans (adj.)	240	367	289	NO			2.30	98.08
292	iechyd	E (n.)	670	614	145	YES		iach, sâl, salwch, gwael, tost, anabl, anabledd	no data	68.02
293	digon	Rha (pron.)	320	263	284	YES			no data	101.03
294	cofnod	E (n.)	1185	1344	119	NO	advanced		3.30	40.09
295	cân	E (n.)	193	314	420	NO			1.33	96.58
296	grŵp	E (n.)	284	359	261	YES			no data	98.67
297	ystyried	B (v.)	213	566	288	NO	advanced		3.19	87.13

298	achos	Ar (prep.)	570	100	831	YES	homograph		no data	66.30
299	dal	B (v.)	223	287	397	YES			no data	99.17
300	sôn	B (v.)	318	172	455	NO			2.63	96.39
301	byw	Ans (adj.)	256	215	441	YES	homograph		no data	99.26
302	pedwar/pedair	Rhi (num.)	304	181	480	NO			1.00	95.34
303	prosiect	E (n.)	427	912	179	NO	advanced		1.96	66.50
304	sir	E (n.)	438	471	199	NO			1.56	86.08
305	digwyddiad	E (n.)	470	713	177	NO	include under "digwydd"		2.37	73.19
306	gyrru	B (v.)	248	315	375	YES			no data	98.38
307	golygu	B (v.)	279	420	276	NO	advanced		3.07	95.76
308	problem	E (n.)	412	217	330	YES			no data	97.22
309	profiad	E (n.)	258	487	290	NO	advanced		2.56	91.91
310	bore	E (n.)	278	178	598	YES			no data	90.84
311	cefnogi	B (v.)	587	772	166	NO			2.37	67.56
312	gwir	Ans (adj.)	292	309	335	YES	homograph		no data	100.00
313	peidio	B (v.)	321	351	281	NO			1.77	98.88
314	hir	Ans (adj.)	245	338	379	YES			no data	98.42
315	gogledd	E (n.)	443	468	212	YES		dwyrain, gorllewin, de	no data	87.49
316	ymchwil	E (n.)	261	848	264	NO	advanced		3.26	74.60
317	gofyn	B (v.)	195	265	616	YES		holi	no data	90.92
318	canol	E (n.)	276	327	352	YES			no data	99.92
319	colli	B (v.)	257	291	412	YES			no data	99.76
320	pell	Ans (adj.)	280	360	322	YES	homograph		no data	99.84

321	deg	Rhi (num.)	515	206	323	NO			1.00	93.97
322	i gyd	Rha (pron.)	356	179	504	YES			no data	94.55
323	ateb	E (n.)	299	371	308	YES			no data	99.30
324	o ran	Ar (prep.)	341	249	390	NO	advanced		no data	99.38
325	hwnnw	Ban (det.)	1202	86	1530	YES	advanced		no data	40.89
326	ymateb	E (n.)	401	519	238	NO	include as derivative under "ateb"		2.81	87.65
327	anodd	Ans (adj.)	347	256	387	YES			no data	99.30
328	medru	B (v.)	293	214	534	YES			no data	95.75
329	cymorth	E (n.)	479	1273	191	NO	advanced		2.41	57.38
330	cymuned	E (n.)	360	452	266	NO	advanced		2.22	93.61
331	rhifyn	E (n.)	962	3154	147	NO	include under "rhif"		3.04	28.34
332	yn ogystal	Adf (adv.)	267	1203	273	NO	advanced		3.28	63.44
333	mwynhau	B (v.)	331	405	298	YES			no data	97.38
334	trafod	E (n.)	367	347	301	NO			2.30	99.01
335	prifysgol	E (n.)	369	439	270	NO			1.56	94.70
336	clwb	E (n.)	450	497	244	YES			no data	87.73
337	tua	Ar (prep.)	362	344	327	NO			no data	98.35
338	plaid	E (n.)	507	671	211	NO	overrepresented		2.64	77.80
339	sawl	Ban (det.)	325	426	314	YES			no data	96.48
340	nodi	B (v.)	209	1011	385	NO	advanced		2.76	69.29
341	de	Ans (adj.)	864	116	807	YES			no data	63.39

342	o fewn	Adf (adv.)	291	485	337	NO	include under "mewn"		2.88	93.84
343	awr	E (n.)	350	267	414	YES			no data	99.85
344	oes	E (n.)	319	341	378	NO			1.40	99.55
345	dosbarth	E (n.)	194	340	692	YES			no data	87.48
346	de	E (n.)	382	242	431	YES		chwith	no data	98.75
347	popeth	Rha (pron.)	373	237	442	YES			no data	99.19
348	gofal	E (n.)	728	1202	183	NO	advanced		2.33	55.70
349	trio	B (v.)	577	154	650	NO			1.16	80.13
350	swydd	E (n.)	432	475	274	NO			1.24	91.15
351	dan	Ar (prep.)	244	536	415	YES			no data	90.50
352	mater	E (n.)	469	723	236	NO	advanced		2.47	78.47
353	cyffredinol	Ans (adj.)	405	670	263	NO	advanced		2.76	82.97
354	tal	B (v.)	386	279	398	YES			no data	99.93
355	canolfan	E (n.)	521	715	230	YES			no data	77.29
356	diolch	Ebych (interj.)	918	171	419	YES			no data	75.66
357	gwyb	Ans (adj.)	621	333	272	NO			1.36	89.84
358	ochr	E (n.)	298	284	509	NO			1.68	98.78
359	parhau	B (v.)	462	841	234	NO	advanced		2.60	74.96
360	oen	E (n.)	2397	1791	146	NO	include only as compound with "cig"		1.88	29.78
361	rhestr	E (n.)	838	775	195	NO			1.56	65.61
362	pa	Ban (det.)	399	216	527	YES			no data	96.12

363	cerdded	B (v.)	315	410	388	YES			no data	98.30
364	dyfodol	E (n.)	475	600	259	YES			no data	85.19
365	llaw	E (n.)	204	406	639	YES		bys, bawd	no data	90.06
366	prynu	B (v.)	383	349	369	YES			no data	99.79
367	barn	E (n.)	391	542	295	NO			2.20	91.69
368	olaf	Ans (adj.)	250	399	526	NO			1.80	95.19
369	adroddiad	E (n.)	701	751	214	NO	specialized		2.76	71.57
370	hyd yn oed	Adf (adv.)	363	390	377	NO			2.56	98.60
371	noson	E (n.)	277	443	443	YES			no data	96.58
372	munud	E (n.)	488	224	471	YES			no data	95.48
373	torri	B (v.)	312	346	465	YES			no data	99.72
374	rhyfel	E (n.)	346	669	306	NO	advanced		2.60	87.72
375	arwain	B (v.)	316	590	365	NO	advanced		2.64	90.71
376	sul	E (n.)	388	522	320	YES		gwener, mawrth, mercher	no data	93.33
377	cymreig	Ans (adj.)	473	982	253	NO		seisnig, celtaidd, ewropeaidd	2.72	71.26
378	rhyw	E (n.)	410	193	722	NO	homograph		2.24	88.25
379	yn unig	Adf (adv.)	307	743	361	NO			2.12	83.99
380	bwrdd	E (n.)	500	571	269	YES			no data	87.80
381	canu	E (n.)	247	305	756	YES			no data	89.74
382	tipyn	E (n.)	387	221	591	YES			no data	96.46
383	trefnu	B (v.)	365	526	344	NO		trefn	1.76	94.40
384	gwyn	Ans (adj.)	187	556	761	YES			no data	80.49

