**Guidelines for isolating stressed units and morphs in CorPH**

Elliott Lash and Fangzhe Qiu

last update: 13/11/2019

**1. Structure of a Stressed Unit**

A stressed unit is minimally made up of stressed item, represented by S in (1).

1. S

The stressed item can be preceded and followed by unstressed items, represented by U in (2). This means that unstressed items may be either proclitics or an enclitics. The ellipses represent the fact that there may be more than one unstressed item that cliticizes to the stressed item.

2. (U…)S(U…)

The ‘stressed unit’ field must be filled with the stressed item and any associated proclitic or enclitic items. A new stressed unit is created when the last enclitic item of the previous stressed unit is reached. For example, consider the following (partially made up) gloss:

3. *Bríathar as·berar fo óenaicciund, ros·failsigestar-som-side ocus is deithbir dun-ni cid cert-som.*

“A word which one says under a single accent, he has clarified it; and it is clear to us that he is right.”

This is made up of the following stressed units:

4. Unit Reason

*bríathar* **Noun**

*as·berar* Preverb + **Verb Root**

*fo óenaicciund* Preposition + **Numeral Compound**

*ros·failsigestar-som-side* Verbal Complex

(Augment + Pronoun + **Verb Root** + Nota Augens + Clitic Anaphor)

*ocus is deithbir* Conjunction + Copula + **Adjective**

*dun-ni*  **Conjugated Preposition** + Nota Augens

*cid cert-som* Copular Form + **Adjective** + Nota Augens

**2. Stressed Items**

1. Adjectives

2. Adjective Compounds (**A**-N, **A**-A)

3. Adverbs

4. Nouns (**N**)

5. Compound Nouns (**N**-N)

6. Conjugated Prepositions

7. Pronominal Demonstratives: *só*, *sé*, *sin* (including *síu* in *a n-í síu*, see L. Breatnach’s chapter in the ChronHib proceedings)

8. Pronouns: *mé*, *tú*, *é*, *sí*, *sní*, *sí/sib*, *é*

9. Pronominal Quantifiers: *cách*, *nech*, *ní*, *alaile*

10. Prop-word (meaning ‘one/thing’): *í*

11. Simple Verbs (**V**)

12. Complex Verbs (PV+[PV2-5]**V**)

13. Verb Compounds (**A**-V): there are at least two items that form such compounds: **bith**-V and **sain**-V (see these in Castillero 2014.71).

14. Grammaticalized compounds: *in-****dé***, *in-****díu***, *al-****aile***, etc.

15. Genitive Pronouns: *muí*, *tuí*, *aí*

16. Independent numbers (*óen*, *dó*, *trí*, *ceithir, cóic*) (these are usually preceded by the particle *a*)

17. Numeral Compounds (**Num**-N)

18. patronymic elements (úa, mac, moccu)

**3. Unstressed Items**

**3.1 Proclitics**

1. Prepositions

2. Preverbs

3. Articles

4. Quantifiers: *cech*, *nach*, *na*

5. *ala*

6. Complementizers

7. Negatives Particles

8. Posssessive Pronouns

9. Copula forms

10. The quotative particle *ol*

11. The predicative particles *os* and *ot*

12. Numerals as determiners (*da, tri, ceithir)*

13. Some Conjunctions

Note that only some conjunctions are proclitics. In particular, only conjunctions that do not include or consist of a stressed item are proclitics. For instance, *óndí* ‘in as much as, because’ includes the stressed item *í* (a prop-word, see section 2, ex. 10). The conjunction is therefore a strssed unit on its own. Likewise, the conjunction *úaire* ‘since’ is probably a stressed item, since it is originally a noun and nouns are stressed units. However, *intan* ‘when’ is most likely unstressed, given the depalatalisation of *tain*.

