

# Scottish Gaelic constituency treebanking guidelines v. 0.1

Colin Batchelor

November 11, 2016

## Contents

|          |                                       |          |
|----------|---------------------------------------|----------|
| <b>1</b> | <b>Introduction</b>                   | <b>1</b> |
| <b>2</b> | <b>Constituents</b>                   | <b>2</b> |
| 2.1      | ADJ . . . . .                         | 2        |
| 2.2      | ADV . . . . .                         | 2        |
| 2.3      | CONJ . . . . .                        | 2        |
| 2.4      | COSUB, COSUB0 . . . . .               | 2        |
| 2.5      | NP . . . . .                          | 2        |
| 2.6      | PP . . . . .                          | 2        |
| 2.7      | sent . . . . .                        | 3        |
| 2.8      | S0, S0DEP . . . . .                   | 3        |
| 2.9      | SASP . . . . .                        | 3        |
| 2.10     | SANN, SANNBAR . . . . .               | 3        |
| 2.11     | SBAR and SREL . . . . .               | 3        |
| 2.12     | SDCL, SINT, SNEG, SNEGINT . . . . .   | 3        |
| 2.13     | SDEP . . . . .                        | 4        |
| 2.14     | SGUN . . . . .                        | 4        |
| 2.15     | SINF . . . . .                        | 4        |
| 2.16     | SSMALL . . . . .                      | 4        |
| 2.17     | SSUB . . . . .                        | 4        |
| 2.18     | utt . . . . .                         | 4        |
| 2.19     | V, VDEP . . . . .                     | 4        |
| <b>3</b> | <b>Special cases</b>                  | <b>5</b> |
| 3.1      | Punctuation-absorbing rules . . . . . | 5        |
| 3.2      | Numbers . . . . .                     | 5        |

## 1 Introduction

The guidelines here assume the ARCOSG tagging scheme.

There is one minor change to the tokenization: for simplicity we collapse any preposition containing a hyphen, for example *a-measg* and *a-rèir*, into a single token. In the text that follows we give an explanation of each kind of constituent and an example taken from elsewhere in ARCOSG, specifying the file and the

line in which the example text occurs. The aim is to produce completely binary trees.

## 2 Constituents

### 2.1 ADJ

These are cases where a predicative adjective is modified by, say *cho* or *nas*, or with a PP headed by *ri*. In the *cho... ri...* case, *cho* goes on the inside. Example:

(ADJ (ADJ (Rg cho) (Ap mòr)) (PP (Sp ri) (Ncsmd each))) (p01:120)

### 2.2 ADV

These are modifiers, typically beginning with the adverbializer *gu* (Ua). Example:

(ADJ (ADV (Ua gu) (Rg math)) (Ap comhfhurtail)) (s08:69)

If the adverb is modifying a clause rather than an adjective, then if it is an adverb of place it can modify a small clause SSMALL but if it is an adverb of time then it should modify the closest accessible tensed clause, for example SBAR or SDCL.

### 2.3 CONJ

These are headed by a coordinator, tagged Cc, or by a comma, Fi in the context of a list. Example:

(CONJ (Cc agus) (SDCL (S0 (V-s fhuaire) (Nt Yugoslavia)) (Pp3sfi))) (s10:53)

Also use CONJ for constructions like *air sgàth 's gu bheil...*

### 2.4 COSUB, COSUB0

These are for cosubordination, where the coordinators *agus*, *is* or *'s* stand in place of the verb *bi* and describe the preceding clause. Example:

(COSUB (COSUB0 (Cc 's) (Pp3sm e)) (SASP (Sa g') (Nv iasgach))) (n05:39)

### 2.5 NP

Work out from the noun N\* in the following order: euphonious prefixes, preceding adjectives and numbers, determiners, modifying NPs and PPs. Example:

(NP (Tdsm an) (NP (Mc aon) (NP (Uo t-) (Ncsmd sagart)))) (pw01:18)

### 2.6 PP

These are headed by a preposition, tagged Sp, or the fossilized noun *airson* (Nf). For compound prepositions, for example, *as dèidh*, we preserve the tokenization and treat *dèidh* as an ordinary noun. Examples:

- (PP (Sp tro') (NP (NP (Tdpm na) (Ncpmd soluis)) (Aq-pmd dhearg))) (ns10:2)

- (PP (Sp air) (NP (Nf ais) (PP (Sp gu) (Nn Nicholl)))) (s10:150)

## 2.7 sent

Treat the sentence within as SDCL (or SINT, SNEG, SNEGINT) and the final punctuation as a modifier. Example:

```
(sent (SDCL (SDCL (V-f Thèid) (SINF (Ncsfn binn) (SINF (Ug a) (SSMALL
(Nv thoirt) (Pr3sm air)))))) (Rt a-rithist)) (Fe .)) (ns07:31)
```