385	ble	Rha (pron.)	442	223	543	YES			no data	96.49
386	cefn	E (n.)	237	416	590	YES			no data	94.49
387	siop	E (n.)	490	448	302	YES		siopa	no data	94.87
388	gweinidog	E (n.)	531	1007	247	NO		offeiriad, imam, rabi	2.48	70.22
389	pump	Rhi (num.)	561	176	656	NO			1.00	86.65
390	blog	E (n.)	3917	2402	157	NO	overrepresented		1.57	21.69
391	diwethaf	Ans (adj.)	514	335	366	YES			no data	97.23
392	papur	E (n.)	332	458	407	YES			1.08	98.60
393	neb	Rha (pron.)	285	277	679	YES			no data	95.98
394	rhy	Adf (adv.)	333	301	542	YES			no data	100.41
395	fan hyn	Adf (adv.)	1416	132	752	NO	include under "man"		2.30	57.12
396	yn erbyn	Adf (adv.)	308	463	444	YES			no data	98.22
397	cymaint	Rha (pron.)	403	321	445	NO			no data	101.50
398	drws	E (n.)	239	372	688	YES			no data	93.43
399	digon	Ban (det.)	343	400	433	YES	homograph		no data	101.42
400	gobeithio	B (v.)	552	272	411	NO		gobaith	1.40	97.72
401	penderfynu	B (v.)	265	502	520	NO			1.88	94.71
402	galw	E (n.)	402	544	347	YES	homograph		no data	94.54
403	person	E (n.)	330	308	561	NO			1.16	100.67
404	tywydd	E (n.)	198	632	691	YES			no data	83.03

405	chdi	Rha (pron.)	1643	111	1135	NO	include under "ti"		no data	47.51
406	cymdeithasol	Ans (adj.)	527	663	279	NO	advanced		2.52	85.81
407	sefydlu	B (v.)	496	790	282	NO	advanced		3.24	81.42
408	canrif	E (n.)	352	601	383	NO	advanced		2.60	93.15
409	meithrin	Ans (adj.)	2609	4077	167	NO	overrepresented		1.98	21.35
410	tir	E (n.)	392	645	350	NO			2.40	90.69
411	ymuno	B (v.)	288	792	421	NO			2.28	85.12
412	nadolig	E (n.)	474	610	318	YES			no data	90.25
413	rheswm	E (n.)	404	467	394	NO			2.16	98.34
414	neges	E (n.)	554	422	340	YES			no data	95.42
415	dŵr	E (n.)	400	337	488	YES			no data	101.31
416	gosod	B (v.)	324	626	422	NO			2.75	92.58
417	achos	E (n.)	466	313	454	YES	homograph		no data	101.17
418	pwyllgor	E (n.)	773	866	246	NO	advanced		2.84	71.12
419	cyfeiriad	E (n.)	445	817	317	YES			no data	82.86
420	agos	Ans (adj.)	351	370	508	YES			no data	102.03
421	erioed	Adf (adv.)	322	438	500	YES			no data	100.19
422	o dan	Ar (prep.)	305	483	518	NO	include under "dan"		no data	97.51
423	ymweld	B (v.)	358	1581	342	NO	include as derivative under "gweld"		1.80	60.79
424	adnabod	B (v.)	372	281	602	YES			no data	101.09
425	ychedig	Rha (pron.)	384	540	399	NO			no data	97.04
426	unwaith	Adf (adv.)	349	354	537	YES			no data	102.47



427	diddordeb	E (n.)	522	604	326	YES			no data	90.24
428	hawdd	Ans (adj.)	503	382	405	YES			no data	99.62
429	safon	E (n.)	764	811	262	NO	advanced		2.56	74.26
430	swyddog	E (n.)	558	1019	283	NO			2.00	73.61
431	cynhyrchu	B (v.)	297	1060	446	NO	advanced		3.17	75.75
432	golwg	E (n.)	314	684	469	NO			2.72	90.32
433	cerddoriaeth	E (n.)	492	376	430	NO			1.44	100.06
434	ffaith	E (n.)	471	490	403	NO			2.16	96.27
435	llais	E (n.)	313	469	573	NO			1.88	96.98
436	rhiant	E (n.)	563	510	353	YES			no data	93.16
437	gŵyl	E (n.)	749	725	278	NO			1.36	78.51
438	gwybod	E (n.)	564	394	401	YES			no data	97.29
439	dathlu	B (v.)	483	774	345	NO			1.48	85.02
440	gwerth	E (n.)	530	495	376	NO			2.80	95.24
441	gwobr	E (n.)	859	1444	245	NO			1.88	56.99
442	cyfieithu	B (v.)	232	864	671	YES			no data	78.66
443	cerdd	E (n.)	455	619	386	NO	advanced		2.00	92.56
444	heno	Adf (adv.)	484	289	562	YES		neithiwr	no data	99.82
445	erthygl	E (n.)	725	1117	268	NO	advanced		2.48	67.84
446	cytuno	B (v.)	397	616	428	NO			1.60	94.08
447	anfon	B (v.)	489	750	355	YES		danfon	no data	86.64
448	tra	Cys (conj.)	418	499	450	NO	advanced		no data	98.62
449	lawr	Adf (adv.)	1089	140	1193	YES			no data	60.51
450	ariannol	Ans (adj.)	1238	1545	224	NO	specialized		3.28	49.89
451	drwg	Ans (adj.)	423	322	588	YES			no data	101.22

452	swyddfa	E (n.)	581	702	325	YES			no data	86.79
453	defnydd	E (n.)	364	1026	410	NO	advanced		2.96	79.00
454	adref	Adf (adv.)	390	364	578	YES			no data	101.84
455	i gyd	Ban (det.)	792	187	774	YES	homograph		no data	81.10
456	adran	E (n.)	336	603	515	NO	specialized		2.29	95.10
457	sgwrs	E (n.)	536	419	423	YES			no data	99.58
458	diddorol	Ans (adj.)	662	350	426	NO			1.36	96.32
459	penodol	Ans (adj.)	345	594	513	NO	advanced		3.28	95.72
460	aml	Ans (adj.)	377	492	511	NO			1.56	100.00
461	i mewn	Adf (adv.)	416	257	865	NO			1.40	91.57
462	cynnar	Ans (adj.)	381	635	456	YES			no data	95.15
463	côr	E (n.)	420	447	506	YES			no data	100.96
464	capel	E (n.)	301	656	579	YES		mosg, teml, synagog	no data	92.16
465	athro	E (n.)	339	637	524	YES		athrawes	no data	94.17
466	pwynt	E (n.)	876	276	459	NO			1.80	88.85
467	cysylltiad	E (n.)	618	799	331	NO			2.88	83.06
468	darn	E (n.)	378	461	544	NO			2.00	101.25
469	proses	E (n.)	326	636	557	NO	advanced		2.28	93.84
470	ymddangos	B (v.)	294	1105	538	NO	include as derivative under "dangos"		3.00	76.44
471	gwrando	B (v.)	491	298	628	YES			no data	99.77
472	ar draws	Adf (adv.)	434	515	478	NO	advanced		2.32	99.36
473	myfyriwr	E (n.)	334	633	552	NO		myfyrwraig	1.76	94.50
474	safle	E (n.)	847	638	309	NO	advanced		2.08	82.23

