**Morphology tagset for the CorPH database**

Fangzhe Qiu, Elliott Lash, Bernhard Bauer, David Stifter

latest version: 03/12/2019

**1. Introduction**

This document details guidelines for tagging the various attributes in the ‘Morphology’ table in the CorPH database. For the structure of the database and all the columns in the ‘Morphology’ table, please refer to the Guideline for the structure of the CorPH database. The attributes under discussion here include ‘Analysis’, ‘Relativity’, ‘Transitivity’, ‘Dependency’, ‘Deponent’, ‘Contraction’, ‘Augment’, ‘Hiatus’, ‘Mutation’, ‘Causing\_Mutation’, ‘Onomastic\_Complex’ and ‘Onomastic\_Usage’.

For each attribute, a limited set of tags are used. Most tags are self-explanatory, only in some cases the full meaning of the tag is explained. **Ø** means null value for the tag or no tag.

**2. Analysis**

2.1 Main principles

The tag for ‘Analysis’ is given on the basis of the grammatical function of the morph, even though sometimes the surface form of a morph resembles another grammatical form due to allomorphs (synchronic or diachronic), scribal error or manuscript defect, e.g.

*dona doínib* ‘for the people’, morph *na* is tagged **dat.pl.** (not **acc.pl.**)

*in fer… rodas·marb in láech* ‘the man… the warrior killed him’, morph *das·* is given the lemma **3sg.masc.inf.pron.** and tagged as **A.acc.**

*leis-sa* (allomorph for *som*), morph *sa* is tagged **3sg.masc./neut.**

*leis-si* (scribal error for *siom*), morph *si* is tagged **3sg.masc./neut.**

*Berthar in chloch* ‘the bell is carried’, *berthar* is NOT tagged as **rel.**

However, when the agreement rules deviate from the general principles of the standard grammar, or when the agreement in standard Old Irish allows for such variation, or when shift of paradigm or grammatical categories happen, the morphological tag is given on the basis of the agreement, e.g.

*Táncatar in slúag ‘*the host came’, morph *táncatar* is tagged **3pl.pres.ind.** (variation in agreement of plurality between subject and verb)

*atanneirrig* ‘who emends us’, morph *tan* is tagged **B.acc.** (Class B and C are both possible in this context)

*in biad* ‘the food’, morph *in* is tagged **nom.sg.masc.** (the gender of *biad* has shifted)

When it cannot be judged by the surface form whether a shift of paradigm or grammatical categories has happened, the tag is given on the basis of the standard grammar.

Tags should be concatenated without blank space but with dots, e.g. **nom.sg.neut.**; **augm.3sg.pres.ind.pass.rel.** When multiple tags are possible, they are all given and separated by a forward slash (/), e.g. **3sg.masc./neut.**, when the context is not clear which is the referent of the pronominal affix or the nota augens.

2.2 For nouns, the tag has the format of case + number:

case:

**nom. gen. dat. acc. voc. abl. loc.**

(abl. and loc. are for proper nouns using Latin case endings)

number:

**sg. du. pl.**

for compositional forms in a nominal composition, use **compos.**, e.g. *ríg* in *rígdomnae*, whereas *rígdomnae* and *domnae* are tagged as **nom.sg.** and **nom.sg.**.

2.3 For adjectives in the positive degree, verbal\_participle and definite\_article, the tag has the format of case + number + gender.

case and number tags are those used for nouns in 2.2, genders are:

**masc. neut. fem.**

In case of uncertain case/number/genders: in the order of the list above, e.g. masc./neut./fem., whichever genders are possible.

adjective\_numeral, adjective\_numeral\_noun and numeral are tagged in the format stress + case + number + gender.

stress:

**stress.** indicates that the numeral is stressed (i.e. when it is the head of its own phrase), otherwise **Ø**. When the numeral is stressed, the gender is tagged as **neut.**

2.4 For equative, comparative and superlative degrees of adjectives, use the following tag:

**equ. comp. sup.**

2.5 For verbs, see Stifter, David in *The Celtic Languages* (2nd ed.), pp. 85-88 for inflectional categories. The tag has the format of augment + verbal person + tense/ mood + voice + relativity

augment:

**augm. Ø**

All augmented verbal forms (or ‘perfect’) should be tagged with **augm.**, including forms that are not augmented by augment particles such as *do·uccai* ‘s/he brings’. For forms that are augmented by augment particles, such as *do·chuaid* ‘s/he has gone’ or *do·rigni* ‘s/he has made’, the augment should receive its own morph, and the ‘augment’ field in the ‘morphology’ table should be filled as **Yes**, see 3.6 below.

verbal person:

**1sg. 2sg. 3sg. 1pl. 2pl. 3pl.**

tense/mood:

**pres.ind.** = present indicative

**hab.**  = habitual or ‘consuetudinal’

**impf.** = imperfect

**pret.** = preterite (including the augmented ‘perfect’)

**fut.** = future

**pres.subj.** = present subjunctive

**past.subj.** = past subjunctive

**cond.** = conditional or ‘secondary future’

**impv.** = imperative

voice:

**pass. Ø**

relativity:

**rel. Ø**

the rel. tag is only used when the verb takes the absolute relative ending *-s, -mai, -tae*, etc.

e.g. *bes* ‘which would be’: **3sg.pres.subj.rel.**; ·roilgisid ‘ye have studied’: **augm.2pl.pret.**

2.6 For simple prepositions, enter the case that it governs in the context, namely **acc.** **dat.** or **gen.**

2.7 For conjugated prepositions, the format is person + case:

person:

**1sg. 2sg. 3sg.masc. 3sg.neut. 3sg.fem.**

**1pl. 2pl. 3pl.**

e.g. *linn* **1pl.acc.**; *for* **3sg.masc./neut.dat.** (when the context is unclear whether the referent is masculine or neuter)

2.8 For pronouns, each subcategory has its own tags:

2.8.1 For pronoun\_demonstrative, pronoun\_anaphoric, the format is case + person. Since these are all 3rd person, the letter ‘3’ is omitted (see below).

Further rules are specified here, i.e.

a) as the subject of copular sentences, the demon.pron. is always **nom.sg.neut.**

b) when used with the 'propword' í, it is always **dat.sg.neut.**

c) in other constructions, one needs to see whether it agrees with plural verbal forms (as subject), plural infixed/suffixed pronouns (as object), or, when singular, whether it refers to masc. or fem. referents. The default gender is neuter.

e.g. *airi* ***sin*** ‘before her’ **dat.sg.fem.**

*It ē mnā in* ***sō*** *nā dlegut lōg n-eneach* ‘these are the women who are not entitled to the honour-price’ **nom.sg.neut.**

*ian oil...a lleth in* ***sin*** *ar llestar colpthaige* ‘a pale of an *ól*...half of that for a two-year-old-heifer hive’, **gen.sg.fem.** (referring to *ían*)  
*deligfitir* ***sin*** *tall isind eséirgi* ‘they will be separated beyond in the Resurrection’ **nom.pl.**

2.8.2 For pronoun\_independent, there is no tag, except in the following cases:

*messe*, *sisi* etc. are analysed as *mé-se, sí-si*, i.e. indp.pron. + nota augens (2 morphs), where the nota augens is tagged accordingly;

stressed forms *aí, moí, toí* have the lemmata ‘3sg.neut.indp.pron.’, ‘1sg.indp.pron.’ etc, and they are tagged as **gen.**

2.8.3 For pronoun\_interrogative, pronoun\_quantifier, tag the case (**nom.** etc.)

2.8.4 For pronoun\_infixed, the format is class + case

class: **A. B. C.**

2.8.5 For pronoun\_suffixed, enter the case **acc.** or **dat.**

2.8.6 For pronoun\_propword *í*, enter the case + person (see 2.8.1).

2.8.7 For pronoun\_emphatic, enter the person (see 2.7):

2.9 For particles, each subcategory has its own tags:

2.9.1 For particle\_anaphoric (all 3rd person), the format is case + person (see 2.8.1).

2.9.2 For particle\_pronominal (notae augentes), enter the person (see 2.7):

2.10 Pure (lexical) adverbs do not have morphological tags, e.g. *beus* ‘yet, still’, *sís* ‘below’.

For adverbs derived from adjectives, the morph’s POS is still ‘adjective’ and they have the following tags in Analysis:

**dat.sg.neut.** or **acc.sg.neut.**

The *-aid, -aith* adverbial forms derived from -*de* adjectives have been given separate lemmata, e.g. *immḟolngidid* ‘effectively’, POS = **adverb**, analysis = **Ø**.

3. The verbal attributes

These are all attributes of verbs and should only be filled when the morph has a lemma whose POS = verb. Currently they are not universally tagged in CorPH but drop-down lists containing the following tags are offered on the website (editor mode) for future tagging.

3.1 Relativity: **Yes** = the verb heads a relative clause, **No** = the verb does not head a relative clause, **Maybe** = unclear case.

3.2 Transitivity: **transitive intransitive passive unclear**

3.3 Dependency: **absolute conjunct deutorotonic prototonic**

3.4 Deponent: **Ø** = non-deponent verb, **Yes** = deponent ending (regardless if the verb is deponent or non-deponent originally), **No** = active ending for verb otherwise attested as deponent.