**3.2 Enclitics**

1. Notae Augentes (-*sa*, -*su*, -*som*, -*si*, -*ni*, -*si*, -*som*)

2. Unstressed anaphor: *-side* (enclitic)

3. Unstressed demonstratives: *-so, -sin, -se* (enclitic)

**4. Definition of a ‘morph’**

In CorPH, a ‘morph’ is the smallest analysable unit that has a lemma. Apart from the usual lexical words, it includes the following categories:

1. in the case of a compound noun or a compound adjective, all compound elements that has a lexical meaning, together with the whole compound itself, are counted as separate morphs. e.g.:

*athlaoch* ‘an ex-layman’, should be analysed as containing the following morphs:

*ath* ‘re-, ex-‘ (Lemma = ‘ath-‘, POS = particle\_prefix)

*laoch* ‘warrior, layman’ (Lemma = ‘láech’, POS = noun)

*athlaoch* ‘ex-layman’(Lemma = ‘athláech’, POS = noun)

*saindiles* ‘special, one’s own’, should be analysed as containing the following morphs:

*sain* ‘different, distinct’ (Lemma = ‘sain’, POS = adjective)

*diles* ‘own, proper, appropriate’ (Lemma = ‘díles’, POS = adjective)

*saindiles* ‘special, one’s own’ (Lemma = ‘saindíles’, POS = adjective)

2. in the case of a verbal complex, all conjunctions, complementisers, particles, prefixes, preverbs, augments, demonstratives, infixed and suffixed pronouns, together with the whole lexical verb itself, are separate morphs. In the morph of the lexical verb proper, the augment and relative mutation are written, and the infixed pronoun is included there as well, if it occurs between the augment or the preverb and the verbal root. e.g.

*as·berar*, morphs:

*as·*

*as·berar*

*ros·failsigestar-som-side*, morphs:

*ro*

*s·*

*ros·failsigestar*

*som*

*side*

*ma nis·tinib* ‘if he would not breathe on them’, morphs:

*ma*

*ni*

*s·*

*·tinib* (*do·infet*)

*·t* (preverb *do·*)

*in* (preverb *in·*)

3. in the combination of prep. + def. art., the *s* which results from earlier \*-*Cs-* on the morpheme boundary always belongs to the morph of the def.art., e.g.

*frisin claideb* ‘towards the sword’, morphs:

*fri*

*sin*

*claideb*

even when the *s* is analogous and not historical, e.g.

*forsind loch* ‘upon the lake’, morphs:

*for*

*sind*

*loch*

4. Exceptions:

a) Proper nouns are not separated into morphs, although many of them originated as compound nouns. e.g.

*Fergus* ‘man-prowess’, *Dublind* ‘dark-pool’.

b) Lexicalised nouns and adjectives are not treated as compounds, e.g.

*follus* (ety. < *fo* + *lés*) ‘bright’, morph = *follus*

*sóer* (ety. < *so* + *fer*?) ‘free, noble’, morph = *sóer*

c) The following lexicalised adverbs and conjunctions have their own lemma and are not analysed as conjugated prepositions or compounds:

*cosé* ‘until’, *lasé* ‘when’, etc. (lemma = *cosé*, etc., no morph *co* or *sé*)

*etir* ‘at all’ (lemma = *etir*, analysis = NULL)

*íarum* ‘afterwards’ (lemma = *íarum*, analysis = NULL)

*as* ‘out’ (lemma = *as*, analysis = NULL)

However, *arindí* ‘because’, *óndí*, *intan* ‘when’ etc. are not treated as lexicalised, therefore:

*arindí*, morphs = *ar*

*ind*

*í*

*intan*, morphs = *in*

*tan*

d) Verbal nouns and verbal of necessities are not separated into morphs, though many of them, like their verb counterparts, contains preverbs. e.g.

*dígbál* ‘act of taking away’, morph = *dígbál* (no morph *dí*),

*eiperthi* ‘to be said’, morph = *eiperthi* (no morph *ei*)

e) Conjugated prepositions are regarded as single morphs. e.g.

*leu* ‘by them’, morph = *leu* (no morph *le* or *u*)

f) derivative suffixes are not treated as separate morphs, nor are the word on the basis of which the derivatives are formed, e.g.

*bochtán* ‘little poor one’, morph = *bochtán* (no morph *bocht* or *án*)

*tírthat* ‘agellus’, morph = *tírthat* (no morph *tír* or *that*)

**5. Presentation of the Stressed Unit and the Morph**

The sentences in CorPH are copied from the published editions and keep the spelling, diacritics and punctuation of the edited texts. In principle, the Stressed Unit and the Morph are copied from the sentence, and thus also keeping the macrons, sine fada, diaraesis etc. in the edition.