475	corff	E (n.)	396	577	497	NO			1.72	97.46
476	draw	Adf (adv.)	437	328	655	NO			2.14	100.47
477	chwech	Rhi (num.)	672	201	874	NO			1.00	84.56
478	llafur	E (n.)	600	1168	321	NO	overrepresented		3.14	72.58
479	system	E (n.)	411	555	498	NO	advanced		2.09	98.47
480	agor	B (v.)	379	442	584	YES		agored	no data	102.05
481	coed	E (n.)	422	705	451	NO			1.28	92.81
482	cyhoeddus	Ans (adj.)	727	904	313	NO			2.28	77.81
483	sefyllfa	E (n.)	393	512	541	NO	advanced		2.68	100.17
484	mynydd	E (n.)	669	803	341	YES			no data	82.92
485	rheoli	B (v.)	708	766	334	NO			2.48	83.25
486	adeilad	E (n.)	692	804	339	NO			1.84	82.34
487	darparu	B (v.)	553	1097	362	NO	advanced		3.40	76.17
488	chwerthin	B (v.)	459	195	1786	NO			1.68	64.38
489	prydain	E (n.)	476	835	409	NO		prydeinig	1.75	87.48
490	effaith	E (n.)	546	816	381	NO	advanced		2.72	86.64
491	llyfrgell	E (n.)	1390	1364	252	NO			1.08	53.63
492	gwyllo	B (v.)	493	482	479	YES			no data	101.25
493	rhedeg	B (v.)	465	343	625	YES			no data	102.65
494	archif	E (n.)	2422	3680	217	NO	specialized		3.20	26.92
495	cam	E (n.)	782	944	307	NO	advanced		2.32	76.51
496	gan gynnwys	Ar (prep.)	506	2019	356	NO	advanced		no data	56.21
497	coleg	E (n.)	532	597	424	YES			no data	96.65
498	gwe	E (n.)	1298	574	294	NO		ar-lein, cyfrifiadur	1.62	72.75

499	disgybl	E (n.)	262	988	763	NO			2.32	77.69
500	e-bost	E (n.)	1438	608	285	YES			no data	68.42
501	casglu	B (v.)	354	673	558	NO			1.92	95.65
502	canlyniad	E (n.)	406	1121	447	NO	advanced		2.68	79.42
503	gweithredu	B (v.)	517	762	413	NO	advanced		3.56	90.81
504	rownd	E (n.)	389	286	996	NO			1.58	91.93
505	yn ôl	Ar (prep.)	329	784	587	YES	homograph		no data	90.75
506	staff	E (n.)	519	486	489	NO			1.23	101.34
507	o gwmpas	Adf (adv.)	415	365	732	YES			no data	100.50
508	bwyta	B (v.)	567	437	482	YES		brecwast, cinio, te, swper, yfed, coffi	no data	102.13
509	saesneg	E (n.)	264	647	951	NO			1.00	84.51
510	hapus	Ans (adj.)	508	331	646	YES			no data	102.52
511	adeiladu	B (v.)	787	537	382	NO			1.44	91.38
512	mas	Adf (adv.)	888	161	1615	YES			no data	61.94
513	coch	Ans (adj.)	468	587	491	YES		du, brown, glas, gwyrdd, llwyd, melyn, oren, pinc, piws, porffor	no data	99.46
514	mil	E (n.)	821	205	859	NO			1.52	84.16
515	fath	Ban (det.)	289	1008	660	NO	tagging error		3.13	81.61
516	gwreiddiol	Ans (adj.)	477	646	470	YES			no data	97.46
517	sefyll	B (v.)	361	450	716	NO			1.42	101.15

518	duw	Ebych (interj.)	655	229	881	YES	homograph		no data	89.64
519	llwyddo	B (v.)	317	789	633	NO	advanced		2.36	90.93
520	methu	B (v.)	566	355	560	NO		ffaelu	2.16	104.27
521	llygad	E (n.)	270	768	875	YES		clust	no data	84.05
522	lefel	E (n.)	417	454	619	NO			1.32	104.08
523	byth	Adf (adv.)	439	363	737	YES			no data	101.47
524	teithio	B (v.)	449	735	481	YES			no data	95.11
525	criw	E (n.)	303	634	801	NO	colloquial		1.92	91.85
526	ar ôl	Ar (prep.)	357	414	820	YES	homograph		no data	99.15
527	mân	Ans (adj.)	731	324	545	NO	tagging error	replace with "man"	2.76	98.84
528	cwm	E (n.)	338	759	609	NO			1.40	93.77
529	manylyn	E (n.)	679	1175	357	NO	include as "manylyion"		3.52	75.03
530	posib	Ans (adj.)	698	391	494	NO			1.52	100.21
531	diweddar	B (v.)	983	1665	292	NO	advanced		3.52	58.33
532	ychwanegu	B (v.)	776	1136	343	NO			2.88	74.09
533	testun	E (n.)	310	846	699	NO	advanced		3.04	87.98
534	fferm	E (n.)	518	672	461	YES			no data	97.33
535	perthynas	E (n.)	300	847	757	NO	advanced		2.56	86.30
536	comisiynydd	E (n.)	1746	1807	256	NO	overrepresented		3.16	46.36
537	unigolyn	E (n.)	424	666	550	NO	advanced		3.24	98.35
538	undeb	E (n.)	441	893	495	NO	overrepresented		3.16	89.76
539	fan yna	Adf (adv.)	2663	126	3293	YES	include under "man"		2.20	29.99

540	tref	E (n.)	467	1222	448	NO			1.08	78.66
541	gweithgaredd	E (n.)	569	657	452	NO	include as derivative under "gwaith"		2.40	97.07
542	ar hyn o bryd	Adf (adv.)	993	387	438	YES			no data	90.79
543	pell	Adf (adv.)	463	921	476	YES			no data	89.17
544	rhag	Ar (prep.)	436	969	503	NO			no data	87.36
545	paratoi	B (v.)	440	665	547	NO			1.96	98.98
546	ennill	E (n.)	464	585	555	YES	homograph		no data	101.63
547	dadl	E (n.)	504	703	486	NO	advanced		3.08	97.24
548	darganfod	B (v.)	355	1034	597	NO	advanced		3.04	84.91
549	megis	Cys (conj.)	366	2692	523	NO	advanced		no data	50.09
550	yna	Ban (det.)	1047	169	1419	YES	homograph		no data	66.34
551	gêm	E (n.)	302	459	1161	YES			2.12	87.76
552	gŵr	E (n.)	340	620	793	YES			no data	95.13
553	cof	E (n.)	398	500	751	NO			2.76	100.36
554	gwerthu	B (v.)	721	521	466	YES			no data	97.56
555	eistedd	B (v.)	348	381	1142	YES			no data	90.37
556	ffilm	E (n.)	461	945	519	YES			no data	88.31
557	diogelwch	E (n.)	1273	1452	297	NO	advanced		2.79	59.24
558	sefydliad	E (n.)	845	889	380	NO	advanced		3.28	81.65
559	tan	Ar (prep.)	638	378	603	YES	advanced		no data	102.87
560	cyfrannu	B (v.)	795	1254	371	NO	advanced		3.40	72.68
561	o gwbl	Adf (adv.)	625	325	726	NO			1.38	100.20

