# Bangor - Welsh

This section documents the transcription conventions specific to the Bangor Siarad, Pilot, and Bangor 2 corpora. The three sections are: main tier, gloss tier, and tags.

## A MAIN TIER

1. Layout of transcription

1.1. Since we are primarily interested in clauses, the data is divided into clauses as far as possible. Where an utterance contains two main clauses, each clause in that utterance is written on a separate main tier. Complex clauses are treated as one clause and therefore subordinate clauses are included in the same tier as their main clauses. Adverbial clauses are also written on the same main tier as their related main clause.

1.2. Each main tier is divided into units which we call, for the purposes of these conventions, 'words'. With some exceptions (see C.1.3) a word is considered to be a continuous sequence of characters containing no spaces, as found in Geiriadur Prifysgol Cymru (Thomas 1950-2004), Geiriadur yr Academi (Griffiths & Jones1995), Cysgeir (2004) or the Oxford English Dictionary online (2008). These are referred to as GPC, GyrA, Cysgeir and OED respectively throughout this document. Where items are treated as hyphenated by these reference dictionaries, they are connected by underscore in the transcripts. When one of the reference dictionaries offers more than one alternative (e.g. 'minibus' 'mini-bus' or 'mini bus'), or when the reference dictionaries differ from each other, the most compact alternative is chosen ('minibus' in this case).

1.3. Other items which are treated as words are:

i. interjections and interactional markers, e.g. ah, er, um etc.
ii. proper names (including names of books, films, organisations etc.), a sequence of words being connected by underscores, e.g. Elton_John, Hong_Kong, One_Flew_Over_the_Cuckoo's_Nest
iii. abbreviations (connected by underscore), e.g. N_S_P_C_C
iv. numbers between eleven and ninety-nine in Welsh and between twenty-one and ninety-nine in English, e.g. pedwar_deg_pump, forty_five. Note that other numbers such as those containing 'hundred', 'thousand' etc. are transcribed as separate words, e.g. one hundred and seventy_three, cant saith_deg_tri
v. some prepositions and adverbs, usually represented as two words, whose individual parts are meaningless or difficult to translate in isolation, e.g. oddi_wrth. See a full list below in C.3.4 vii.

1.4. Contractions that do not have entries in one of the Welsh-language reference dictionaries (namely GPC, GyrA or Cysgeir) or in King (2003), are transcribed in full, but the unpronounced parts are bracketed. For example, the pronunciation of 'fel yna' (like that) as [vɛla] in speech is represented in the transcripts as 'fel (yn)a'.

1.5. There are some continuous sequences of characters in the main tier which are not treated as words. These include simple events such as '&=laugh' (see CHAT 7.6.1), 'xx' or 'xxx' for unintelligible sounds, or the use of an ampersand plus phonetic characters for intelligible sounds without clear meaning (see CHAT 6.4 for both).

1.6. Note that we do not follow the guidelines for collocations, compounds and linkages given in the CHAT manual (see CHAT 6.6.2). We consider collocations and compounds to be single items only if they are considered words according to the definition given in C.1.2 and C.1.3. For example, CHAT gives the option of writing 'peanut butter' or 'peanut+butter', and 'Star Wars' or 'Star+Wars'. According to our conventions, the first phrase should be transcribed as two separate words ('peanut butter') as this is how the phrase appears in the OED. The second case should be transcribed as 'Star_Wars' as it is a film title. Note that we do not use the plus sign within words under any circumstances.

2. Language marking

2.1. Each word in the main tier is assigned a language marker, which consists of @s: plus one or two other letters which denote its language: @s:w = Welsh, @s:e = English, @s:u = undetermined, @s:ew = word with first morpheme(s) English, second morpheme(s) Welsh, @s:we = word with first morpheme(s) Welsh, second morpheme(s) English. Other languages have been coded as they have arisen and have been included in the depfile, e.g. @s:f = French

2.2. A word or morpheme is considered to be Welsh if it can be found in any of the Welsh-language reference dictionaries or in King (2003).

2.3. Words which contain two or more morphemes from different languages are marked as mixed-language words, e.g. 'concentrate_io@s:ew' (to concentrate). However, where a word containing at least one English morpheme and at least one Welsh morpheme is included in one or more of the Welsh-language reference dictionaries, it is marked as a Welsh word. For example, the English word 'use' forms the basis of the Welsh word 'iwsio' (to use) but we mark the entire word as Welsh ('iwsio@s:w') because it is included in one of the Welsh-language reference dictionaries.

