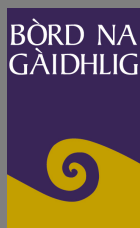


Tha <V-p> dùil <Ncsfn> ri <Sp> tuilleadh <Ncsmd>
 phlèanaichean <Ncpmg> an-diugh <Rt> fhathast
 <Rt> aig <Sp> port-adhair <Ncsmd> Istanbul <Nt> le
 <Sp> comhair <Ncsfd> dhan <Sp> na <Tdpf> mìltean
 <Ncpfd> a <Q-r> tha <V-p> gun <Sp> dachaidh
 <Ncsfd> anns <Sp> an <Tdsf> Tuirc <Nt> . <Fe> A
 <Sp> bharrachd <Ncsfd> air <Sp> teantaichean
 <Ncpmd> , <Fi> bidh <V-f> iad <Pp3p> a' <Sa>
 giulain <Nv> biadh <Ncsmg*> agus <Cc>
 cungaidhean <Ncpmg*> leighis <Ncsmg> . <Fe>

Scottish Gaelic Part-of-Speech Annotation Guidelines

Susanna Naismith and William Lamb

A manual for interpreting and using the Scottish Gaelic PAROLE-type tagset
 Revised 7 April 2020



THE UNIVERSITY
of EDINBURGH

University of Edinburgh, Celtic and Scottish Studies

Table of Contents

REVISIONS	4
1. INTRODUCTION	6
2 LIST OF PARTS OF SPEECH WITH CORRESPONDING TAG	8
2.1 NOUN (N)	8
a) Common Nouns	8
b) Names	8
c) Toponyms, verbal nouns and fossilised lexemes	8
2.2 VERB (V)	9
a) Present	9
b) Past	9
2.3 ADJECTIVE (A)	10
a) Qualificator (attributive)	10
b) Predicative	10
c) Preposed and verbal	10
2.4 PRONOUN (P)	11
a) Personal	11
b) Prepositional	11
c) Numerical, demonstrative, reflexive	11
2.5 DETERMINER (D)	12
a) Possessive	12
b) Demonstrative	12
c) Quantifier	12
2.6 ARTICLE (T)	12
2.7 ADVERB (R)	12
2.8 ADPOSITION (S)	13
a) Preposition	13
b) Aspectual	13
2.9 CONJUNCTION (C)	14
2.10 NUMERALS (M)	14
2.11 INTERJECTION (I)	14
2.12 UNIQUE CLASS MEMBERSHIP	14
2.13 RESIDUALS (X)	15
a) Spoken	15
b) Foreign	15
c) Other	15
2.14 PUNCTUATION	15
2.15 ABBREVIATION (Y)	16
2.16 COPULA (W)	16
a) Present/ Future	16
b) Past/ Conditional	17
2.17 VERBAL PARTICLE (Q)	17
a) Negative	17
b) Interrogative/ affirmative	17

3. CONFUSING PARTS OF SPEECH19

Adjective or adverb: <Aq> or <R>	19
Demonstratives: <Dd> or <Pd>.....	19
Numerals: <Mc> or <Pn>	19
Adverb or Interjection: <Rg> or <I>	20
Adverb or Common Noun: <Rg> or <Nc>.....	20
Adverb of Time or Common Noun: <Rt> or <Nc>	20
Aspectual Adposition or Preposition: <Sa> or <Sp>	20
Preposition or Possessive Determiner: <Sp> or <Spp3sm>/<Spp3sf>	21
Imperative Verb or Subordinate Conjunction: <Vm-2s> or <Cs>	21
Hybrid English-Gaelic Words: <Xfe> or <Nc>	21
Abbreviation or Foreign Word: <Y> or <Xfe>	21

4. PROBLEMATIC WORDS AND COLLOCATIONS22

<i>ach</i>	22
<i>a dh'aindeoin</i>	22
<i>air choreigin</i>	22
<i>air falbh</i>	22
<i>airson</i>	22
<i>a-nochd</i>	22
<i>aon</i>	22
<i>ceart</i>	22
<i>chionn</i>	23
<i>co-dhiù</i>	23
<i>diofar</i>	23
<i>eile</i>	23
<i>fa-near/ fainear</i>	23
<i>fad</i>	23
<i>far</i>	23
<i>fodha</i>	24
<i>gach</i>	24
<i>gu</i>	24
<i>ma/mu</i>	24
<i>rè</i>	24
<i>reimhe/ roimhe</i>	24
<i>ri linn</i>	25
<i>rudeigin</i>	25
<i>seach</i>	25
<i>thoireadh/ (a) thoradh</i>	25
<i>timcheall</i>	25
<i>uile</i>	26

5. FUSED FORMS27

5.1 Nouns.....	27
5.2 Prepositions	27
5.3 Verbal particles	27

5.4 Conjunctions	27
5.5 Question words	28
5.6 Adverbs	28
6 TAGGING CONVENTIONS.....	29
6.1 Part of speech and syntactic function	29
6.2 Asterisks.....	29
6.3 Capitalised words	29
7. DEFAULT TAGSET WITH CORRESPONDING PART OF SPEECH	30
8. SIMPLIFIED TAGSET WITH CORRESPONDING PART OF SPEECH.....	37

REVISIONS

7 April 2020

- Removed hyphens in the verbal tags of the Simplified Tagset (e.g. V-f > Vf)
- In the new release of ARCOSG, tags use the Brown format ('/') and multi-word expressions that receive a single tag are joined with underscores ('_'), e.g. *Dùn_Èideann/Nt*

19 Dec 2014

- Quantifiers <Dq> reinstated as a category in the full (N=246) and short (N= 41) tagsets

1. INTRODUCTION

This manual provides guidelines and examples on annotating Gaelic texts by part of speech, otherwise known as ‘tagging’. Each part of speech has its own corresponding abbreviation or ‘tag’. The first letter of each tag is capitalised and indicates which category the part of speech belongs to. There are seventeen categories in the tagset altogether.

These categories and the letter corresponding to each are as follows:

1. Noun (N), 2. Verb (V), 3. Adjective (A), 4. Pronoun (P), 5. Determiner (D), 6. Article (T), 7. Adverb (R), 8. Adposition (S), 9. Conjunction (C), 10. Numerals (M), 11. Interjection (I), 12. Unique Class Membership (U), 13. Residuals (X), 14. Punctuation (F), 15. Abbreviation (Y), 16. Copula (W), 17. Verbal Particle (Q).

Section 2 contains a list of the parts of speech along with their corresponding tags and definitions. This section allows you to find the appropriate tag for a part of speech according to the category to which it belongs. Section 3 lists all of the tags in alphabetical order with their corresponding part of speech. Sections 4 and 5 give examples of problematic forms. Section 4 provides guidelines on how to tag parts of speech that are easily confused while Section 5 focuses on specific problematic words and collocations. Section 6 discusses examples of concatenated tags for fused word forms and Section 7 provides further information on tagging conventions.

The tags referred to in this manual come from the default tagset of 246 tags (see Section 7), which is based upon the PAROLE tagset for Irish (see Uí Dhonnchadha 2009: Appendix). A simplified tagset, with 41 tags, has been developed for basic applications, and this can be found in Section 8. The two tagsets are incorporated in an automatic part-of-speech tagger for Scottish Gaelic, which is available at the following address: <https://www.wiki.ed.ac.uk/display/gaelicpostagger/Software+releases>.

For further information

Danso, Samuel, and William Lamb. 2014. ‘Developing an automatic part-of-speech tagger for Scottish Gaelic’, *Proceedings of the 1st Celtic Language Technology Workshop*: 1-5.

Uí Dhonnchadha, Elaine. 2009. ‘Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar’, PhD, Dublin City University.

ACKNOWLEDGEMENTS

We would like to acknowledge the generous funding that we received for this project from The Carnegie Trust for the Universities of Scotland and Bòrd na Gàidhlig. *Ceud mile taing dhuibh*. We would also like to acknowledge the support and contributions of the project’s collaborators: Mr Stephen Barrett (University of Glasgow), Prof Rob Dunbar (University of Edinburgh), Prof Mirella Lapata (University of Edinburgh), Mr Alistair Lawson (Edinburgh Napier University), Prof Colm Ó Baoill (University of Aberdeen), Dr Mark McConville (University of Glasgow), Prof Wilson McLeod (University of Edinburgh),

Dr Caoimhín Ó Donnaíle (Sabhal Mòr Ostaig), Prof Roibeard Ó Maolalaigh (University of Glasgow), Prof Boyd Robertson (Sabhal Mòr Ostaig) and Dr Moray Watson (University of Aberdeen). We are also indebted to our unofficial collaborators, Mr Michael Bauer and Dr Elaine Uí Dhonnchadha, for much advice and help. Finally, a special word of thanks to Dr Sammy Danso for developing the project software, and to Dr Sharon Arbuthnot, for her help in devising the tagset and re-tagging the corpus.