562	oh	Ebych (interj.)	3580	133	3043	NO	spoken discourse marker		1.22	28.10
563	tudalen	E (n.)	585	667	485	YES			no data	97.50
564	neis	Ans (adj.)	1995	161	1322	YES			no data	52.65
565	map	E (n.)	1843	1076	293	NO			1.31	56.72
566	rhagor	Rha (pron.)	781	977	400	NO			no data	81.16
567	casgliad	E (n.)	733	1148	402	NO	advanced		3.00	77.35
568	lle	Ans (adj.)	413	449	853	YES	homograph		no data	99.31
569	cychwyn	E (n.)	604	609	507	YES			no data	99.21
570	yn barod	Adf (adv.)	565	373	710	NO			1.48	103.03
571	rhain	Rha (pron.)	867	336	600	NO	advanced		no data	95.58
572	ambell	Ban (det.)	458	588	648	NO			no data	100.95
573	croeso	E (n.)	624	379	652	YES			no data	103.12
574	gardd	E (n.)	460	869	567	YES			no data	91.94
575	ei gilydd	Rha (pron.)	374	432	1009	NO	advanced		no data	95.60
576	cost	E (n.)	940	1569	354	YES			no data	64.02
577	eang	Ans (adj.)	719	1212	408	NO	advanced		3.16	76.85
578	dal	Adf (adv.)	542	374	790	YES	homograph		no data	101.25
579	cwci	E (n.)	8052	12348	250	NO	overrepresented		1.98	9.71
580	sesiwn	E (n.)	870	549	453	NO			1.65	93.78
581	tuag at	Ar (prep.)	452	625	653	NO	advanced		no data	100.49
582	personol	Ans (adj.)	516	814	540	NO	include under "person"		1.60	94.15

583	ysbyty	E (n.)	661	818	460	YES		moddion, doctor, meddyg, nyrs, triniaeth	no data	91.38
584	gwir	E (n.)	693	295	850	YES		celwydd, gau	no data	95.84
585	sioe	E (n.)	544	736	535	YES			no data	97.02
586	pwnc	E (n.)	495	728	575	NO			1.76	97.95
587	diwydiant	E (n.)	1367	1643	310	NO	advanced		2.72	56.85
588	dim byd	Rha (pron.)	685	246	1152	YES			no data	86.49
589	awdur	E (n.)	743	980	429	NO		awdures	1.76	84.18
590	troed	E (n.)	425	527	817	YES			no data	99.90
591	marchnad	E (n.)	1797	1593	296	NO			1.56	51.93
592	siaradwr	E (n.)	431	819	647	NO			1.94	94.36
593	dull	E (n.)	395	1174	642	NO	advanced		3.24	82.68
594	i lawr	Adf (adv.)	385	528	916	YES	include under "lawr"		no data	97.65
595	cyfeirio	B (v.)	453	858	612	NO	advanced		3.33	93.66
596	cais	E (n.)	766	1067	425	NO	advanced		2.60	81.51
597	cau	B (v.)	573	506	627	YES			no data	104.09
598	ynghyd	Adf (adv.)	512	2056	490	NO	specialized		3.68	62.27
599	cefnogaeth	E (n.)	862	1685	389	NO	advanced		3.20	64.71
600	gwraig	E (n.)	353	653	944	YES		partner, priod, sboner, wejen	no data	93.20
601	sant	E (n.)	613	933	493	NO		santes	1.56	89.78
602	awgrymu	B (v.)	446	828	657	NO	advanced		2.96	94.26
603	do	Adf (adv.)	1315	166	2215	NO	advanced		1.00	52.70



604	calon	E (n.)	448	489	878	NO			1.52	99.72
605	blynyddol	Ans (adj.)	1120	1991	349	NO	advanced		2.84	56.17
606	cyffredin	Ans (adj.)	637	854	505	NO			2.40	92.12
607	llythyr	E (n.)	421	770	739	YES		llythyren	no data	94.98
608	ti'n gwybod	Ebych (interj.)	2907	150	2884	NO	spoken discourse marker		1.50	33.98
609	trafodaeth	E (n.)	671	639	533	NO	advanced		3.00	99.11
610	bron	Adf (adv.)	523	474	776	NO			2.16	102.60
611	ymarfer	E (n.)	375	732	867	YES			no data	93.67
612	ar hyd	Ar (prep.)	435	802	730	NO			no data	94.09
613	llall	Rha (pron.)	426	478	1004	NO	advanced		no data	96.74
614	dinas	E (n.)	562	605	626	YES			no data	102.20
615	gwella	B (v.)	818	681	484	NO			2.04	93.82
616	sut	Ban (det.)	505	629	677	YES			no data	101.61
617	ynglŷn	Ar (prep.)	686	397	743	NO	advanced		no data	101.03
618	hwyl	Ebych (interj.)	337	733	1052	YES			no data	88.81
619	rheini	Rha (pron.)	763	299	950	YES	advanced		no data	93.17
620	maint	E (n.)	793	375	682	NO			2.28	100.33
621	lladd	B (v.)	576	659	601	NO	advanced		2.32	101.12
622	sadwrn	E (n.)	904	513	531	YES			no data	96.22
623	lloegr	E (n.)	751	737	512	YES			no data	94.17
624	sŵn	E (n.)	501	317	1380	NO			1.64	86.83
625	teledu	E (n.)	498	648	712	YES			no data	100.65

626	lot	E (n.)	1472	197	1291	NO	non-standard		1.50	66.72
627	sgil	E (n.)	674	1031	499	NO			2.64	86.98
628	gwleidyddol	Ans (adj.)	704	579	581	NO	advanced		3.28	100.79
629	enwedig	Ans (adj.)	680	565	593	NO	advanced		2.84	102.15
630	penderfyniad	E (n.)	837	899	468	NO	advanced		3.08	87.34
631	arweinydd	E (n.)	650	922	532	NO	advanced		3.12	91.10
632	am ddim	Adf (adv.)	1879	813	370	YES	include as compound under "dim"		1.36	65.23
633	sicr	Ans (adj.)	603	580	643	NO	advanced		2.04	103.29
634	iâ	E (n.)	3016	158	2754	NO	tagging error		1.32	35.31
635	hollol	Ans (adj.)	760	357	804	NO			2.56	99.14
636	cryf	Ans (adj.)	550	546	754	NO		gwan	1.56	102.56
637	nod	E (n.)	548	990	592	NO	advanced		2.36	90.86
638	duw	E (n.)	808	306	958	YES		duwies	no data	93.21
639	tyfu	B (v.)	629	518	684	NO			1.64	103.89
640	rhif	E (n.)	642	440	772	YES			no data	102.92
641	tu allan	Adf (adv.)	775	642	553	NO		tu fas, tu fewn	1.52	97.80
642	wy	E (n.)	789	273	1149	YES			no data	88.53
643	arfer	E (n.)	606	859	574	NO			1.68	95.17
644	eleni	Adf (adv.)	683	641	594	YES			no data	100.50
645	cednll	E (n.)	646	951	548	NO	advanced		2.96	91.29
646	blaenorol	Ans (adj.)	1597	2476	346	NO	advanced		3.32	47.36
647	brawd	E (n.)	409	593	1048	YES		chwaer	no data	95.23
648	eithaf	Adf (adv.)	1278	231	1086	YES			no data	77.41
649	cymeriad	E (n.)	482	874	733	NO			2.76	93.93

650	yno	Ban (det.)	429	1003	811	NO			no data	88.36
651	hyfryd	Ans (adj.)	900	434	621	YES			no data	99.78
652	cywir	Ans (adj.)	732	491	669	NO		anghywir	1.60	102.78
653	naturiol	Ans (adj.)	657	749	586	NO			2.08	98.43
654	fideo	E (n.)	1466	1521	374	YES			no data	61.70
655	ci	E (n.)	408	436	1450	YES			no data	87.20
656	llai	Rha (pron.)	828	472	636	NO			no data	101.30
657	braf	Ans (adj.)	592	498	797	YES		diflas	no data	103.70
658	lot	Rha (pron.)	1901	198	1250	NO	non-standard		1.50	62.24
659	marw	B (v.)	529	543	835	NO			1.60	103.04
660	tymor	E (n.)	596	708	637	NO			1.42	101.61
661	w	Ebych (interj.)	819	318	959	NO	spoken discourse marker		1.45	95.16
662	cofnodi	B (v.)	992	2034	417	NO	advanced		3.44	60.99
663	radio	E (n.)	1053	473	570	YES			no data	95.41
664	cynllunio	B (v.)	1127	718	467	NO	advanced		3.08	87.63
665	haf	E (n.)	599	586	725	YES		gaeaf, gwanwyn, hydref	no data	103.71
666	dwyieithog	Ans (adj.)	407	795	974	NO			2.44	92.69
667	prysur	Ans (adj.)	702	607	635	YES			no data	102.41
668	stryd	E (n.)	767	807	551	YES			no data	94.89
669	llwybr	E (n.)	807	1248	492	NO			2.00	80.93
670	diwylliant	E (n.)	537	794	727	NO	advanced		2.84	97.84