## 2.8 S0, SODEP

These are constituents which contain a verb and its subject. They can contain other constituents as long as that constituent is a modifier. Example:

```
(S0 (V-p Tha) (Pp3sm e)) (pw11:48)
```

However, if there is a passive sentence using the infinitive construction with *rach* (or *chaidh*, *deach*, etc.), then treat the object as part of the infinitive clause SINF. Example:

```
(SDCL (V-s Chaidh) (SINF (NP (Tdsm am) (Ncsmn prògram)) (SINF (Ug
a) (SSMALL (Nv dhèanamh) (PP (Sp le) (Nn Hummingbird)))))) (pw07:61)
```

## 2.9 SASP

These are clauses that begin with an aspect marker, tagged *Sa* in the ARCOSG scheme, and typically contain a small clause SSMALL. Example:

```
(SASP (Sa air) (SSMALL (Dp1s mo) (Nv mhealladh))) (pw10:22)
```

## 2.10 SANN, SANNBAR

These are to mark fronted constituents. The rest of the sentence will be a SREL but the structure of the first part is, for example:

```
(SANN (SANNBAR (Wp-i 'S) (Pr3sm ann)) (PP (Sp le) (Ncsmd cianalas)))
(pw06:65)
```

## 2.11 SBAR and SREL

These are to be used in relative clauses, introduced by the relative particle *a* (Q-r).

Example:

```
(SREL (Q-r a) (SBAR (V-p tha) (SASP (Sa air) (SSMALL (Dp3sm a)
(Nv neartachadh)))) (pw01:39)
```

## 2.12 SDCL, SINT, SNEG, SNEGINT

These are clauses that are usually complete sentences in themselves but do not contain the final full stop or quotation mark. The POS-tag of the verb will usually be *V-s* or *V-p* where it is a declarative sentence. Typically they will contain an S0 clause followed by an NP, a PP, an SGUN or an SINF, depending on the subcategorization of the verb.

```
(SDCL (PP (Sp Às) (NP (Nf aonais) (NP (Tdpmg nan) (Ncpmg Gàidheal))))
(SDCL (S0 tha/V-p i/Pp3sf) (SASP (Sa ag) (SSMALL (Nv ràdh) (SGUN (Qn
```

nach) (SDEP (SODEP (V-h--d biodh) (Ncsfn sgir-easbaig)) (Pr3sm ann)))))) (pw01:99)

### 2.13 SDEP

Typically they will contain an SODEP clause followed by an NP, a PP, an SGUN or an SINF, depending on the subcategorization of the verb.

(SSUB (Cs mura) (SDEP (SODEP (V-h--d cosnadh) (Pp3p iad)) (Pp3sm e))) (pw04:70)

### 2.14 SGUN

These are clauses that are headed by the particle *gu*, tagged **Qa**. These almost always contain a dependent clause SDEP.

(SGUN (Qa gu) (SDEP (SODEP (V-p--d bheil) (Pd sin)) (Ap ceart))) (s10:27)

### 2.15 SINF

These are clauses that either begin with an agreement particle, tagged **Ug**, or with an NP which is the object of the clause.

(SINF (Ug a) (SSMALL (Nv dhol) (SINF (Ug a) (SSMALL (Nv bruidhinn) (PP (Sp ri) (Ncsmd cuideigin)))))) (p01:154)

### 2.16 SSMALL

These are clauses headed by a verbal noun, **Nv**. They contain objects, other arguments and modifiers of the verbal noun. The object, be it a determiner tagged **Dp3sm** or similar, or an NP in the genitive, sits inside SSMALL.

(SASP (Sa ') (SSMALL (Nv cumail) (Rg a-mach))) (s08:117)

### 2.17 SSUB

These are clauses introduced by subordinators **Cs**. Example:

(SSUB (Cs mura) (SDEP (V-f0-d lorgar) (NP (Ncpmn camaidhean) (Aq-pmn eile)))) (fp04:43)

### 2.18 utt

This is to be used to mark an utterance in conversation. It can contain anything. It begins with an indication of the speaker, tagged **Xsc**, if it is at the start of a turn. Examples:

- (utt (Xsc [1]) (sent (SDCL (Uq dè)) (Fg ?))) (c09:122)
- (utt (Xsc [4]) (SDCL (V-f feumaidh) (SGUN (Wpdia gur) (Pr3sm ann)))) (c09:121)

## 2.19 V, VDEP

These are special cases to handle where a verb is preceded by a particle such as *dh'* (Uo).

(S0 (V (Uo dh') (V-s fheuch)) (Pp3p iad)) (s09:117)

## 3 Special cases

### 3.1 Punctuation-absorbing rules

In newswire, there are a lot of stray dashes and commas. The rule here is that you use your judgement to work out which preceding clause should absorb them and treat them as a modifier.

### 3.2 Numbers

Treat numbers as NPs and try to keep them together.