2 LIST OF PARTS OF SPEECH WITH CORRESPONDING TAG

2.1 NOUN (N)

a) Common Nouns

Common nouns are tagged according to number, gender and case in that order.

EXAMPLE: *tha dà bhròig <Ncdfn> agamsa*

The first letter 'N' stands for 'noun'. The second letter indicates the type of noun (e.g. c=common), the third letter indicates number (d=dual, p=plural and s=singular), the fourth letter indicates gender (f=feminine, m=male) and the fifth letter indicates case (d=dative, g=genitive, n=nominative and v=vocative). Thus, *bhròig* is a common feminine noun, with dual number, in the nominative case.

If the noun is emphasised, the contrast is indicated with a sixth letter (e=emphatic).

EXAMPLE: *mo thrì bhràithrean-sa <Ncpmne>*

N.B. Dual nouns are tagged only when expressed with the slender ending. Genitive dual nouns are normally identical to the genitive singular. Therefore, there is no tag for genitive dual nouns. This is a departure from our normal practice, whereby we tag nouns according to their syntactic function, rather than their form.

b) Names

Surnames are tagged simply as <Nn>. For first names, gender and case are indicated.

EXAMPLE: *Dòmhnall <Nn-mn> Eachainn <Nn-mg>*

Thus, *Dòmhnall* is a masculine noun in the nominative case and *Eachainn* is a masculine noun in the genitive case.

Due to their complexity, and the prevalence of foreign elements, we have omitted gender and case with surnames. Both Gaelic and English surnames are tagged similarly (e.g. *MacDhòmhnaiill* <Nn>; *Black* <Nn>). The effective tagging of surnames in Gaelic texts would benefit from further study.

c) Toponyms, verbal nouns and fossilised lexemes

Toponyms are tagged <Nt>, whether in Gaelic or in English.

EXAMPLE: *Beinn na Faoghla <Nt>*

Thus, toponyms are not broken down into their constituents but appear as one token.

Verbal nouns are tagged <Nv>. Emphatic forms are tagged <Nv---e>, e.g. *leughadh-ne*.

<Nf> indicates fossilised lexemes such as *am measg*; *a rèir*, *air feadh*. In most cases, these are nouns, historically. Overall, they are non-declinable, e.g. *cò eile*. Emphatic forms are tagged <Nf---e>, e.g. *air mo shon-sa*.

2.2 VERB (V)

a) Present

Verbs in the present tense are tagged <V-p> (p=present).

Dependent forms are tagged <V-p--d> (d=dependent), e.g. *chan eil*.

Impersonal or passive forms are tagged <V-p0> (0=passive), e.g. *thathar ag innse dhuinn*.

Dependent passive forms are tagged <V-p0-d> (d=dependent), e.g. *sgoiltean far a bheilear a' bruidhinn Gàidhlig*.

b) Past

Verbs in the past tense are tagged <V-s> (s=past).

Dependent forms are tagged <V-s--d> (d=dependent), e.g. *cha robh*.

Impersonal or passive forms are tagged <V-s0> (0=passive), e.g. *bhathar a' reic*.

Dependent passive forms are tagged <V-s0-d> (d=dependent) e.g. *chan fhacas*.

c) Future/ present habitual

Verbs in the future or present habitual tense are tagged <V-f> (f=future).

Dependent forms are tagged <V-f--d> (d=dependent), e.g. *chan fhaod*.

Impersonal or passive forms are tagged <V-f0> (0=passive), e.g. *Feumar a ràdh*.

Dependent passive forms are tagged <V-f0-d> (d=dependent), e.g. *cha ghabhar*.

Relative forms are tagged <V-f--r> (r=relative), e.g. *dh'fheumas*.

d) Conditional/ past habitual

Verbs in the conditional or past habitual tense are tagged <V-h> (h=past habitual).

Dependent forms are tagged <V-h--d> (d=dependent), e.g. *cha leigeadh*.

Impersonal or passive forms are tagged <V-h0> (0=passive), e.g. *ma dh'fhaodte*.

Dependent passive forms are tagged <V-h0-d> (d=dependent), e.g. *feuch an tillte e*.

Person and number are indicated for the first person singular and first person plural.

EXAMPLE: *bhithinn* <V-h1s> (1= first person and s=singular)

bhitheamaid <V-h1p> (1=first person and p=plural)

Dependent forms are tagged <V-h1sd> (d=dependent), e.g. *chan fhaicinn* or <V-h1pd>, e.g. *far am biomaid a' coinneachadh*.

Emphatic forms are tagged in the following way:

bhithinn-sa <V-h1s-e> (e=emphatic)

nam bithinn-sa <V-h1sde> (d=dependent, e=emphatic)

e) Imperative

Verbs in the imperative are tagged according to person and number.

Thus, the second person singular is tagged <Vm-2s> (m=imperative, 2= second person and s=singular), e.g. *slaod a-steach*.

The third person is identical in the singular and the plural and is therefore tagged <Vm-3> (3=third person), e.g. *na biodh eagal oirbh*.

The first person plural is tagged <Vm-1p> (1=first person and p=plural), e.g. *Cuimhnicheamaid*.

The second person plural is tagged <Vm-2p> (2=second person and p=plural), e.g. *Fuirichibh*.

2.3 ADJECTIVE (A)

a) Qualificator (attributive)

Attributive adjectives which qualify English nouns are tagged according to number: <Aq-s> (s=singular) and <Aq-p> (p=plural).

Attributive adjectives qualifying Gaelic nouns are tagged according to number, gender and case in the same way as nouns (see § 1).

EXAMPLE: *fhuair mi leabhar* <Ncsmn> *ùr* <Aq-smn>

As *leabhar* is tagged <Ncsmn>, its qualifying adjective *ùr* is tagged <Aq-smn> (m=male, n=nominative).

Qualifying words or expressions which do not indicate number, gender or case, e.g. *taobh a-mach*, *duine sam bith*, *an turas mu dheireadh* are tagged simply as <Aq>.

b) Predicative

Predicative adjectives are tagged <Ap>, e.g. *tha e math*.

Comparative or superlative predicative adjectives are tagged <Apc>, e.g. *nas fheàrr*.

Second comparatives are tagged <Aps>, e.g. *b' fheàrrde*.

c) Preposed and verbal

Preposed adjectives such as *deagh* and *seann* are tagged <Ar>.

EXAMPLE: *a h-uile* <Ar> *sian* <Ncsmn>

N.B. *a h-uile* is tagged as a single token.

Verbal adjectives such as *fosgailte*, which are sometimes referred to as past-participles in the literature, are tagged <Av>.

2.4 PRONOUN (P)

a) Personal

Personal pronouns are tagged according to person, number and gender.

EXAMPLES: *mi* <Pp1s> (p=personal, 1=first person, s=singular)

i <Pp3sf> (p=personal, 3=third person, s=singular, f=feminine)

The emphatic forms are tagged in the following way:

mise <Pp1s--e>

ise <Pp3sf-e> (e=emphatic)

Case can also be indicated for rare¹ occasions of nominative pronouns such as:

se <Pp3smn>

siad <Pp3p-n>

b) Prepositional

Prepositional pronouns are also tagged according to person, number and gender.

EXAMPLES: *ann* <Pr3sm> (r=prepositional, 3=third person, s=singular, m=masculine)

bhuainn <Pr1p> (r=prepositional, 1=first person, p=plural)

Emphatic forms are tagged in the following way:

dhàsan <Pr3sm-e>

oirnne <Pr1p--e> (e=emphatic)

c) Numerical, demonstrative, reflexive

Demonstrative pronouns are tagged <Pd>, e.g. *siud*, *sin*, *seo*.

EXAMPLE: *Gheibh* <V-f> *thu* <Pp2s> *sin* <Pd>

Numerical pronouns are tagged <Pn>.

EXAMPLE: *bha* <V-s> *ceithir* <Pn> *againn*

Reflexive pronouns such as *fhèin* and *chèile* are tagged <Px>.

¹ N.B. Not attested in the representative corpus.

2.5 DETERMINER (D)

This category includes possessive pronouns and demonstrative determiners.

a) Possessive

Possessive pronouns are tagged according to person, number and gender.

EXAMPLES: *do* <Dp2s> (p=possessive, 2=second person, s=singular)

ar <Dp1p> (p=possessive, 1= first person, p=plural)

b) Demonstrative

Demonstrative determiners are tagged <Dd>, e.g. *sin*, *seo*, *ud*.

EXAMPLE: *an fheadhainn* <Ncsfn> *ud* <Dd>

c) Quantifier

Two quantifiers are tagged in the corpus, the tokens *gach* and *a h-uile*. These are in complementary distribution with the definite article (similar to possessives).