671	teg	Ans (adj.)	765	380	870	NO			1.88	99.78
672	adeg	E (n.)	560	631	791	NO	advanced		2.92	101.36
673	sylwad	E (n.)	1560	1124	396	NO	advanced		3.48	68.52
674	i fyny	Adf (adv.)	574	417	1002	YES			no data	101.13
675	ffermwr	E (n.)	1075	1311	435	YES	specialized		no data	74.40
676	galw	B (v.)	711	407	860	YES			no data	102.06
677	cynnwys	B (v.)	589	1478	577	NO			2.33	79.08
678	ymyrch	E (n.)	1061	1460	434	NO	advanced		3.60	71.61
679	cymry	E (n.)	551	967	670	NO		cymro, cymraes, sais	2.20	93.81
680	oedolyn	E (n.)	633	985	599	NO	include as derivative under "oed"		1.96	92.85
681	isdeitl	E (n.)	236	6412	2811	NO	overrepresented		2.45	23.98
682	gweddill	E (n.)	533	998	690	NO	advanced		2.80	92.94
683	rygbi	E (n.)	607	796	651	YES			no data	99.66
684	pentref	E (n.)	612	1532	569	NO	advanced		1.20	77.96
685	ffurflen	E (n.)	1731	1388	384	NO			1.72	61.85
686	bardd	E (n.)	647	865	618	NO	specialized		2.48	96.91
687	ennill	B (v.)	611	560	783	YES			no data	104.61
688	ystafell	E (n.)	511	1185	693	YES		ymolchi, golchi	no data	87.80
689	pecyn	E (n.)	1308	1410	416	NO	advanced		2.00	68.84
690	yr un	Adf (adv.)	832	332	1006	NO	include under "un"		1.98	95.83
691	lleoliad	E (n.)	689	1119	576	NO	include as derivative under "lle"		2.68	88.30

692	isel	Ans (adj.)	715	761	614	NO			2.00	99.29
693	ffurf	E (n.)	447	1308	825	NO	advanced		3.08	82.50
694	union	Ans (adj.)	676	479	845	NO	tagging error		2.68	103.43
695	ffôn	E (n.)	634	445	928	YES		ffonio	no data	103.25
696	post	E (n.)	1189	1244	440	YES			no data	75.17
697	clir	Ans (adj.)	659	652	714	NO	advanced		1.92	102.71
698	hyrwyddo	B (v.)	925	1391	483	NO	advanced		3.60	77.15
699	cefndir	E (n.)	920	476	675	NO	advanced		2.96	100.97
700	adnodd	E (n.)	880	1269	510	NO	advanced		3.04	81.01
701	ger	Ar (prep.)	753	1140	559	NO	advanced		no data	87.21
702	anifail	E (n.)	545	842	784	YES		cath, ci	no data	97.25
703	cylch	E (n.)	879	777	566	NO	overrepresented		1.92	95.40
704	categori	E (n.)	1468	3043	392	NO	advanced		2.48	46.30
705	afon	E (n.)	622	798	687	YES	overrepresented		no data	100.21
706	saith	E (n.)	1020	260	1339	NO			1.00	82.73
707	nofel	E (n.)	1121	776	514	NO	advanced		1.16	89.19
708	o hyd	Adf (adv.)	559	700	838	YES	advanced		no data	101.00
709	canu	B (v.)	584	396	1177	YES			no data	98.66
710	cariad	E (n.)	513	583	1013	YES		annwyl, caru, dwli	no data	100.75
711	llwyfan	E (n.)	555	1137	720	NO	advanced		2.36	89.60
712	ardal	Ans (adj.)	881	891	556	YES	homograph		no data	92.58
713	taliad	E (n.)	1977	3789	372	NO	include as derivative under "talu"		2.80	37.84
714	diogelu	B (v.)	1191	2011	439	NO	advanced		3.16	61.89

715	gwely	E (n.)	487	606	1079	YES			no data	98.79
716	geni	B (v.)	632	782	741	NO			1.96	99.59
717	iau	E (n.)	868	615	644	YES			2.20	100.87
718	cwrdd	E (n.)	583	676	821	YES			no data	102.98
719	o hyd i	Ar (prep.)	525	1421	746	NO	advanced		no data	82.02
720	profi	B (v.)	582	741	802	NO	advanced		2.52	101.32
721	senedd	E (n.)	1784	1186	418	NO			2.16	66.70
722	llawer	Ban (det.)	628	1109	654	YES	homograph		no data	91.53
723	dysgwr	E (n.)	541	548	1032	YES	include as derivative under "dysgu"		1.08	101.86
724	disgwyl	E (n.)	549	845	826	NO			2.26	97.98
725	tiwtor	E (n.)	380	1081	1147	YES			no data	85.01
726	agwedd	E (n.)	540	1079	788	NO	advanced		3.36	91.43
727	pryhawn	E (n.)	614	429	1089	YES			no data	101.89
728	cyflym	Ans (adj.)	901	662	617	YES		araf	no data	100.04
729	datblygiad	E (n.)	677	1082	638	NO	include as derivative under "datblygiad"		3.32	92.10
730	caled	Ans (adj.)	635	530	907	NO			1.72	104.85
731	sydyn	Ans (adj.)	756	296	1683	NO			1.92	82.04
732	eisoes	Adf (adv.)	694	1439	604	NO	advanced		3.44	82.09
733	cynulleidfa	E (n.)	571	857	814	NO	advanced		2.92	98.19
734	coginio	B (v.)	999	1536	496	YES		cegin	no data	75.02
735	ymdrech	E (n.)	486	1189	873	NO	advanced		2.96	87.85
736	deunydd	E (n.)	485	1640	823	NO	advanced		3.28	77.12

737	mynnu	B (v.)	645	1039	683	NO	advanced		3.20	94.04
738	safbwynt	E (n.)	675	685	782	NO	advanced		3.40	102.83
739	cerdyn	E (n.)	1495	1021	458	YES			no data	76.79
740	cyfrifol	Ans (adj.)	724	1154	620	NO	advanced		3.04	89.96
741	natur	E (n.)	481	1005	956	NO			1.58	91.90
742	prawf	E (n.)	456	873	1058	NO	advanced		1.60	93.90
743	datganiad	E (n.)	1284	1157	474	NO	advanced		3.48	78.57
744	ugain	Rhi (num.)	745	389	1114	NO			1.28	99.25
745	lan	Adf (adv.)	1578	212	1916	NO			1.48	63.21
746	cystadleuaeth	E (n.)	586	1353	711	NO	advanced		2.16	85.93
747	canfod	B (v.)	945	2800	502	NO	advanced		3.72	55.82
748	presennol	Ans (adj.)	822	1515	563	NO		gorffennol	1.60	79.41
749	cyflawni	B (v.)	595	1220	728	NO	advanced		3.44	89.48
750	licio	B (v.)	1428	225	1806	YES			no data	67.76

## Appendix B: the basic 500-word list, without additions

Note that all headwords in the list are given without a part of speech tag, and without separation of homographs with different etymologies. As a result, for example, the list item *de* covers the meanings “South” and “right”, and *ysgol* covers both “school” and “ladder”.