2.4. The language marker @s:u marks words that occur in the lexicon of both languages, (as determined by the Welsh-language reference dictionaries for Welsh or by the OED for English), that are pronounced in a way that is possible both in Welsh and in English, e.g. ['ʌŋkl] / ['əŋkl] ('uncle' in English or 'yncl' in Welsh) or [mat] ('mat' in both languages). @s:u also marks a specified list of interjections and interactional markers, e.g. ah, ahhah, aw, er, ey, hmm, ho, mmm, mmhm, oh, ooh, ow, ugh, um. Other interjections and interactional markers are assigned language markers according to their inclusion or not in the reference dictionaries. For example, 'ych' (a marker of disgust equivalent to 'yuk' in English) is marked @s:w as it is only found in the Welsh-language reference dictionaries.

2.5. Where a lexeme could belong to both languages, but its pronunciation in a specific occurrence belongs unambiguously to one language only, it will be marked @s:w or @s:e (and written in the respective orthography) according to the pronunciation, e.g. ‘problem@s:w’ for the specifically (north-west) Welsh pronunciation of the second vowel as [a], but ‘problem@s:u’ where the second vowel is [ə] or [ɛ], which is possible in both English and Welsh; ‘toast@s:e’ where the word is pronounced with [əʊ] / [oʊ] as in English only, and ‘tost@s:w’ where it is pronounced with [ɒ] as in southern Welsh, but ‘toast@s:u’ where the word is pronounced with [o:] as in northern Welsh or some varieties of welsh English.

2.6. Proper names and titles are marked as undetermined unless there are alternatives in each language in general use, e.g. Elton_John@s:u, One_Flew_Over_the_Cuckoo’s_Nest@s:u, Hong_Kong@s:u, Tebot_Piws@s:u (a Welsh-language pop group, literally meaning ‘purple teapot’) but Cardiff@s:e, Caerdydd@s:w (the Welsh word for ‘Cardiff’).

2.7. According to GPC, the -s plural ending is an established loan in the Welsh lexicon. Any plural formed with the -s ending is assigned the language marker of the previous morpheme. For example, ‘pregethwrs@s:w’ from ‘pregethwr@s:w’ (preacher), ‘dolphins@s:u’ from ‘dolphin@s:u’ and ‘dogs@s:e’ from ‘dog@s:w’.

2.8. In multi-word phrases, each word is tagged separately, regardless of the phrase’s internal syntax. For example, in ‘traffic@s:u lights@s:e’ ‘traffic’ is coded as undetermined, although the syntax of the whole phrase comes from English.

3. Orthography

3.1. Words marked as English are transcribed in standard English orthography, including contractions, such as ‘isn’t’. Some non-standard spellings for colloquial forms such as ‘gonna’ are used.

3.2. Words whose language is undetermined are transcribed in English rather than in Welsh orthography, e.g. ‘acid@s:u’ rather than ‘asid@s:u’. This is in order to make the corpus more accessible to non-Welsh-speakers who might use the data.

3.3. When words marked as English or undetermined are mutated (where the sound of an initial consonant is changed depending on the grammatical context, see for example King 1993:14-20), the initial (mutated) sound is written in Welsh orthography and the rest in English, e.g. ei@s:w firthday@s:e = his birthday; ei@s:w goat@s:u = his coat . In the case of words that begin with ‘qu’ in English but that are mutated in the data, the mutated sound and the following [w] are written in Welsh orthography, e.g. question (unmutated), gwestion (soft mutation), chwestion (aspirate mutation), nghwestion (nasal mutation).

3.4. Words marked as Welsh are transcribed in Welsh orthography. We have not represented regional variation in the transcripts, except in cases which have orthographic

representation in the Welsh-language reference dictionaries or in King (2003).