EXAMPLE: *gach* <Dq> *duine* <Ncsmn>

2.6 ARTICLE (T)

Definite articles preceding English nouns are tagged according to number: <Tds> (d=definite, s=singular) and <Tdp> (d=definite, p=plural). N.B. There is no indefinite article in Gaelic.

Definite articles preceding Gaelic nouns are tagged according to number and gender.

EXAMPLES: *dhèanadh* <V-h> *e* <Pp3sm> *an* <Tdsm> *gnothach* <Ncsmn>

thiodhlaic <V-s> *iad* <Pp3p> *na* <Tdpf> *sligean* <Ncpfn>

If the article precedes a noun tagged in the genitive case this is also indicated:

EXAMPLES: *fad* <Ncsmn> *an* <Tdsmg> *latha* <Ncsmg>

airson <Nf> *nan* <Tdpfg> *lèabagan* <Ncpfg>

2.7 ADVERB (R)

General adverbs are tagged <Rg>.

Spatial adverbs are tagged <Rs>, e.g. *an seo*, *suas*, *a-mach*.

Temporal adverbs are tagged <Rt>, e.g. *an uairsin*, *an-dràsta*, *an-diugh*.

2.8 ADPOSITION (S)

a) Preposition

Prepositions are tagged <Sp> (p=preposition), regardless of whether they are the head of a clause or a noun phrase:

EXAMPLES: CLAUSE *cò <Uq> às <Sp> a <Q-r> bha <V-s> iad <Pp3p> , <Fi>*

NP fear <Ncsmn> de <Sp> na <Tdpm> bàird <Ncsmd>

Prepositions with an article are tagged according to number: <Spa-s> (s=singular) and <Spa-p> (p=plural).

Prepositions with possessive determiners are tagged according to person, number and gender.

EXAMPLES: *bha <V-s> Garaidh <Nn-mn> na <Spp3sm> chadal <Ncsmd> air <Sp> an <Tdsm> ùrlar <Ncsmd>*

bha <V-s> iad <Pp3p> nan <Spp3p> suidhe <Nv>

The first example is in the third person singular and the gender is masculine; it is therefore tagged <Spp3sm> (p=possessive, 3=third person, s=singular, m=male).

The second example is in the third person plural and is therefore tagged <Spp3p> (p=possessive, 3=third person, p=plural).

Prepositions fused with a relative particle are tagged <Spr>.²

Prepositions fused with a verbal particle are tagged <Spv>.

EXAMPLE: *an <Tdsf> dòigh <Ncsfn> san <Spv> deach <V-s--d> i <Pp3sf> a-steach <Rs>*

N.B. Each new line marks the start of a new grammatical clause.

b) Aspectual

Aspectual particles which appear before verbal nouns are tagged <Sa>.

EXAMPLE: *tha <V-p> thu <Pp2s> a' <Sa> tuigsinn <Nv>*

tha <V-p> an <Tdsm> saoghal <Ncsmn> air <Sa> atharrachadh <Nv>

N.B. When the aspectual particle *a* is only indicated by an apostrophe, it is still tagged <Sa>.

² N.B. This tag does not occur in the representative corpus.

EXAMPLE: *tha* <V-p> *mi* <Pp1s> *'* <Sa> *smaointinn* <Nv>

Aspectual particles with possessive determiners are tagged according to person, number and gender.

EXAMPLES: *bha* <V-s> *ise* <Pp3sf-e> *an sin* <Rs> *ga* <Sap3sm> *fheitheamh* <Nv>

bidh <V-f> *mise* <Pp1s--e> *gur* <Sap2p> *fàgail* <Nv>

2.9 CONJUNCTION (C)

Coordinating conjunctions are tagged <Cc>, e.g. *agus*, *ach*, *no*.

Subordinating conjunctions are tagged <Cs>, e.g. *ged*, *mar*, *far*, *nuair*.

Subordinating conjunctions with the copula are tagged <Csw>.

EXAMPLE: *mas* <Csw> *e* <Pp3sm>

2.10 NUMERALS (M)

Cardinal numerals are tagged <Mc>, e.g. *aon*, *dhà*, etc.

Symbolic numerals (actual numbers) are tagged <Mn>, e.g. '1, 2, 3 etc'. Symbolic numerals followed by the plural ending *-an* are simply tagged <Mn>, e.g. '1630*an*'; '1970*an*'.

Ordinal numerals are tagged <Mo>. Symbolic numerals followed by the ordinal ending *-mh* are tagged as ordinal numerals <Mo>, e.g. *16mh*.

Roman numerals are tagged <Mr>.

Mathematical operators are tagged <Ms>.

2.11 INTERJECTION (I)

Interjections are tagged <I>, e.g. *àidh*, *obh*, *ò*, *uill*.

2.12 UNIQUE CLASS MEMBERSHIP

The adverbial particle *gu* is tagged <Ua>.

EXAMPLE: *tha* <V-p> *iad* <Pp3p> *gu* <Ua> *math* <Rg> *riatanach*

The comparative and superlative particles *nas* and *as* are tagged <Uc>, e.g. *nas fhearr*.

Old composite verbs and fixed copular multi-word expressions such as *is dòcha*, *bu chòir*, *chan urrainn dhomh* are tagged <Uf>.

The tag <Ug> is used to indicate agreement, e.g. *chaidh na caoraich a rùsgadh*.

Phrasal complementisers such as *na* are tagged <Um>, e.g. *tha e nas fheàrr na am fear eile*.

Morphophonemic particles such as *h-*, *n-*, *t-* and *a seachd* are tagged <Uo>.

Patronyms are tagged <Up>.

EXAMPLE: *Mac <Up> Iain <Nn-mg> 'ic <Up> Sheumais <Nn-mg>*

Question words are tagged <Uq>, e.g. *càite, dè, cuine*. N.B. *co mheud* is tagged as one token.

The vocative particle *a* is tagged <Uv>, e.g. *a Chaluim*.

2.13 RESIDUALS (X)

a) Spoken

In transcripts of spoken word, a speaker or communicator (indicated by a number in brackets, e.g. [1]) is tagged <Xsc> (s=spoken, c=communicator).

An event such as a cough or laugh is tagged <Xsev> (s=spoken, ev=event).

An incomplete word is tagged <Xsi> (s=spoken, i=incomplete).

A phonetic element is tagged <Xsp>.

EXAMPLE: *cuag <Xsp> cuag <Xsp>*

b) Foreign

Foreign words are tagged <Xf> (f=foreign); English words are tagged <Xfe> (e=English).

c) Other

Acronyms are tagged <Xa>.

Dates are tagged <Xd>.

E-mail addresses are tagged <Xe>.

Website addresses are tagged <Xw>.

Unknown or unintelligible parts of speech are tagged <Xx>.

Symbols are tagged <Xy>.

2.14 PUNCTUATION

Full stops are tagged <Fe>.

Commas, colons and semi-colons are tagged <Fi> (i=internal).

Opening quotation marks are tagged <Fq>; closing quotation marks are tagged <Fz>.

Hyphens, underscores, dashes, ellipsis and parentheses are tagged <Fb>.

Question marks are tagged <Fg>; exclamation marks are tagged <Fu>.

2.15 ABBREVIATION (Y)

Abbreviations are tagged <Y>.

EXAMPLE: *an* <Tdsm> *t-* <Uo> *Oll.* <Y> *Urr.* <Y> *Iain* <Nn-mn> *MacAonghais* <Nn>

2.16 COPULA (W)

a) Present/ Future

The present indicative is tagged <Wp-i> (p=present, i=indicative)

EXAMPLES: *'s* <Wp-i> *e* <Pp3sm>

's <Wp-i> *ann* <Pr3sm>

Although it is not standard usage, *'Se* (*Se*) can be found as one word, in which case it is tagged <Wp-i-3> (3=third person pronoun).

Similarly, *'Sann* (*Sann*) can also be found as one word, in which case it is tagged <Wp-i-x> (x=existential *ann*).

The negative present indicative is tagged <Wp-in> (i=indicative, n=negative).

EXAMPLE: *chan* <Wp-in> *e* <Pp3sm>

chan <Wp-in> *ann* <Pr3sm>

Affirmative dependent forms are tagged <Wpdia> (d=dependent, i=indicative, a=affirmative) and negative dependent forms are tagged <Wpdin> (n=negative).

EXAMPLES: *gur* <Wpdia> *e* <Pp3sm>

nach <Wpdin> *ann* <Pr3sm>

Affirmative interrogative forms are tagged <Wpdqa> (d=dependent, q=interrogative, a=affirmative) and negative interrogative forms are tagged <Wpdqn> (n=negative).

EXAMPLES: *an* <Wpdqa> *e* <Pp3sm> ? <Fg>

nach <Wpdqn> *ann* <Pr3sm> ? <Fg>

Relative forms are tagged <Wpr>.