Lemma	Geirfa Graidd?	Additions?	Tutors' survey average
<b>a</b>	YES		1.64
<b>â</b>	NO		no data
<b>achos</b>	YES		no data
<b>addysg</b>	NO		2.19
<b>adeilad</b>	NO		1.84
<b>adeiladu</b>	NO		1.44
<b>adnabod</b>	YES		no data
<b>adref</b>	YES		no data
<b>aelod</b>	NO		2.19
<b>agor</b>	YES	agored	no data
<b>agos</b>	YES		no data
<b>ail</b>	NO		1.67
<b>allan</b>	YES		no data
<b>am</b>	YES		no data
<b>ambell</b>	NO		no data
<b>aml</b>	NO		1.56
<b>amser</b>	YES		no data
<b>anfon</b>	YES	danfon	no data
<b>angen</b>	YES		no data
<b>anifail</b>	YES	cath, ci	no data
<b>anodd</b>	YES		no data
<b>ar</b>	YES		no data
<b>ar gael</b>	YES		no data
<b>ar gyfer</b>	NO		no data
<b>ar hyd</b>	NO		no data
<b>ar hyn o bryd</b>	YES		no data
<b>ar ôl</b>	YES		no data
<b>arall</b>	YES		no data
<b>arbennig</b>	NO		1.56
<b>ardal</b>	YES		no data
<b>arfer</b>	NO		1.68
<b>arian</b>	YES		no data
<b>aros</b>	YES		no data



<b>at</b>	YES		no data
<b>ateb</b>	YES		no data
<b>athro</b>	YES	athrawes	no data
<b>awdur</b>	NO	awdures	1.76
<b>awr</b>	YES		no data
<b>bach</b>	YES		1.00
<b>barn</b>	NO		2.20
<b>beth</b>	YES		no data
<b>ble</b>	YES		no data
<b>blwyddyn</b>	YES		no data
<b>bod</b>	YES		no data
<b>bore</b>	YES		no data
<b>braf</b>	YES	diflas	no data
<b>brawd</b>	YES	chwaer	no data
<b>bron</b>	NO		2.16
<b>busnes</b>	NO		1.59
<b>bwrdd</b>	YES		no data
<b>bwyd</b>	YES		no data
<b>bwyta</b>	YES	brecwast, cinio, te, swper, yfed, coffi	no data
<b>byd</b>	NO		1.38
<b>byth</b>	YES		no data
<b>byw</b>	YES		no data
<b>bywyd</b>	NO		2.07
<b>cadw</b>	YES		no data
<b>cael</b>	YES		no data
<b>caled</b>	NO		1.72
<b>calon</b>	NO		1.52
<b>cân</b>	NO		1.33
<b>canol</b>	YES		no data
<b>canolfan</b>	YES		no data
<b>canu</b>	YES		no data
<b>capel</b>	YES	mosg, teml, synagog	no data
<b>car</b>	YES		no data
<b>cariad</b>	YES	annwyl, caru, dwli	no data
<b>cartref</b>	YES	adref, gartref	no data
<b>casglu</b>	NO		1.92
<b>cau</b>	YES		no data
<b>cefn</b>	YES		no data

<b>cefnogi</b>	NO		2.37
<b>ceisio</b>	NO		2.30
<b>ceneditaethol</b>	NO		2.37
<b>cerdded</b>	YES		no data
<b>cerddoriaeth</b>	NO		1.44
<b>cerdyn</b>	YES		no data
<b>ci</b>	YES		no data
<b>cig</b>	YES	llysiau, ffrwythau	no data
<b>clwb</b>	YES		no data
<b>clywed</b>	YES	blasu, aroglu	no data
<b>coch</b>	YES	du, brown, glas, gwyrdd, llwyd, melyn, oren, pinc, piws, porffor	no data
<b>codi</b>	YES		no data
<b>coed</b>	NO		1.28
<b>cof</b>	NO		2.76
<b>cofio</b>	YES		no data
<b>coginio</b>	YES	cegin	no data
<b>coleg</b>	YES		no data
<b>colli</b>	YES		no data
<b>côr</b>	YES		no data
<b>corff</b>	NO		1.72
<b>cost</b>	YES		no data
<b>credu</b>	YES		no data
<b>croeso</b>	YES		no data
<b>cryf</b>	NO	gwan	1.56
<b>cwestiwn</b>	YES		no data
<b>cwm</b>	NO		1.40
<b>cwmni</b>	NO		1.63
<b>cwrdd</b>	YES		no data
<b>cwrs</b>	YES		no data
<b>cychwyn</b>	YES		no data
<b>cyfarfod</b>	YES		no data
<b>cyfeiriad</b>	YES		no data
<b>cyffredin</b>	NO		2.40
<b>cyfieithu</b>	YES		no data
<b>cyfle</b>	NO		2.33
<b>cyflym</b>	YES	araf	no data
<b>cyhoeddus</b>	NO		2.28
<b>cymaint</b>	NO		no data

<b>cymdeithas</b>	NO		2.33
<b>cymeriad</b>	NO		2.76
<b>cymraeg</b>	YES	saesneg	no data
<b>cymreig</b>	NO	seisnig, celtaidd, ewropeaidd	2.72
<b>cymru</b>	YES	lloegr, yr alban, iwerddon, gogledd iwerddon, ewrop	no data
<b>cymry</b>	NO	cymro, cymraes, sais	2.20
<b>cymryd</b>	YES		no data
<b>cyn</b>	YES		no data
<b>cyngor</b>	NO		1.63
<b>cynnar</b>	YES		no data
<b>cynnig</b>	NO		2.33
<b>cynnwys</b>	NO		2.33
<b>cyntaf</b>	NO		1.52
<b>cyrraedd</b>	YES		no data
<b>cysylltiad</b>	NO		2.88
<b>cysylltu</b>	NO		2.30
<b>cytuno</b>	NO		1.60
<b>cywir</b>	NO	anghywir	1.60
<b>chi</b>	YES		no data
<b>chwarae</b>	YES		no data
<b>chwech</b>	NO		1.00
<b>chwerthin</b>	NO		1.68
<b>chwilio</b>	NO		1.67
<b>da</b>	YES		no data
<b>dal</b>	YES		no data
<b>dan</b>	YES		no data
<b>dangos</b>	NO		1.67
<b>darllen</b>	YES		no data
<b>darn</b>	NO		2.00
<b>datblygu</b>	NO		2.74
<b>dathlu</b>	NO		1.48
<b>dau/dwy</b>	NO		1.00
<b>de</b>	YES	chwith	no data
<b>deall</b>	YES		no data
<b>dechrau</b>	YES		no data
<b>defnyddio</b>	YES	defnyddiol	no data
<b>deg</b>	NO		1.00