3.5 Contraction: **Yes** = contracted form (e.g. *tic* ‘he arrives’), **No** = not contracted form [###FQ: does this include cases where contraction is not expected to happen at all, i.e. not CV-V…?], **Maybe** = unclear case.

3.6 Augment: **Yes** = augmented by augment particles (cf. 2.5 above), **No** = not augmented, **Maybe** = unclear case. [###FQ: I am not sure about the use of this column anymore]

4. Mutation and Causing\_Mutation

The tags used in these two columns are:

**+ Nasalization - Nasalization**

**+ Lenition - Lenition**

**+ Gemination - Gemination**

Currently they are not universally tagged in CorPH but drop-down lists containing the following tags are offered on the website (editor mode) for future tagging.

These tags are given in the following contexts:

4.1 when the initial consonant of a morph is mutated, or *n-* or *h-* is added before an initial vowel of a morph, and these mutations are shown in the spelling of the morph, the + tags are given in this field.

e.g. *chíall* **+ Lenition**; *mmolad* **+ Nasalization**; *h-ech* **+ Gemination**.

note: in *ind fir* ‘the men’, *fir* **+ Lenition**, because the lenited *f* is indicated by the -*d* in the definite article.

note: prepositions prefixed with *h* are not treated as being affected by germination, e.g. *hi* ‘in’, *hiter* ‘between’.

4.2 when morpho-syntactic rules demand that the initial consonant or vowel of a morph be mutated (or added *n-* or *h-*), and standardised Old Irish orthography (as adopted for the Lemmata) allows these mutations to be indicated, but such mutations are not shown in the spelling of the morph, the - tags are given.

Mutations allowed by the standardised orthography include:

nasalization: n-V, mb, nd, ng, ll, rr

germination: hV, mm, nn, pp, bb, ss, ll, gg, rr

lenition: ch, ph, th.

e.g. *la amrae* ‘with a miracle’, *amrae* **- Gemination**

*no·guidim* ‘I beseech him’, *·guidim* **- Nasalization**

*a carpat* ‘his chariot’, *carpat* **- Lenition**

*anro·chésasat* ‘that which they had suffered’, *ro·* **- Nasalization**

4.3 when mutations are indicated in a way that the standardised Old Irish orthography does not allow, e.g. in later manuscripts or with idiosyncratic spelling, the + tags are used, but it should be indicated in the ‘comment’ that this is not the standardised Old Irish spelling. e.g.

*di gheinntib* ‘by the heathens’, *gheinntib* **+ Lenition**, comment: ‘non-standard spelling’

*a ṡarugud* ‘violating him’, *ṡarugud* **+ Lenition**, comment: ‘non-standard spelling’

*i bpein* ‘in pain’, *bpein* **+ Nasalization**, comment: ‘non-standard spelling’

*con·dánicc* ‘until he came’, *·dánicc* **+ Nasalization**, comment: ‘non-standard spelling’

4.4 When mutations are indicated in the morph which are against the morpho-syntactic rules, they are marked with the + tags, but it should be indicated in the ‘comment’ that these mutations are not grammatical/expected. e.g.

*co chéill* ‘with sense’, *chéill* **+ Lenition**, comment: ‘unexpected lenition’

4.5 When a morph is tagged in the ‘Mutation’ column, the preceding morph which causes such a mutation or non-appearance of expected mutation is tagged with the same tag, but in the ‘Causing\_Mutation’ column, e.g.

*ind fir* ‘the men’, *ind* **+ Lenition** (cf. 4.1)

*la amrae* ‘with a miracle’, *la* **– Gemination** (cf. 4.2)

*i bpein* ‘in pain’, *i* **+ Nasalization**, comment: ‘non-standard spelling’ (cf. 4.3)

*co chéill* ‘with sense’, *co* **+ Lenition**, comment: ‘unexpected lenition’ (cf. 4.4)

5. Onomastic\_Complex and Onomastic\_Usage

Currently they are not universally tagged in CorPH but drop-down lists containing the following tags are offered on the website (editor mode) for future tagging.

5.1 Onomastic\_Complex

This field is for marking the whole proper name, when the proper name consists of several morphs.

e.g. under each morph *glenn*, *dá* and *locha*, one marks in this column ‘Glenn dá Locha’, also *Cú cen Máthair*; *Uí Fidgenti*; *Glúin Duib*; *Mac an Bhaird* etc.

If a proper name consists of only one morph, this field is left unfilled, e.g. *Breg; Díarmait; Lífe; Dergfholt*.

5.2 Onomastic Usage:

When the Onomastic\_Complex column is populated, choose from the following tags:

**anthroponym epithet ethnonym hydronym toponym**