EXAMPLES: *'se* <Wp-i-3> *sin* <Pd> *as* <Wpr> *coireach* <Ap>

a' <Tdsf> *chlach* <Ncsfn> *as* <Wpr> *àirde* <Apc>

dè <Uq> *as* <Wpr> *urrainn* <Uf> *dhasan* <Pr3sm-e> *a* <Ug> *dhèanamh* <Nv>

b) Past/ Conditional

The past/ conditional tense is tagged <Ws>.

EXAMPLE: *bu* <Ws> *mhath* <Uf> *leam* <Pr1s>

Cha <Qn> *b'* <Wsd> *e* <Pp3sm>

2.17 VERBAL PARTICLE (Q)

a) Negative

Negative complementisers are tagged <Qn> (n=negative).

EXAMPLE: *chan* <Qn> *eil* <V-p--d> *mise* <Pp1s--e> *cinnteach* <Ap>

Relative negative complementisers are tagged <Qnr> (n=negative, r=relative).

EXAMPLE: *Carson* <Uq> *nach* <Qnr> *eil* <V-p--d> *e* <Pp3sm> *pòsta* <Av>

Imperative negative complementisers are tagged <Qnm> (n=negative, m=imperative).

EXAMPLE: *Na* <Qnm> *tog* <Vm-2s> *ceàrr* <Rg> *mi* <Pp1s> *idir* <Rg>

b) Interrogative/ affirmative

Interrogative and dependent clause markers are tagged <Qq>.

EXAMPLE: *an* <Qq> *robh* <V-s--d> *esan* <Pp3sm-e> *an seo* <Rs> ? <Fg>

Affirmative complementisers are tagged <Qa> (a=affirmative).

EXAMPLE: *saoilidh* <V-f> *mi* <Pp1s> *gu* <Qa> *bheil* <V-p--d> *cunnart* <Ncsmn> *an seo* <Rs> *an-diugh*

Relative verbal particles are tagged <Q-r> (r=relative).

EXAMPLE: *na* <Tdpm> *leabhraichean* <Ncpmd> *a* <Q-r> *bha* <V-s> *esan* <Pp3sm-e> *a'* <Sa> *leughadh* <Nv>

Verbal particles in the past tense are tagged <Q--s> (s=past).

EXAMPLE: *cha* <Qn> *do* <Q--s> *bhruidhinn* <V-s> *e* <Pp3sm> *rium* <Pr1s> *idir* <Rg>

Verbal particles in the subjunctive are tagged <Q-s> (s=subjunctive).

EXAMPLE: *nan* <Q-s> *innseadh* <V-h--d> *duine* <Ncsmn> *dhomh* <Pr1s>

N.B. Although the subjunctive is a traditional label for verbs that are strongly irrealis, modern Gaelic verbal morphology does not present a strong case for it as a distinct verbal category. Therefore the subjunctive case has been excluded from the tagset, with the exception of the subjunctive verbal particle *nam / nan* <Q-s>. Verbs that traditionally would have been considered subjunctive have been classified as past habitual/conditional (see §2d).

3. CONFUSING PARTS OF SPEECH

This section provides guidelines on how to tag parts of speech that can be easily confused.

Adjective or adverb: <Aq> or <R>

Certain adverbs such as *a-mach*, *a-muigh*, *thall*, *an ear*, *an iar* and *mu dheireadh* are normally tagged as either general, spatial or temporal adverbs (<Rg>, <Rs> or <Rt>). However, they can also function as attributive adjectives <Aq>.

EXAMPLE: *taobh* <Ncsmn> *a-mach* <Aq> *à* <Sp> *Inbhir Nis* <Nt>
 taobh <Ncsmn> *thall* <Aq> *na* <Tdsfg> *pàirce* <Ncsfg>
 Air <Sp> *an* <Tdsf> *oidhche* <Ncsfd> *mu dheireadh* <Aq>

Certain adjectives such as *math* and *dearbh* are normally tagged as attributive, predicative or preposed adjectives (<Aq>, <Ap> or <Ar>). However, when they are preceded by the adverbial particle *gu* <Ua>, they are tagged as general adverbs <Rg>.

Demonstratives: <Dd> or <Pd>

Seo, *sin* and *siud* can be tagged either as demonstrative determiners <Dd> or demonstrative pronouns <Pd>.

Seo/sin/siud are tagged as demonstrative pronouns <Pd> when they stand for noun phrases:

EXAMPLE: *chan* <Wpdin> *e* <Pp3sm> *droch* <Ar> *rud* <Ncsmn> *tha*³ <V-p> *sin* <Pd>

Seo/sin/siud can also be used with the definite noun, like adjectives. In this case, they are tagged as demonstrative determiners <Dd>.

EXAMPLE: *tha* <V-p> *am* <Tdsm> *fear* <Ncsmn> *sin* <Dd> *math* <Ap>

Numerals: <Mc> or <Pn>

If the number indicates a date or time, it is tagged as a cardinal numeral <Mc>.

EXAMPLES: *ochd* <Mc> *deug* <Mc> *dà* <Mc> *fhichead* <Mc> *'s* <Cc> *a* <Uo> *naoi* <Mc>

leth-uair <Ncsfn> *an* <Sp> *dèidh* <Nf> *seachd* <Mc>

If the number is followed by a noun, it is tagged as a cardinal numeral <Mc>.

EXAMPLES: *an* <Tdsf> *dà* <Mc> *chailleach* <Ncsfn>

airson <Nf> *dhà* <Mc> *no* <Cc> *tri* <Mc> *oidhcheannan* <Ncpfg>

an <Tdsm> *aon* <Mc> *rud* <Ncsmn>

³ NB: the relative pronoun *a* expected before *tha* has been elided in this example.

eadar <Sp> *an* <Tdsm> *dà* <Mc> *chnoc* <Ncsmn> *tha*⁴ <V-p> *siud* <Pd>
fhuair <V-s> *e* <Pp3sm> *an* <Tdsm> *dà* <Mc> *àite* <Ncsmn>

Numbers are tagged as numerical pronouns <Pn> when they stand for noun phrases:

EXAMPLES: *Chunnaic* <V-s> *sinn* <Pp1p> *ceithir* <Pn> *air* <Sp> *telebhisean* <Ncsmd>

tha <V-p> *a* <Uo> *dhà* <Pn> *eile* <Aq> *ann* <Pr3sm>

tha <V-p> *sibhse* <Pp2p--e> *a'* <Sa> *riochdachadh* <Nv> *a* <Uo> *dhà* <Pn>
dhiubh <Pr3p>

uair <Ncsfn> *no* <Cc> *dhà* <Pn>

tha <V-p> *ceangal* <Ncsmn> *eadar* <Sp> *na* <Tdp> *dhà* <Pn>

Adverb or Interjection: <Rg> or <I>

Expressions such as *a-nise*, *ma-thà* and *dìreach* are tagged as general adverbs <Rg>, not as interjections <I>.

Adverb or Common Noun: <Rg> or <Nc>

Words that are indicative of quantity such as *beagan* and *tòrr* are tagged as common nouns, not as adverbs <Rg>.

Adverb of Time or Common Noun: <Rt> or <Nc>

Days of the week and months of the year are tagged as a common nouns, not temporal adverbs.

EXAMPLE: *an* <Qq> *d'* <Q--s> *fhuair* <V-s> *sibh* <Pp2p> *crìochnachadh* <Ncsmn>
Disathairne <Ncsmn> ? <Fg>

Aspectual Adposition or Preposition: <Sa> or <Sp>

When a preposition such as *ri* or *air* precedes a verbal noun <Nv>, it is tagged as an aspectual adposition <Sa>, not a preposition <Sp>.

EXAMPLES: *Tha* <V-p> *mi* <Pp1s> *a'* <Sa> *coimhead* <Nv> *air* <Sp> *adhart* <Ncsmd> *ri*
<Sa> *bhith* <Nv> *a'* <Sa> *tilleadh* <Nv> *gun* <Sp> *chomataidh* <Ncsmd>

bu <Ws> *chòir* <Uf> *dhuinn* <Pr1p> *a* <Ug> *bhith* <Nv> *air* <Sa> *ar* <Dp1p>
dìon <Nv>

⁴ NB: the relative pronoun *a* has been elided in this example.

Preposition or Possessive Determiner: <Sp> or <Spp3sm>/<Spp3sf>

When prepositions incorporate the third person singular possessive determiner *a*, they are tagged <Spp3sm> or <Spp3sf> not <Sp>. In the following example, *mu* stands for *mu a*.

EXAMPLE: *Carson* <Uq> *nach* <Qn> *cuala* <V-s--d> *sinn* <Pp1p> *mu* <Spp3sf> *deidhinn* <Nf>

Imperative Verb or Subordinate Conjunction: <Vm-2s> or <Cs>

Verbs such as *feuch* and *saoil* are tagged as imperative verbs <Vm-2s> and not as subordinating conjunctions.