<b>derbyn</b>	NO		1.96
<b>dewis</b>	NO		1.46
<b>diddordeb</b>	YES		no data
<b>diddorol</b>	NO		1.36
<b>digon</b>	YES		no data
<b>digwydd</b>	YES		no data
<b>dilyn</b>	NO		2.15
<b>dim</b>	YES		no data
<b>dim byd</b>	YES		no data
<b>dim ond</b>	YES		no data
<b>dinas</b>	YES		no data
<b>diolch</b>	YES		no data
<b>disgwyl</b>	NO		2.26
<b>disgybl</b>	NO		2.32
<b>diwedd</b>	NO		1.30
<b>diweddar</b>	NO		2.67
<b>diwethaf</b>	YES		no data
<b>diwrnod</b>	YES		no data
<b>dod</b>	YES		no data
<b>dosbarth</b>	YES		no data
<b>draw</b>	NO		2.14
<b>drwg</b>	YES		no data
<b>drws</b>	YES		no data
<b>duw</b>	YES	duwies	no data
<b>dweud</b>	YES		no data
<b>dŵr</b>	YES		no data
<b>dwyieithog</b>	NO		2.44
<b>dy</b>	YES		no data
<b>dydd</b>	YES	dyddiad	no data
<b>dyfodol</b>	YES		no data
<b>dyma</b>	YES		no data
<b>dyn</b>	YES	dynes	no data
<b>dyna</b>	YES		no data
<b>dysgu</b>	YES		no data
<b>e-bost</b>	YES		no data
<b>edrych</b>	YES		no data
<b>efallai</b>	YES		1.60
<b>efo</b>	YES		no data
<b>eglwys</b>	YES		no data

<b>ei</b>	YES		no data
<b>eich</b>	YES		no data
<b>ein</b>	YES		no data
<b>eisiau</b>	YES	moyn	no data
<b>eistedd</b>	YES		no data
<b>eisteddfod</b>	YES		no data
<b>eithaf</b>	YES		no data
<b>eleni</b>	YES		no data
<b>ennill</b>	YES		no data
<b>enw</b>	YES		no data
<b>erbyn</b>	YES		no data
<b>erioed</b>	YES		no data
<b>ers</b>	YES		no data
<b>eto</b>	YES		no data
<b>eu</b>	YES		no data
<b>fel</b>	YES		no data
<b>felly</b>	NO		1.33
<b>fi</b>	YES		no data
<b>fideo</b>	YES		no data
<b>fo</b>	YES		no data
<b>fy</b>	YES		no data
<b>ffaith</b>	NO		2.16
<b>fferm</b>	YES		no data
<b>ffilm</b>	YES		no data
<b>ffôn</b>	YES	ffonio	no data
<b>ffordd</b>	YES		no data
<b>ffrind</b>	YES		no data
<b>ffurflen</b>	NO		1.72
<b>gadael</b>	YES		no data
<b>gair</b>	YES		no data
<b>gallu</b>	YES		no data
<b>galw</b>	YES		no data
<b>gan</b>	NO		no data
<b>gardd</b>	YES		no data
<b>gêm</b>	YES		2.12
<b>geni</b>	NO		1.96
<b>gobeithio</b>	NO	gobaith	1.40
<b>gofyn</b>	YES	holi	no data
<b>gogledd</b>	YES	dwyrain, gorllewin, de	no data

<b>golwg</b>	NO		2.72
<b>gosod</b>	NO		2.75
<b>grŵp</b>	YES		no data
<b>gwahanol</b>	NO	gwahaniaeth	1.85
<b>gwaith</b>	YES		no data
<b>gwasanaeth</b>	NO		1.81
<b>gwe</b>	NO	ar-lein, cyfrifiadur	1.62
<b>gwefan</b>	NO		1.44
<b>gweinidog</b>	NO	offeiriad, imam, rabi	2.48
<b>gweithio</b>	YES		no data
<b>gweld</b>	YES		no data
<b>gwella</b>	NO		2.04
<b>gwely</b>	YES		no data
<b>gwerth</b>	NO		2.80
<b>gwerthu</b>	YES		no data
<b>gwir</b>	YES	celwydd, gau	no data
<b>gwlad</b>	YES		no data
<b>gwneud</b>	YES		no data
<b>gwobr</b>	NO		1.88
<b>gŵr</b>	YES		no data
<b>gwraig</b>	YES	partner, priod, sboner, wejen	no data
<b>gwrando</b>	YES		no data
<b>gwreiddiol</b>	YES		no data
<b>gwybod</b>	YES		no data
<b>gwybodaeth</b>	NO		1.93
<b>gwyd</b>	NO		1.36
<b>gŵyl</b>	NO		1.36
<b>gwylio</b>	YES		no data
<b>gwyn</b>	YES		no data
<b>gyda</b>	YES		no data
<b>gyrru</b>	YES		no data
<b>haf</b>	YES	gaeaf, gwanwyn, hydref	no data
<b>hanes</b>	NO		1.85
<b>hanner</b>	YES		no data
<b>hapus</b>	YES		no data
<b>hawdd</b>	YES		no data
<b>heb</b>	YES		no data
<b>heddiw</b>	YES	ddoe, yfory	no data
<b>hefyd</b>	YES		no data

helpu	YES		no data
hen	YES		no data
heno	YES	neithiwr	no data
hi	YES		no data
hir	YES		no data
hoffi	YES	lico, casáu	no data
hollol	NO		2.56
hwyl	YES		no data
hyd yn oed	NO		2.56
hyfryd	YES		no data
i	YES		no data
i fyny	YES		no data
i gyd	YES		no data
i mewn	NO		1.40
iaith	NO		1.15
iau	YES		2.20
iawn	YES		no data
iechyd	YES	iach, sâl, salwch, gwael, tost, anabl, anabledd	no data
ifanc	YES		no data
isel	NO		2.00
lan	NO		1.48
lawr	YES		no data
lefel	NO		1.32
licio	YES		no data
llai	NO		no data
llais	NO		1.88
llaw	YES	bys, bawd	no data
llawer	YES		no data
llawn	NO		1.44
lle	YES		no data
lleol	NO		1.48
lloegr	YES		no data
llun	YES		no data
llwybr	NO		2.00
llyfr	YES		no data
llyfrgell	NO		1.08
llygad	YES	clust	no data
llythyr	YES	llythyren	no data
maes	YES		no data

<b>maint</b>	NO		2.28
<b>mam</b>	YES		no data
<b>man</b>	NO		2.76
<b>manylion</b>	NO		3.52
<b>map</b>	NO		1.31
<b>marchnad</b>	NO		1.56
<b>marw</b>	NO		1.60
<b>mas</b>	YES		no data
<b>math</b>	NO		2.25
<b>mawr</b>	YES		no data
<b>meddwl</b>	YES		no data
<b>medru</b>	YES		no data
<b>merch</b>	YES	menyw, mab	no data
<b>methu</b>	NO	ffaelu	2.16
<b>mewn</b>	YES		no data
<b>mil</b>	NO		1.52
<b>mis</b>	YES		no data
<b>mor</b>	NO		2.06
<b>môr</b>	YES		no data
<b>munud</b>	YES		no data
<b>mwy</b>	YES		no data
<b>mwynhau</b>	YES		no data
<b>myfyriwr</b>	NO	myfyrwraig	1.76
<b>mynd</b>	YES		no data
<b>mynydd</b>	YES		no data
<b>nadolig</b>	YES		no data
<b>natur</b>	NO		1.58
<b>naturiol</b>	NO		2.08
<b>nawr</b>	YES		no data
<b>neb</b>	YES		no data
<b>neges</b>	YES		no data
<b>neis</b>	YES		no data
<b>nesaf</b>	YES		no data
<b>neu</b>	YES		no data
<b>newid</b>	YES		no data
<b>newydd</b>	NO		1.00
<b>newyddion</b>	YES		no data
<b>nhw</b>	YES		no data
<b>ni</b>	YES		no data



nifer	NO		2.37
nos	YES	noswaith	no data
noson	YES		no data
o	YES		no data
o gwbl	NO		1.38
o gwmpas	YES		no data
ochr	NO		1.68
oed	YES		no data
oes	NO		1.40
olaf	NO		1.80
ond	YES		no data
os	NO		no data
pa	YES		no data
pam	YES		no data
pan	YES		no data
papur	YES		1.08
paratoi	NO		1.96
pawb	YES		no data
pedwar/pedair	NO		1.00
peidio	NO		1.77
pell	YES		no data
pen	YES		no data
penderfynu	NO		1.88
person	NO		1.16
peth	NO		1.45
plant	YES		no data
pob	YES		no data
pobl	NO		1.00
popeth	YES		no data
posib	NO		1.52
post	YES		no data
presennol	NO	gorffennol	1.60
prif	NO		1.74
prifysgol	NO		1.56
problem	YES		no data
pryd	YES		no data
prydain	NO	prydeinig	1.75
pryhawn	YES		no data
prynu	YES		no data