There are some cases where we differ from the standard orthography:

i. We transcribe some non-standard verb-noun suffixes, e.g. '-ian' in 'swnian' (to grumble) rather than '-io' in the standard form 'swnio'.

ii. We represent non-standard usage of inflected prepositions. Agreement markers for person and number show considerable variation in the spoken language. Thus one may, for example, find several forms for 'to you' (plural/respect form), such as 'wrthoch chi' (the variant found in King 2003), 'wrthych (chi)' (more formal variant, e.g. prescribed in Thomas 1996) as well as 'wrthach chi' (more colloquial, northern variant). The orthography used in transcripts is based on pronunciation (note that the Welsh orthographic system has a fairly regular relationship between sound and speech, so representing sound variation is possible in Welsh in a way that is not in English).

iii. Northern second person singular verb and preposition endings not usually represented in writing are transcribed as '-a' where they are followed by the pronoun 'chdi', e.g. oedda chdi (you were), arna chdi (on you). Where they occur in isolation, they are transcribed as '-achd', e.g. oeddachd (you were/weren't you), arnachd (on you).

iv. We do not represent morpheme-final [v] when it is not pronounced. For example, [pɛntrɛ] (village) is written 'pentre' in the transcripts rather than 'pentref' (as the word is represented in the Welsh-language reference dictionaries).

v. Morpheme-initial /r/ is transcribed as 'r' even when the standard orthography prescribes 'rh' (pronounced [r̥]) as [r̥] is often absent from speakers' phonological systems. 'rh' is only transcribed where [r̥] is clearly discernible.

vi. We have represented mutation (sound change to initial consonants) or its absence without following prescriptive rules. Thus 'in Cardiff' may be transcribed 'yn Caerdydd' and 'yn Gaerdydd' as well as the standard form 'yng Nghaerdydd', according to what is heard. We have also transcribed the aspirate mutation of /m/ and /n/ after the 3rd singular feminine possessive adjective common in northern varieties, e.g. 'ei mham' (her mother), rather than standard 'ei mam'.

vii. We list below the phrases described in C.1.3 v which we transcribe using underscore to link the individual words.

| **Our transcription** | **Standard** | **English translation** |
|---|---|---|
| ar_draws | ar draws | across |
| ar_goll | ar goll | lost |
| ar_gyfer | ar gyfer | for |
| ar_ôl | ar ôl | after |

| Our transcription | Standard | English translation |
|---|---|---|
| cyn_belled | cyn belled | as far |
| dim_byd | dim byd | nothing |
| ein_gilydd, eich_gilydd, ei_gilydd | ein gilydd, eich gilydd, ei gilydd | each other (1st, 2nd and 3rd person plural) |
| er_mwyn | er mwyn | for |
| ers_talwm | ers talwm | in the past, long ago |
| i_fewn | i fewn | in(to) |
| i_ffwrdd | i ffwrdd | away |
| i_fyny | i fyny | up |
| i_gyd | i gyd | all |
| i_lawr | i lawr | down |
| i_mewn | i mewn | in(to) |
| naill_ai | naill ai | either |
| o_gwbl | o gwbl | at all |
| o_gwmpas | o gwmpas | around |
| oddi_ar | oddi ar | off |
| oddi_wrth | oddi wrth | from |
| oni_bai | oni bai | unless |
| pob_dim | pob dim | everything |
| ta_waeth | ’ta waeth | anyway |
| un_ai | un ai | either |
| wrth_gwrs | wrth gwrs | of course |
| yn_erbyn | yn erbyn | against |
| yn_ôl | yn ôl | back |
| yn_ystod | yn ystod | during |

viii. We list below some colloquial forms which are not represented in the Welsh-language reference dictionaries but which we have transcribed as indicated:

| Our spelling | Standard | Meaning | Comments |
|---|---|---|---|
| (r)hein, (r)hain, (r)heiny etc | rhein, rhain, rheiny etc. | these, those etc. | pronounced with initial [h] |
| byswn i, bysa chdi etc. | baswn i, baset ti etc. | I would, you would etc. | very common in northern varieties |
| cynna fi, cynna chdi etc. | | before me, before you etc. | preposition inflected in northern varieties |
| dylen i, bydden i etc. | dylwn i, byddwn i etc. | I should, I would etc. | common in southern varieties |
| gosa | | unless | heard in north-western varieties |