EXAMPLE: *feuch* <Vm-2s> *an* <Qq> *tillte* <V-h0-d> *e* <Pp3sm>

Saoil <Vm-2s> *an* <Wpdqa> *ann* <Pr3sm> *ri* <Sp> *fealla-dhà* <Ncsfd>
a <Q-r> *bha* <V-s> *e* <Pp3sm> ? <Fg>

Hybrid English-Gaelic Words: <Xfe> or <Nc>

English words are tagged <Xfe>. Hybrid words (usually English words with a Gaelicised spelling) such as ‘poilitigs’ and ‘heileacoptair’ are tagged as Gaelic nouns if they are listed in *Am Faclair Beag*. N.B. English surnames and place-names are tagged <Nn> and <Nt> respectively, the same as Gaelic surnames and place-names.

Abbreviation or Foreign Word: <Y> or <Xfe>

English abbreviations are tagged as abbreviations <Y> and not as foreign words <Xfe>.

EXAMPLE: *dè* <Uq> *thuir* <V-s> *Mrs* <Y> *MacQueen* <Nn> *riut* <Pr2s> *an-diugh* <Rt> ? <Fg>

4. PROBLEMATIC WORDS AND COLLOCATIONS

This section contains a list of specific problematic words and collocations in alphabetical order.

ach

Ach is usually tagged as a conjunction <Cc>. However, focus particles are tagged as prepositions <Sp>.

EXAMPLE: *Chan* <Qn> *eil* <V-p--d> *ann* <Pr3sm> *ach* <Sp> *duine* <Ncsmn> *foilleil* <Aq-smn>

a dh'aindeoin

Aindeoin is tagged as a fossilised lexeme <Nf>, not as an adverb <Rg>.

EXAMPLE: *a* <Sp> *dh'* <Uo> *aindeoin* <Nf>

air choreigin

Air choreigin is tagged as an attributive adjective <Aq>.

air falbh

Air falbh is tagged as a spatial adverb <Rs>.

airson

When *airson* appears as one word, it is tagged as a fossilised lexeme <Nf>. When it appears as two words, *air* is tagged as a preposition <Sp> and *son* is tagged as a fossilised lexeme <Nf>.

a-nochd

A-nochd is usually tagged as a temporal adverb <Rt>.

EXAMPLE: *gheibh* <V-f> *sinn* <Pp1p> *fois* <Ncsfn> *a-nochd* <Rt> *co-dhiù* <Rg>

However, it can also function as a noun.

EXAMPLE: '*S* <Wp-i> *e* <Pp3sm> *a-nochd* <Nf> *a* <Q-r> *dhèanadh* <V-h> *dragh* <Ncsmn> *dhòmhsa* <Pr1s--e>

aon

When *aon* means 'the same', it is still tagged as a cardinal numeral, not as a preposed adjective.

EXAMPLE: *an* <Tdsm> *aon* <Mc> *rud* <Ncsmn>

ceart

Ceart can be tagged as an attributive, predicative or preposed adjective and may also function as an adverb.

EXAMPLES: *slaod* <Vm-2s> *a-steach* <Rs> *do* <Dp2s> *shèithear* <Ncsmn> *ceart* <Rg> *faisg* <Ap> *air* <Sp> *a'* <Tdsm> *bhòrd* <Ncsmd>

chionn

When *chionn* is preceded by a preposition <Sp>, it is tagged as a fossilised lexeme <Nf>.

EXAMPLE: *o* <Sp> *chionn* <Nf> *deich* <Mc> *bliadhna* <Ncsfg>

When it stands alone at the start of a new clause, it is tagged as a subordinating conjunction <Cs>.

EXAMPLE: *chionn* <Cs> *gun* <Qa> *robh* <V-s--d> *fearann* <Ncsmn> *ann* <Pr3sm>
co-dhiù

co-dhiù

Co-dhiù is usually tagged as a general adverb <Rg>. However, when it is followed by a verb, the copula or a verbal particle, it is tagged as a subordinating conjunction.

EXAMPLE: *Co-dhiubh* <Cs> *a* <Q-r> *thig* <V-f> *a* <Dp3sm> *leithid* <Ncsfn> *gu* <Sp> *buil* <Ncsfd>

diofar

Diofar is tagged as a masculine common noun, not as a preposed adjective.

EXAMPLE: *bhiodh* <V-h> *diofar* <Ncsmn> *sheòrsachan* <Ncpmg> *mun* <Sp> *cuairt* <Ncsfd>

eile

When *eile* is preceded by a noun or pronoun it is tagged as an attributive adjective <Aq>.

EXAMPLE: *bha* <V-s> *feadhainn* <Ncsfn> *eile* <Aq-sfn> *ann* <Pr3sm>

However, if it is preceded by a conjunction or question word, it is tagged as a fossilised lexeme <Nf>.

EXAMPLES: *cò* <Uq> *eile* <Nf>

ceistean <Ncpfd> *annasach* <Aq-pfd> *is* <Cc> *eile* <Ncpfn>

fa-near/ faineair

Fa-near/ faineair is tagged <Ap> as it functions like a predicative adjective.

fad

Fad is most often tagged as a common noun <Ncsmn> but when it is followed by the coordinating conjunction 's, it is tagged as a subordinating conjunction <Cs>.

far

When *far* is followed by a noun, it is tagged as a preposition <Sp>. When it is followed by a dependent clause marker <Qq> or negative complementiser <Qn>, it is tagged as a subordinating conjunction <Cs>.

EXAMPLES: *bha* <V-s> *sinne* <Pp1p--e> *a'* <Sa> *cur* <Nv> *chlèibh* <Ncpfg> *far* <Sp> *an* <Tdsmg> *eathair* <Ncsmg>

siud <Pd> *far* <Cs> *a* <Qq> *bheil* <V-p--d> *iad* <Pp3p>

fodha

Fodha is tagged as a third person singular prepositional pronoun <Pr3sm>, not as a spatial adverb <Rs>.

EXAMPLE: *chunnaic* <V-s> *e* <Pp3sm> *sgoth* <Ncsfn> *iasgaich* <Ncsmg> *a'* <Sa> *dol* <Nv> *fodha* <Pr3sm> *eadar* <Sp> *Ìle* <Nt> *agus* <Cc> *Èirinn* <Nt>

gach

As *gach* is in complementary distribution with the definite article, it

gu

When followed by a verb, *gu* is tagged as an affirmative complementiser <Qa>. When followed by a noun, *gu* is tagged as a preposition <Sp>.

EXAMPLES: *thuir* <V-s> *am* <Tdsm> *Ministear* <Ncsmn> *gu* <Qa> *robh* <V-s--d> *na* <Tdps> *h-* <Uo> *àireamhan* <Ncpfn> *sìos* <Rs> *is* <Cc> *suas* <Rs> *gach* <Rg> *bliadhna* <Ncsfn>

thàinig <V-s> *iad* <Pp3p> *gu* <Sp> *eilean* <Ncsmd>

ma/mu

When followed by a verb, *ma/mu* is tagged as a subordinating conjunction <Cs>. When followed by a noun, a numeral or an adverb, *ma/mu* is tagged as a preposition <Sp>.

EXAMPLES: *ma* <Cs> *gheibh* <V-f> *mi* <Pp1s> *e* <Pp3sm> . <Fe>

Tha <V-p> *e* <Pp3sm> *coltach* <Ap> *nach* <Qn> *eil* <V-p--d> *dad* <Ncsmn> *amharasach* <Aq-smn> *mu* <Sp> 'n <Tdsf> *chùis* <Ncsfd>

rè

Rè is tagged as a common noun <Ncsmn>, not as a preposition.

reimhe/ roimhe

Reimhe/ roimhe is tagged as a prepositional pronoun <Pr3sm>, not as a temporal adverb <Rt>.

EXAMPLE: *an* <Tdsf> *oidhche* <Ncsfn> *roimhe* <Pr3sm>

ri linn

When *ri linn* appears as one token, it is tagged as subordinating conjunction <Cs>. When it appears as two separate tokens, *ri* is tagged as a preposition <Sp> and *linn* is tagged as a noun in the dative case <Ncsmd>.

rudeigin

Rudeigin is normally tagged as a noun but when it precedes a predicative adjective <Ap> it is tagged as an adverb <Rg>.

EXAMPLE: *bheir* <V-f> *mise* <Pp1s--e> *dhut* <Pr2s> *rudeigin* <Ncsmn>

tha <V-p> *e* <Pp3sm> *rudeigin* <Rg> *òg* <Ap>

seach

When *seach* is followed by a noun, it is tagged as a preposition <Sp>. When it is followed by an affirmative complementiser <Qa>, a negative complementiser <Qn> or the copula, it is tagged as a subordinating conjunction <Cs>.