<b>prysur</b>	YES		no data
<b>pump</b>	NO		1.00
<b>pwnc</b>	NO		1.76
<b>pwyl</b>	YES		no data
<b>pwylt</b>	NO		1.80
<b>pwysig</b>	YES		no data
<b>radio</b>	YES		no data
<b>reit</b>	NO		1.34
<b>rhag</b>	NO		no data
<b>rhaglen</b>	YES		no data
<b>rhagor</b>	NO		no data
<b>rhaid</b>	NO		1.27
<b>rhan</b>	NO		1.65
<b>rhannu</b>	NO		2.00
<b>rhedeg</b>	YES		no data
<b>rheoli</b>	NO		2.48
<b>rhestr</b>	NO		1.56
<b>rheswm</b>	NO		2.16
<b>rhiant</b>	YES		no data
<b>rhif</b>	YES		no data
<b>rhoi</b>	NO		1.11
<b>rhwng</b>	YES		no data
<b>rhy</b>	YES		no data
<b>rhyw</b>	NO		2.24
<b>rhywbeth</b>	YES		no data
<b>rhywun</b>	YES		no data
<b>rownd</b>	NO		1.58
<b>rŵan</b>	YES		no data
<b>rygbi</b>	YES		no data
<b>sadwrn</b>	YES		no data
<b>saesneg</b>	NO		1.00
<b>saith</b>	NO		1.00
<b>sant</b>	NO	santes	1.56
<b>sawl</b>	YES		no data
<b>sefyll</b>	NO		1.42
<b>senedd</b>	NO		2.16
<b>sesiwn</b>	NO		1.65
<b>sgil</b>	NO		2.64
<b>sgwrs</b>	YES		no data

<b>siarad</b>	YES		no data
<b>siaradwr</b>	NO		1.94
<b>sioe</b>	YES		no data
<b>siop</b>	YES	siopa	no data
<b>sir</b>	NO		1.56
<b>siŵr</b>	YES		no data
<b>sôn</b>	NO		2.63
<b>staff</b>	NO		1.23
<b>stori</b>	YES		no data
<b>stryd</b>	YES		no data
<b>sul</b>	YES	gwener, mawrth, mercher	no data
<b>sut</b>	YES		no data
<b>sŵn</b>	NO		1.64
<b>swydd</b>	NO		1.24
<b>swyddfa</b>	YES		no data
<b>swyddog</b>	NO		2.00
<b>sydyn</b>	NO		1.92
<b>symud</b>	YES		no data
<b>syniad</b>	YES		no data
<b>tad</b>	YES		no data
<b>taith</b>	YES		no data
<b>tal</b>	YES		no data
<b>tebyg</b>	NO		2.30
<b>teg</b>	NO		1.88
<b>teimlo</b>	YES		no data
<b>teithio</b>	YES		no data
<b>teledu</b>	YES		no data
<b>teulu</b>	YES		no data
<b>ti</b>	YES		no data
<b>tîm</b>	YES		no data
<b>tipyn</b>	YES		no data
<b>tir</b>	NO		2.40
<b>tiwtor</b>	YES		no data
<b>torri</b>	YES		no data
<b>trafod</b>	NO		2.30
<b>tref</b>	NO		1.08
<b>trefnu</b>	NO	trefn	1.76
<b>tri/tair</b>	NO		1.00
<b>trio</b>	NO		1.16

<b>tro</b>	YES		no data
<b>troed</b>	YES		no data
<b>troi</b>	YES		no data
<b>tros</b>	NO		no data
<b>trwy</b>	YES		no data
<b>tu allan</b>	NO	tu fas, tu fewn	1.52
<b>tua</b>	NO		no data
<b>tudalen</b>	YES		no data
<b>tŷ</b>	YES		no data
<b>tyfu</b>	NO		1.64
<b>tymor</b>	NO		1.42
<b>tynnu</b>	NO		1.63
<b>tywydd</b>	YES		no data
<b>uchel</b>	NO		1.89
<b>ugain</b>	NO		1.28
<b>un</b>	NO	naw, wyth, cant	1.00
<b>unrhyw</b>	NO		no data
<b>unwaith</b>	YES		no data
<b>wedyn</b>	YES		no data
<b>wrth</b>	YES		no data
<b>wrth gwrs</b>	YES		no data
<b>wy</b>	YES		no data
<b>wythnos</b>	YES		no data
<b>y</b>	YES		no data
<b>ychwanegu</b>	NO		2.88
<b>ychydig</b>	NO		no data
<b>yma</b>	YES		no data
<b>ymarfer</b>	YES		no data
<b>ymlaen</b>	YES		no data
<b>ymuno</b>	NO		2.28
<b>yn</b>	YES		no data
<b>yn barod</b>	NO		1.48
<b>yn erbyn</b>	YES		no data
<b>yn ôl</b>	YES		no data
<b>yn unig</b>	NO		2.12
<b>yn ystod</b>	NO		no data
<b>yna</b>	YES		no data
<b>yno</b>	NO		no data
<b>ysbyty</b>	YES	moddion, doctor, meddyg, nyrs, triniaeth	no data

<b>ysgol</b>	YES		no data
<b>ysgrifennu</b>	YES		no data
<b>ystafell</b>	YES	ymolchi, golchi	no data

**Appendix C:** the working list of additions, as an alphabetical list

adref	dwyrain	moddion
agored	dyddiad	mosg
anabl	dynes	moyn
anabledd	ddoe	myfyrwraig
anghywir	Ewrop	naw
annwyl	Ewropeaidd	neithiwr
ar-lein	ffaelu	noswaith
araf	ffonio	nyrs
aroglu	ffrwythau	offeiriad
athrawes	gaeaf	oren
awdures	gartref	partner
bawd	gau	pinc
blasu	glas	piws
brecwast	gobaith	porffor
brown	Gogledd Iwerddon	priod
bys	golchi	Prydeinig
cant	gorffennol	rabi
caru	gorllewin	Sais
casáu	gwael	sâl
cath	gwahaniaeth	salwch
cegin	gwan	santes
celtaidd	gwanwyn	sboner
celwydd	gwener	Seisnig
ci	gwyrdd	siopa
cinio	holi	swper
clust	hydref	synagog
coffi	iach	te
cyfrifiadur	imam	templ
Cymraes	Iwerddon	tost
Cymro	lico	trefn
chwaer	Lloegr	triniaeth
chwith	llwyd	tu fas
danfon	llysiau	tu fewn
de	llythyren	wejen
defnyddiol	mab	wyth
diflas	mawrth	yfed
doctor	meddyg	yfory
du	melyn	ymolchi
duwies	menyw	Yr Alban
dwli	mercher	