The following punctuations, however, are added to facilitate the users to grasp the prosodic profile of the data in a more unified manner:

1. Pretonic element(s) and the stressed part are separated by a raised dot in a verbal complex, but by a space elsewhere, including between a copula and the stressed predicate. The raised dot is written in the morph as well when the morph is immediately next to the stress boundary. The pretonic elements themselves are not separated by space or punctuation, except between a copula and the following unstressed elements.

e.g. Sentence *conachaepert…* ‘so that he did not say it (fem.)’

> SU *conacha·epert*

> Morphs *co*

*nach*

*a·*

*·epert*

*·e*

Sentence *innathaig…* ‘in his house’

>SU *inna thaig*

> Morphs *i*

*nna*

*thaig*

Sentence *cid liambríathraib…* ‘though it be with their words’

>SU *cid lia mbríathraib*

> Morphs *ci*

*d*

*li*

*a*

*mbríathraib*

A single morph is not separated by any raised dot or space, even when the stress does not fall on the first syllable, e.g. adverbs such as *immallé*

2. Post-tonic enclitic is separate from the stressed element, and from each other, by a hyphen.

e.g. Sentence *ro-s:failsigestar-somside*

>SU: *ros·failsigestar-som-side*

Sentence *infearso* ‘this man’

>SU: *in fear-so*

This hyphen is not written in the morphs.

e.g. SU: *in fear-so*

>morphs: *in*

*fear*

*so*

3. A hyphen is written between a mutation n and the following vowel, but not after other mutation letters.

e.g. Sentence *anás-mbeir* ‘when he says’

>SU: *a n-ás·mbeir*

This hyphen is not written in the morphs.

e.g. SU: *a n-ás·mbeir*

>morphs: *a*

*nás·*

*nás·mbeir*

4. A conjunction is always followed by a space. This rule overrides all the above ones.

e.g. Sentence *ocusinfearso*

>SU: *ocus in fear-so*

Sentence *manīs:tinib=side* ‘if the abovementioned person would not breathe on them’

>SU: *ma nīs·tinib-side*

4. sometimes one or more morph is inserted into another morph, due to re-analysis of morpheme boundaries, e.g. GOI §529, or to tmesis. In such cases, ‘…’ is used to represent the space between the parts of the same morph, e.g.

*fo·rrorbris* ‘whom He had routed', Ml. 67b24, morphs:

*fo·rrorbris*

*fo·…r*

*·rro*

*ø* (the nasalising particle)

*ad· cruth caín ·cichither* ‘the fair form will be seen’, SU *ad·…·cichither*, morphs:

*ad·…·cichither*

*ad·*

Vowels that are unexpected lost are represented by an apostrophe (‘), excluding contracted forms of verbs, in both SU and morph, e.g.

*M’druad*, morphs:

*M’*

*druad*

*ho’ inni* ‘from its essence’, morphs:

*ho*

*‘*

*inni*

but *tucc* ‘he gave’, morphs:

*tucc*

*t*

5. When the edition, on which CorPH is based, has indicated editorial changes to the manuscript text, the edited text should be used in the SU. Reading that is not in the MS but in the edited text is put between square brackets ‘[ ]’; reading that is in the MS but not in the edited text is put between round brackets ‘( )’. In practice, esp. in Ml., this means that the text after *leg.* should be used, and what is different between the *leg.* text and the MS reading is put between the two types of brackets.

For Morph, the brackets are omitted and the expected reading (that of *leg.* in Ml.) is used.

e.g.

Ml. 021c01 edition: *(im)baloin [leg. imbabaloin]* (*im* is legible on MS but disagrees with Thes.)

SU *i mba[ba]loin*

Morph *i*

*mbabaloin*

Ml. 014b03 edition: *.i. med(o)ntestimin* (*o* illegible on MS)

SU *.i. med[o]ntestimin*

Morph *medon*

*testimin*

*medontestimin*

**6. The Expected Morph**

For each ‘morph’ in the Morphology table, CorPH provides an ‘expected morph’. This serves, firstly, as a ‘normalised’ form for the user’s reference as to how this form should have been spelled and inflected in an idealised standard of Old Irish; and secondly, as a guidance for the CorPH editors to evaluate the deviation of the actual morph from that idealised standard and to facilitate variation tagging.