EXAMPLES: *chan* <Qn> *eil* <V-p--d> *tadhal* <Ncsmn> *aig* <Sp> *taobh* <Ncsmd> *seach* <Sp> *taobh* <Ncsmn> *fhathast* <Rt>

seach <Cs> *gu* <Qa> *bheil* <V-p--d> *blas* <Ncsmn> *Uibhist* <Nt> *air* <Sp> *mo* <Dp1s> *Ghàidhlig* <Ncsfd>

thoireadh/ (a) thoradh

When *thoradh* means 'since/ because', it is tagged as a coordinating conjunction <Cc>.

EXAMPLE: *thoradh* <Cc> *chan* <Qn> *fhac'* <V-s--d> *iad* <Pp3p> *gin* <Ncsfn> *dheth* <Pr3sm> *riamh* <Rt>

When it is preceded by the preposition *mar*, it is tagged as a noun.

EXAMPLE: *mar* <Sp> *thoradh* <Ncsmd> *air* <Sp> *a-seo* <Pd>

N.B. *thoireadh* can also be the past habitual/ conditional tense of the verb *thoir*, in which case it would be tagged <V-h>.

timcheall

When *timcheall* is followed by the preposition *air* <Sp> or the prepositional pronoun *air* <Pr3sm>, it is tagged as a predicative adjective <Ap>.

EXAMPLE: *tha* <V-p> *a'* <Tdsf> *ghealach* <Ncsfn> *a'* <Sa> *toirt* <Nv> *timcheall* <Ap> *air* <Sp> *an* <Tdsm> *t-* <Uo> *saoghal* <Ncsmd>

When it is followed by a noun, it is tagged as a fossilised lexeme <Nf>.

EXAMPLE: *tha* <V-p> *an* <Tdsm> *saoghal* <Ncsmn> *a'* <Sa> *toirt* <Nv> *timcheall* <Nf> *na* <Tdsfg> *grèine* <Ncsfg>

Otherwise, it is tagged a spatial adverb <Rs>.

EXAMPLE: *thòisich* <V-s> *am* <Tdsm> *batal* <Ncsmn> *timcheall* <Rs> *mun* <Sp> *an* <Tdsm> *teampall* <Ncsmd>

uile

When *uile* is preceded by a pronoun (personal, prepositional or demonstrative), it is tagged as an adverb <Rg>.

EXAMPLE: *tha* <V-p> *iad* <Pp3p> *uile* <Rg> *air* <Sa> *am* <Dp3p> *fastadh* <Nv>

However, when it is preceded by a noun or article, it is tagged as a fossilised lexeme <Nf>.

EXAMPLE: *mar* <Cs> *is* <Wp-i> *aithne* <Uf> *dha* <Sp> *na* <Tdpf> *h-* <Uo> *uile* <Nf>

5. FUSED FORMS

Allowance has been made for concatenated tags for a small number of fused word forms. This section discusses examples of such cases.

5.1 Nouns

Certain nouns can be fused with a prepositional pronoun. Though they appear as one token, they must be treated within the tag as two discrete words.

EXAMPLES: *Tha* <V-p> *cuimhneam* <Ncsfn+Pr1s>

Chan <Qn> *eil* <V-p--d> *fhiosam* <Ncsmn+Pr1s>

5.2 Prepositions

A preposition can be fused with a verbal particle. In the following example, the preposition *anns* is fused with the relative verbal particle *na*.

EXAMPLE: *cha* <Qn> *robh* <V-s--d> *mòran* <Ncsmn> *brìgh* <Ncsfg> *sna* <Sp+Q-r> *thubhairt* <V-s> *e* <Pp3sm>

A preposition can also be fused with a possessive pronoun. In the following example, the preposition *air* is fused with the possessive pronoun *do* to form *orra*.

EXAMPLE: *Orra* <Sp+Dp2s> *shocair* <Ncsfd> , <Fi> *'ille* <Ncsmv>

5.3 Verbal particles

Verbal particles can be fused together. In the following example, the interrogative verbal particle *an* is fused with the past tense verbal particle *do* to form *na*.

EXAMPLE: *gus* <Cs> *na* <Qq+Q--s> *chuir* <V-s> *e* <Pp3sm> *a-staigh* <Rs> *a h-uile* <Ar> *gin* <Ncsfn> *dhiubh* <Pr3p>

Sometimes the past tense verbal particle is dropped altogether. In the following example, the affirmative complementiser *gun* and the past tense verbal particle *do* have been contracted to *gun*.

EXAMPLES: *gun* <Qa+Q--s> *sgriobh* <V-s> *e* <Pp3sm>

5.4 Conjunctions

A subordinating conjunction can be fused with an interrogative verbal particle. In the following example, the subordinating conjunction *gus* has been fused with the interrogative verbal particle *an* to form *gun*.

EXAMPLE: " <Fq> *gun* <Cs+Qq> *tig* <V-f--d> *sibh* <Pp2p> *a-mach* <Rs> *ann a'* <Sp> *soilleireachd* <Ncsfd> . <Fe> " <Fz>

In the following example, the subordinating conjunction *mur* has been fused with the interrogative verbal particle *am* to form *mum*.

EXAMPLE: *tha* <V-p> *trì* <Mc> *rudan* <Ncpmn> *aige* <Pr3sm> *ri* <Sa> *dhèanamh* <Nv>
mum <Cs+Qq> *faigh* <V-f--d> *e* <Pp3sm> *i* <Pp3sf> *airson* <Nf> *a* <Dp3sf>
pòsadh <Nv>

5.5 Question words

The question word *càite* can be fused with the present dependent verb *bheil*.

EXAMPLE: *càil* <Uq+V-p--d> *thu* <Pp2s> *a'* <Sa> *dol* <Nv> ? <Fg>

5.6 Adverbs

When the adverb *fiù* is fused with the conjunction *'s* (> *fiùs*), it is tagged <Rg+Cc>.

6 TAGGING CONVENTIONS

6.1 Part of speech and syntactic function

In general, parts of speech are tagged according to form than function. This has important implications, especially where nouns are concerned.

When plural nouns appear in their singular form, they are tagged as singular nouns.

EXAMPLES: *thar* <Sp> *nan* <Tdpmg> *còig* <Mc> *là* <Ncsmg>

ceithir <Mc> *bliadhn'* <Ncsfn> *deug* <Mc> *thar* <Sp> *fhichead* <Mc>

When nouns indicate possession, they are tagged simply as nouns.

EXAMPLES: *Ach* <Cc> *dh'* <Uo> *fheumadh* <V-h> *e* <Pp3sm> *cumail* <Nv> *ri* <Sp> *fhacal* <Ncsmd> . <Fe>

Thuir <V-s> *e* <Pp3sm> *ri* <Sp> *mhàthair* <Ncsfd>

NB: In spoken Gaelic, and sometimes written Gaelic, as well, unstressed vowels are elided in certain environments. In some cases, full words are elided. For example, in the example immediately above, it is clear that the third-person, masculine possessive (i.e. *a*) has been elided between *ri* and *fhacal* and also *ri* and *mhàthair*. Elided words have not been inserted or tagged in the corpus.

6.2 Asterisks

Some tokens in the Gold-standard Corpus are marked with an asterisk at the end of their corresponding tags, to indicate that the token is nonconcordant with conservative grammar. This most often occurs when the form of an article, noun or adjective does not correspond to the case it ought to be in. In the following example, *uallach* follows a verbal noun and, in some conservative registers, would be in the genitive case, taking the form *uallaich*.

EXAMPLE: *tha* <V-p> *e* <Pp3sm> *a'* <Sa> *cur* <Nv> *uallach* <Ncsmg*> *orm* <Pr1s> *cà* <Uq> *deach* <V-s--d> *i* <Pp3sf>

6.3 Capitalised words

If a single word or series of words is capitalised because it is used as a title, it should be tagged as a name <Nn>, e.g. *Sradagan* and *Dé a-Nis*?

N.B. Although languages, days of the week and months of the year are capitalised, they are tagged as common nouns, not names.

7. DEFAULT TAGSET WITH CORRESPONDING PART OF SPEECH

This section contains a list of the full list of tags, in alphabetical order, with corresponding parts of speech.