| Our spelling | Standard | Meaning | Comments |
|---|---|---|---|
| m | ’m | my | usually connected by apostrophe to a preceding vowel |
| mag | mae | | 3rd singular present form of ‘bod’ (to be) heard in south-western varieties |
| molchi | ymolchi | wash oneself | mutates to ‘folchi’ |
| mynedd | amynedd | patience | mutates to ‘fynedd’ |
| na i etc. | a i etc. | I will go etc. | heard in the Caernarfon area |
| nunman | unman | nowhere | widespread |
| oedd nhw, wneith nhw etc. | oedden nhw, wnan nhw etc. | they were, they will etc. | 3rd person singular verb forms used with plural pronouns |
| penwsnos | penwythnos | weekend | GPC has an entry for ‘wsnos’ |
| tes i (ddi)m | es i’m | I didn’t go | some northern varieties |
| w | ’w | his/her/ their | usually appears with apostrophe, e.g. ‘i’w’, but we transcribe as ‘i w’ (to his/her/its) |
| wannwyl | | dear Lord | a contraction of ‘Duw annwyl’ |
| whi | hwyaid | ducks | heard in some southern varieties |
| y fi | rwy i | I am | southern Welsh |

## B GLOSS TIER

1. Principles

1.1. Each word (see C.1.2 and C.1.3) in the main tier is given a gloss in the gloss tier (%gls). Non-words (see C.1.5) are not glossed, with the exception of ‘xx’ and ‘xxx’, which are represented by the same characters in the gloss.

1.2. With the exception of proper names (see below), all words are glossed with the closest English-language equivalent (in lower case). In Welsh or mixed-language words, certain morphological information is included in the gloss (in upper case, see D.2.1). For example:

wasn’t@s:e : wasn’t

soup@s:u : soup
hefyd@s:w : also
recharge_io@s:ew : recharge.NONFIN

Some words marked as Welsh are glossed only with morphological information, such as 'POSS.2S' for the 2nd singular possessive adjective 'dy'.

Proper names (including names of books, films, organisations etc.) marked as English or undetermined are glossed as they appear in the main tier. For example, 'Hong_Kong@s:u' is glossed as 'Hong_Kong', 'Cardiff@s:e' is glossed as 'Cardiff' and 'Tebot_Piws@s:u' is glossed as 'Tebot_Piws'. However, proper names marked as Welsh are glossed with their English-language equivalents. For example, 'Caerdydd@s:w' is glossed as 'Cardiff'.

1.3. Lexical information always precedes morphological information in the gloss. A full stop '.' is used to separate morphological information from lexical information (e.g. go.NONFIN) and also to separate morphological information (e.g. PRON.3S).

1.4. The underscore is used on the gloss tier to connect more than one lexical item in a gloss, where the English translation of a single Welsh word involves more than one word. For example, 'neithiwr' is glossed as 'last_night' .

2. Specific glosses

2.1. The following glosses are used for morphological information:

| **Gloss** | **Use** |
|---|---|
| 1,2,3 | 1st, 2nd, 3rd person |
| CONDIT | conditional/habitual past |
| DET | determiner |
| F | feminine |
| FUT | future/habitual present (verb 'bod' (to be) only) |
| IM | interactional marker/exclamation, e.g. 'um', 'oh' |
| IMP | imperfect (verb 'bod' (to be) only) |
| IMPER | imperative |
| IMPERSONAL | impersonal |
| INT | interrogative |
| M | masculine |
| NEG | negative |
| NONFIN | nonfinite |
| NONPAST | nonpast tense (used for present/habitual/future) |
| PL | plural |
| PAST | past tense |
| POSS | possessive |
| POSSD | possessed |
| PRES | present tense (verb 'bod' (to be) only) |

| Gloss | Use |
|---|---|
| PRON | pronoun |
| PRT | particle |
| REL | relative |
| S | singular |
| SUBJ | subjunctive |

2.2. Gender-specific adjectives in Welsh are not marked for gender in the gloss. For example, ‘gwyn’ (used to modify masculine nouns) and ‘wen’ (used to modify feminine nouns) are both glossed as ‘white’.

2.3. Numerals are glossed for gender where appropriate. For example, ‘dau’ and ‘dwy’ are glossed as ‘two.M’ and ‘two.F’ respectively.

2.4. Welsh collective nouns are glossed by the English plural. For example, ‘moch’ (singular collective noun indicating ‘pigs’) will have the gloss ‘pigs’.