The expected morph is based on the orthography and morphology of, as said, an ideally standardised form of Old Irish, and should not be taken as reflecting or confirming the actual existence of such forms. In general one should refer to *Sengoídelc* by David Stifter (2006) and *The Early Irish Verb* by Kim McCone (2 edn. 1997) for the standardised orthography and morphology. Etymology, phonological rules and regular inflectional patterns provide guidance to the standardisation, but they are not the sole considerations. If all attestations of a given word show systematic deviation from the considerations above due to analogy or some unknown process, and the word itself is very frequent, the expected morph should be formulated according to the attested morph. e.g. the expected morph of the augment in forms of *do·gní*, *do·sluindi* and *do·léici* should be based on the attested morph, i.e. *ri* in the case of *do·gní*, *rí* in the case of *do·sluindi*, *rei* in the case of *do·léici*.

A few specific points are further clarified here:

1. The punctuation of the expected morph follows that of the morph. The correct mutation should be kept in restored for the e.m. These mean that no hyphen is used in the expected morph, either before enclitics or after nasalisation *n*.

2. The longer form *inna* for gen.sg.fem., nom.pl.fem./neut., acc.pl., gen.pl. of the definite article is always restored, save for after a preposition, e.g.

SU *na cíalla* ‘(acc.) the meanings’, morph *na*, e.m. *inna*

SU *etirna túatha* ‘between the *túath*s’, morph *na*, e.m. *na*.

3. For variants of the definite article with or without *s* after prepositions *for, tar, eter*, the e.m. follows the type of the attested morph, e.g.

SU *forsna cenéla* ‘onto the tribes’, morph *sna*, e.m. *sna*

SU *etirna túatha* ‘between the *túath*s’, morph *na*, e.m. *na*.

4. In relative clauses Class A and C infixed pronouns are both used for the 1st and 2nd persons; in these cases, the e.m. follows the class of the attested morph.

5. For (allomorph) variants within the same class of infixed pronoun, the e.m. follows the type of the attested morph (e.g. *did* or *d* in Class C 3sg.neut.), but the vowel in the 1st and 2nd persons is always spelled *o* (*dom, tot, don, tob* etc., even *tom* though that is rarer), in the 3rd persons excluding Class C 3sg.masc./neut.) always *a* (*a, da, ta* etc.).

6. The lower vowels *o* is used in prepositions and preverbs in a pretonic complex, e.g. *do(nda·)*, *no(s·), ceso, maso* etc., but *i* used in *di* (=*do* and *de*) before *nd* and *a*.

7. For the copula: The present subjunctive conjunct will be treated like the allomorphs of the infixed pronouns Class C, i.e. following the type of the actual morph *(p, dip, po*) subject to orthographic and the low vowel rules (i.e. *p* instead of *b* between vowels, *o* instead of *u*). With the conjunction *ce* ‘though’ it is written *cid*. The past subjunctive is *bid*.

8. 1sg. pres.ind. of -*aigidir* verbs ends in -*aigiur*.

9. 3pl. of the secondary ending is written as *tis* and *tais*, depending on the quality of the preceding consonant. The distribution of syncopated/unsyncopated forms is a question worth studying.

10. The form of the preverb *ar·* before non-syllabic infixed pronouns is written *ara*, except before *(n)d* in the 3sg., where it is written *ari*.

11. The place of augment *ro* in the verbal complex follows the hierarchy given in McCone EIV and Castillero (Trans. of the Philo. Soc. 111:1 (2013)). When the augment is placed differently in the expected morph than in the morph, the stressed part of the verbal complex needs to be adjusted according to the normal syncope and phonotactic rules. e.g. morph: *asru·lensat* ‘they have soiled’, expected morph *as·roilnesat*.

12. expected morph of all morphological/orthographical variants of the enclitic *side* is written as *side*. Details of declension can be found in Griffith ‘Irish suide / -side ‘the aforementioned’ and by comparing the actual morphs with this expected morph.