NO	TAG	PART OF SPEECH
1	Ap	Adjective, predicative
2	Apc	Adjective, predicative, comparative or superlative
3	Aps	Adjective, predicative, second comparative
4	Aq	Adjective, qualifier (attributive)
5	Aq-dfd	Adjective, qualifier, dual, feminine, dative
6	Aq-dfn	Adjective, qualifier, dual, feminine, nominative
7	Aq-p	Adjective, qualifier, plural
8	Aq-pfd	Adjective, qualifier, plural, feminine, dative
9	Aq-pfg	Adjective, qualifier, plural, feminine, genitive
10	Aq-pfn	Adjective, qualifier, plural, feminine, nominative
11	Aq-pfv	Adjective, qualifier, plural, feminine, vocative
12	Aq-pmd	Adjective, qualifier, plural, masculine, dative
13	Aq-pmg	Adjective, qualifier, plural, masculine, genitive
14	Aq-pmn	Adjective, qualifier, plural, masculine, nominative
15	Aq-pmv	Adjective, qualifier, plural, masculine, vocative
16	Aq-s	Adjective, qualifier, singular
17	Aq-sfd	Adjective, qualifier, singular, feminine, dative
18	Aq-sfg	Adjective, qualifier, singular, feminine, genitive
19	Aq-sfn	Adjective, qualifier, singular, feminine, nominative
20	Aq-sfv	Adjective, qualifier, singular, feminine, vocative
21	Aq-smd	Adjective, qualifier, singular, masculine, dative
22	Aq-smg	Adjective, qualifier, singular, masculine, genitive
23	Aq-smn	Adjective, qualifier, singular, masculine, nominative
24	Aq-smv	Adjective, qualifier, singular, masculine, vocative
25	Ar	Adjective, preposed
26	Av	Adjective, verbal
27	Cc	Conjunction, coordinate
28	Cs	Conjunction, subordinate
29	Cs+Qq	Fused Form: subordinating conjunction fused with interrogative verbal particle
30	Csw	Conjunction, subordinate with copula
31	Dd	Determiner, demonstrative
32	Dp1p	Determiner, possessive, first person plural
33	Dp2p	Determiner, possessive, second person plural
34	Dp3p	Determiner, possessive, third person plural
35	Dp1s	Determiner, possessive, first person singular
36	Dp2s	Determiner, possessive, second person singular
37	Dp3sf	Determiner, possessive, third person singular, feminine
38	Dp3sm	Determiner, possessive, third person singular, masculine

39	Dq	Determiner, quantifier
40	Fb	Punctuation, hyphen/ underscore/ dash/ ellipsis/ parentheses
41	Fe	Punctuation, sentence final (full stop)
42	Fg	Punctuation, question mark
43	Fi	Punctuation, sentence internal (comma, colon, semi-colon)
44	Fq	Punctuation, quote (opening quotation mark)
45	Fu	Punctuation, exclamation mark
46	Fz	Punctuation, quote (closing quotation mark)
47	I	Interjection
48	Mc	Numeral, cardinal
49	Mn	Numeral, symbolic
50	Mo	Numeral, ordinal
51	Mr	Numeral, roman
52	Ms	Numeral, operator
53	Ncdfd	Noun, common, dual, feminine, dative
54	Ncdfde	Noun, common, dual, feminine, dative, emphatic
55	Ncdfn	Noun, common, dual, feminine, nominative
56	Ncdfne	Noun, common, plural, feminine, nominative, emphatic
57	Ncpfd	Noun, common, plural, feminine, dative
58	Ncpfde	Noun, common, plural, feminine, dative, emphatic
59	Ncpfg	Noun, common, plural, feminine, genitive
60	Ncpfge	Noun, common, plural, feminine, genitive, emphatic
61	Ncpfn	Noun, common, plural, feminine, nominative
62	Ncpfne	Noun, common, plural, feminine, nominative, emphatic
63	Ncpfv	Noun, common, plural, feminine, vocative
64	Ncpmd	Noun, common, plural, masculine, dative
65	Ncpmde	Noun, common, plural, masculine, dative, emphatic
66	Ncpmg	Noun, common, plural, masculine, genitive
67	Ncpmge	Noun, common, plural, masculine, genitive, emphatic
68	Ncpmn	Noun, common, plural, masculine, nominative
69	Ncpmne	Noun, common, plural, masculine, nominative, emphatic
70	Ncpmv	Noun, common, plural, masculine, vocative
71	Ncsfd	Noun, common, singular, feminine, dative
72	Ncsfde	Noun, common, singular, feminine, dative, emphatic
73	Ncsfg	Noun, common, singular, feminine, genitive
74	Ncsfge	Noun, common, singular, feminine, genitive, emphatic
75	Ncsfn	Noun, common, singular, feminine, nominative
76	Ncsfn+Pr1s	Fused Form: singular feminine noun fused with first person singular prepositional pronoun
77	Ncsfne	Noun, common, singular, feminine, nominative, emphatic
78	Ncsfv	Noun, common, singular, feminine, vocative
79	Ncsmd	Noun, common, singular, masculine, dative
80	Ncsmdde	Noun, common, singular, masculine, dative, emphatic
81	Ncsmg	Noun, common, singular, masculine, genitive
82	Ncsmge	Noun, common, singular, masculine, genitive, emphatic

83	Ncsmn	Noun, common, singular, masculine, nominative
84	Ncsmn+Pr1s	Fused Form: singular masculine noun fused with first person singular prepositional pronoun
85	Ncsmne	Noun, common, singular, masculine, nominative, emphatic
86	Ncsmv	Noun, common, singular, masculine, vocative
87	Nf	Noun, fossilised lexeme
88	Nf---e	Noun, fossilised lexeme, emphatic
89	Nn	Noun, name (surname)
90	Nn-fd	Noun, name, feminine, dative
91	Nn-fg	Noun, name, feminine, genitive
92	Nn-fn	Noun, name, feminine, nominative
93	Nn-fv	Noun, name, feminine, vocative
94	Nn-md	Noun, name, masculine, dative
95	Nn-mg	Noun, name, masculine, genitive
96	Nn-mn	Noun, name, masculine, nominative
97	Nn-mv	Noun, name, masculine, vocative
98	Nt	Noun, toponym
99	Nv	Noun, verbal
100	Nv---e	Noun, verbal, emphatic
101	Pd	Pronoun, demonstrative
102	Pn	Pronoun, numerical
103	Pp1p	Pronoun, personal, first person plural
104	Pp1p--e	Pronoun, personal, first person plural, emphatic
105	Pp2p	Pronoun, personal, second person plural
106	Pp2p--e	Pronoun, personal, second person plural, emphatic
107	Pp3p	Pronoun, personal, third person plural
108	Pp3p--e	Pronoun, personal, third person plural, emphatic
109	Pp3p-n	Pronoun, personal, third person plural, nominative
110	Pp3p-ne	Pronoun, personal, third person plural, nominative, emphatic
111	Pp1s	Pronoun, personal, first person singular
112	Pp1s--e	Pronoun, personal, first person singular, emphatic
113	Pp2s	Pronoun, personal, second person singular
114	Pp2s--e	Pronoun, personal, second person singular, emphatic
115	Pp3sf	Pronoun, personal, third person singular, feminine
116	Pp3sf-e	Pronoun, personal, third person singular, feminine, emphatic
117	Pp3sfn	Pronoun, personal, third person singular, feminine, nominative
118	Pp3sfne	Pronoun, personal, third person singular, feminine, nominative, emphatic
119	Pp3sm	Pronoun, personal, third person singular, masculine
120	Pp3sm-e	Pronoun, personal, third person singular, masculine, emphatic
121	Pp3smn	Pronoun, personal, third person singular, masculine, nominative
122	Pp3smne	Pronoun, personal, third person singular, masculine, nominative, emphatic
123	Pr1p	Pronoun, prepositional, first person plural

124	Pr1p--e	Pronoun, prepositional, first person plural, emphatic
125	Pr2p	Pronoun, prepositional, second person plural
126	Pr2p--e	Pronoun, prepositional, second person plural, emphatic
127	Pr3p	Pronoun, prepositional, third person plural
128	Pr3p--e	Pronoun, prepositional, third person plural, emphatic
129	Pr1s	Pronoun, prepositional, first person singular
130	Pr1s--e	Pronoun, prepositional, first person singular, emphatic
131	Pr2s	Pronoun, prepositional, second person singular
132	Pr2s--e	Pronoun, prepositional, second person singular, emphatic
133	Pr3sf	Pronoun, prepositional, third person singular, feminine
134	Pr3sf-e	Pronoun, prepositional, third person singular, feminine, emphatic
135	Pr3sm	Pronoun, prepositional, third person masculine
136	Pr3sm-e	Pronoun, prepositional, third person masculine, emphatic
137	Px	Pronoun, reflexive
138	Qa	Verbal particle, affirmative
139	Qa+Q--s	Fused Form : affirmative verbal particle fused with past tense verbal particle
140	Qn	Verbal particle, negative complementiser
141	Qnm	Verbal particle, negative complementiser, imperative
142	Qnr	Verbal particle, negative complementiser, relative
143	Qq	Verbal particle, interrogative/ dependent clause marker
144	Qq+Q--s	Fused Form: interrogative verbal particle fused with past tense verbal particle
145	Q-r	Verbal particle, relative
146	Q-s	Verbal particle, subjunctive
147	Q--s	Verbal particle, past tense
148	Rg	Adverb, general
149	Rg+Cc	Fused Form: adverb fused with coordinating conjunction
150	Rs	Adverb, spatial
151	Rt	Adverb, temporal
152	Sa	Adposition, aspectual
153	Sap1p	Adposition, aspectual with possessive determiner, first person plural
154	Sap2p	Adposition, aspectual with possessive determiner, second person plural
155	Sap3p	Adposition, aspectual with possessive determiner, third person plural
156	Sap1s	Adposition, aspectual with possessive determiner, first person singular
157	Sap2s	Adposition, aspectual with possessive determiner, second person singular
158	Sap3sf	Adposition, aspectual with possessive determiner, third person singular, feminine