2.5. In third person singular possessive constructions, the gender of the possessor is marked only where there is positive evidence of that gender (i.e. either when the possessed noun is mutated, or when a gender-specific pronoun follows the possessed noun, specifically referring to the possessor). The gender is marked on the possessive adjective. For example:

‘her mother’
ei mam : POSS.3S mother
ei mham: POSS.3SF mother
ei mam hi : POSS.3SF mother PRON.3SF

‘his mother’
ei fam : POSS.3SM mother
ei fam e : POSS.3SM mother PRON.3SM
ei mam e : POSS.3SM mother PRON.3SM

The above applies also to possessive constructions involving non-finite verbs preceded by ‘ei’. For example:

‘he was born’
gaeth (e) ei eni: get.3S.PAST (PRON.3SM) POSS.3SM bear.NONFIN

‘he/she was shot’
gaeth ei saethu: get.3S.PAST POSS.3S shoot.NONFIN

2.6. When a possessive construction in the first person singular is marked only by mutation of the noun, the possessed noun, in the gloss, is followed by ‘.POSSD.1S’. For example, ‘nhad’ (my father) is glossed as ‘father.POSSD.1S’ . Note that this gloss is used only if

there is no possessive adjective preceding or pronoun following the possessed noun (‘fy nhad’ or ‘nhad i’ are glossed ‘POSS.1S father’ and ‘father PRON.1S’ respectively, and ‘fy nhad i’ is glossed ‘POSS.1S father PRON.1S’).

## C TAGS

1. There are certain phrases used in Welsh, usually at the end of an utterance, but also possible mid-utterance, which are used discursively to engage with the listener, to gauge whether he/she agrees, understands etc. (although the listener is seldom required to reply). We term these ‘tags’. Tags can be agreeing (i.e. they include a verb form that agrees in person, number and tense with the finite verb in the main clause) or they can be non-agreeing. Both kinds are particularly problematic in transcription, as they are seldom seen in the written language and therefore there are no fixed conventions for their spelling. They are also often highly contracted in speech and can be problematic for glossing.

2. The following is an incomplete list of agreeing tags that may occur, which serves as a pattern for other agreeing tags (with different verbs, tenses and persons). The table gives the tag as is represented by us in the main tier, and its gloss.

| **Main tier** | **Gloss** |
|---|---|
| byddaf | be.1S.FUT |
| na fyddaf | NEG be.1S.FUT |
| yn_byddaf | be.1S.FUT.NEG |
| medri | can.2S.NONPAST |
| na fedri | NEG can.2S.NONPAST |
| yn_medri | can.2S.NONPAST.NEG |
| dylai | should.3S.CONDIT |
| na ddylai | NEG should.3S.CONDIT |
| yn_dylai | should.3S.CONDIT.NEG |
| ydy, yndy | be.3S.PRES |
| nag ydy, nac (y)dy, na(g) (y)dy etc. | NEG be.3S.PRES |
| yn_dydy, yn_tydy, dydy, tydy | be.3S.PRES.NEG |
| oes e | be.3S.PRES there |
| nag oes e | NEG be.3S.PRES there |
| yn_does e, does e | be.3S.PRES.NEG there |

3. Here is also a list of common non-agreeing tags with their spellings and their glosses. Note that not all occurrences of these words or phrases in the transcripts are tags.

| **Main tier** | **Gloss** |
|---|---|
| felly, (fe)lly | thus |
| wsti, sti | know.2S |
| wchi, (w)chi | know.2PL |
| yli, (y)li | see.2S.IMPER |
| ylwch, (y)lwch | see.2PL.IMPER |
| yn_de, de | TAG |

| Main tier | Gloss |
|---|---|
| yn_do, do | yes |
| yn_dyfe, dyfe | PRT.INT.NEG |
| chimod | know.2PL |
| chwel | see.2PL |
| deud | say.2S.IMPER |
| deuda | say.2S.IMPER |
| deudwch, (deu)dwch | say.2PL.IMPER |
| dywedwch | say.2PL.IMPER |
| dofe | yes |
| dywed, dywad, dŵad | say.2S.IMPER |
| fel | like |
| gwed | say.2S.IMPER |
| iawn | right |
| na | no |
| naci | no |
| naddo | no |
| nag yfe | NEG PRT.INT |
| nage | no |
| ti gweld, ti weld | PRON.2S see.NONFIN |
| ti (y)n gweld | PRON.2S PRT see.NONFIN |
| timod | know.2S |
| twel | see.2S |
| ie,ia | yes |
| yfe | PRT.INT |