159	Sap3sm	Adposition, aspectual with possessive determiner, third person singular, masculine
160	Sp	Adposition, preposition
161	Sp+Dp2s	Fused Form: preposition fused with second person singular possessive pronoun
162	Sp+Q-r	Fused Form: preposition fused with relative verbal particle
163	Spa-p	Adposition, preposition with article, plural
164	Spa-s	Adposition, preposition with article, singular
165	Spp1p	Adposition, preposition with possessive determiner, first person plural
166	Spp2p	Adposition, preposition with possessive determiner, second person plural
167	Spp3p	Adposition, preposition with possessive determiner, third person plural
168	Spp1s	Adposition, preposition with possessive determiner, first person singular
169	Spp2s	Adposition, preposition with possessive determiner, second person singular
170	Spp3sf	Adposition, preposition with possessive determiner, third person singular, feminine
171	Spp3sm	Adposition, preposition with possessive determiner, third person singular, masculine
172	Spr	Adposition, preposition with relative particle
173	Spv	Adposition, preposition fused with verbal particle
174	Tdp	Article, definite, plural
175	Tdpf	Article, definite, plural, feminine
176	Tdpfg	Article, definite, plural, feminine, genitive
177	Tdp-g	Article, definite, plural, genitive
178	Tdpm	Article, definite, plural, masculine
179	Tdpmg	Article, definite, plural, masculine, genitive
180	Tds	Article, definite, singular
181	Tdsf	Article, definite, singular, feminine
182	Tdsfg	Article, definite, singular, feminine, genitive
183	Tds-g	Article, definite, singular, genitive
184	Tdsm	Article, definite, singular, masculine
185	Tdsmg	Article, definite, singular, masculine, genitive
186	Ua	Unique membership class, adverbial
187	Uc	Unique membership class, comparative/superlative
188	Uf	Unique membership class, fixed copular multi-word expressions
189	Ug	Unique membership class, agreement
190	Um	Unique membership class, complementiser (phrasal)
191	Uo	Unique membership class, morphophonemic
192	Up	Unique membership class, patronym
193	Uq	Unique membership class, question words
194	Uq+V-p--d	Fused Form: question word fused with present dependent verb

195	Uv	Unique membership class, vocative
196	V-f	Verb, future/ present habitual
197	V-f--d	Verb, future/ present habitual, dependent
198	V-f--r	Verb, future/ present habitual, relative
199	V-f0	Verb, future/ present habitual, impersonal/ passive
200	V-f0-d	Verb, future/ present habitual, impersonal/ passive, dependent
201	V-h	Verb, past habitual/ conditional
202	V-h--d	Verb, past habitual/ conditional, dependent
203	V-h0	Verb, past habitual/ conditional, passive
204	V-h0-d	Verb, past habitual/ conditional, passive, dependent
205	V-h1p	Verb, past habitual/ conditional, first person plural
206	V-h1pd	Verb, past habitual/ conditional, first person plural, dependent
207	V-h1s	Verb, past habitual/ conditional, first person singular
208	V-h1sd	Verb, past habitual/ conditional, first person singular, dependent
209	V-h1sde	Verb, past habitual/ conditional, first person singular, dependent, emphatic
210	V-h1s-e	Verb, past habitual/ conditional, first person singular, emphatic
211	Vm-1p	Verb, imperative, first person plural
212	Vm-2p	Verb, imperative, second person plural
213	Vm-3	Verb, imperative, third person (singular or plural)
214	Vm-1s	Verb, imperative, first person singular
215	Vm-2s	Verb, imperative, second person singular
216	V-p	Verb, present
217	V-p--d	Verb, present, dependent
218	V-p0	Verb, present, passive
219	V-p0-d	Verb, present, passive, dependent
220	V-s	Verb, past
221	V-s--d	Verb, past, dependent
222	V-s0	Verb, past, passive
223	V-s0-d	Verb, past, passive, dependent
224	Wpdia	Copula, present/ future, dependent, indicative, affirmative
225	Wpdin	Copula, present/ future, dependent, indicative, negative
226	Wpdqa	Copula, present/ future, dependent, interrogative, affirmative
227	Wpdqn	Copula, present/ future, dependent, interrogative, negative
228	Wp-i	Copula, present/ future, indicative
229	Wp-i-3	Copula, present/ future, indicative, third person pronoun
230	Wp-in	Copula, present/ future, indicative, negative
231	Wp-i-x	Copula, present/ future, indicative, existential
232	Wpr	Copula, present/ future, relative
233	Ws	Copula, past/ conditional
234	Xa	Residual, acronym
235	Xd	Residual, date
236	Xe	Residual, e-mail address
237	Xf	Residual, foreign

238	Xfe	Residual, foreign, English
239	Xsc	Residual, spoken, communicator
240	Xsev	Residual, spoken, event
241	Xsi	Residual, spoken, incomplete
242	Xsp	Residual, spoken, phonetic element
243	Xw	Residual, website address
244	Xx	Residual, unknown/ unintelligible
245	Xy	Residual, symbol
246	Y	Abbreviation

8. SIMPLIFIED TAGSET WITH CORRESPONDING PART OF SPEECH

Note: The simplified tagset is, by definition, unable to encode morphological information to the same degree as the default tagset. While this may not impede the basic user, there are a few cases of data loss deserving mention.

In the case of synthetic verbal forms (e.g. *bhithinn*), the pronoun is not specified in simplified tagset. So, whereas we have V-h1s in the default tagset (past-habitual/conditional verb in the 1st person singular), the simplified tagset gives us V-h (past-habitual / conditional verb). Additionally, prepositions encoding the definite article (e.g. *dhan*) are tagged as Spa-s in the default tagset, but as Sp in the simplified tagset, omitting reference to the definite article.

Note also that, in the case of previously fused forms (e.g. Sp+Q-r ‘preposition with relative verbal particle’), the simplified tag is formed by the first element (e.g. Sp).

NO	TAG	PART OF SPEECH
1	Ap	Adjective: predicative
2	Aq	Adjective: qualifactor (attributive)
3	Ar	Adjective: preposed
4	Av	Adjective: verbal (inc past participles)
5	Cc	Conjunction: coordinate
6	Cs	Conjunction: subordinate (e.g. <i>ma</i> ‘if’)
7	Dd	Determiner: demonstrative
8	Dp	Determiner: possessive
9	Dq	Determiner: quantifier
10	F	Punctuation
11	I	Interjection (e.g. <i>shìorraidh</i> ‘heavens’)
12	M	Numeral or numerical operator (e.g. ‘+’)
13	Nc	Noun: common
14	Nf	Noun: fossilised historical (in multi-word expression)
15	Nn	Noun: personal name
16	Nt	Noun: toponym
17	Nv	Noun: verbal
18	Pd	Pronoun: demonstrative
19	Pn	Pronoun: numerical (e.g. <i>aon dhiubh</i> ‘one of them’)
20	Pp	Pronoun: personal
21	Pr	Pronoun: prepositional
22	Px	Pronoun: reflexive
23	Q	Verbal particle (e.g. complementisers, dependent clause markers, interrogative clause markers)
24	Rg	Adverb: general

25	Rs	Adverb: spatial
26	Rt	Adverb: temporal
27	Sa	Adposition: aspectual
28	Sap	Adposition: aspectual with a possessive determiner
29	Sp	Adposition: preposition
30	Td	Article: definite (there is no indefinite article in ScG)
31	U	Unique membership class, e.g. vocative particle 'a'; adverbial particle 'gu'; patronym 'Mac'
32	Vf	Verb: main, future/ present habitual
33	Vh	Verb: conditional/ past habitual
34	Vp	Verb: present
35	V-s	Verb: past
36	Vm	Verb: imperative
37	Wp	Copular verb: present / future
38	Ws	Copular verb: past / conditional
39	X	Residual (e.g. acronyms, dates)
40	Xfe	Residual: foreign, English
41	Y	Abbreviation