

CORPAS FILÍOCHTA CHOLÁISTE NA TRÍONÓIDE

EOIN MAC CÁRTHAIGH AGUS ELAINE UÍ DHONNCHADHA

An aidhm leis an obair atá idir lámha againn ná leagan leictreonach inchoardaithe d'fhillíocht na scol a chur ar fáil don phobal, ag tabhairt isteach a oiread agus is féidir den dá mhíle (nó mar sin) dán siollach ón tréimhse c. 1200–c. 1650 atá tagtha anuas chugainn. San obair seo táimid ag tógáil ar cheithre bhunchloch a leag scoláirí eile, agus tosóimid anseo leo sin.

Bunchlocha

An chéad bhunchloch: Is í Katharine Simms (Comhalta Emerita i Roinn na Staire, Coláiste na Tríonóide) a leag an chéad bhunchloch. Thosaigh sí ar an saothar saoil seo agus í ina scoláire iardhochtúireachta faoi stiúir Bhriain Uí Chuív i Scoil an Léinn Cheiltigh in Institiúid Ard-Léinn Bhaile Átha Cliath, nuair a chuir sí liosta na gcéadlínte i gcatalóg lámhscríbhinní Acadamh Ríoga na hÉireann ar chártaí innéacs. D'fhás a tionscnamh as cuimse thar na blianta agus chuir sí torthaí a cuid oibre i láthair an phobail i léacht reachtúil Scoil an Léinn Cheiltigh sa bhliain 2007, cúpla bliain roimh di éirí as a post mar léachtóir i gColáiste na Tríonóide.

Tá an bunachar seo le fáil ag <https://bardic.celt.dias.ie/>. Is féidir cuardach a dhéanamh faoi cheannteidil éagsúla (féach an íomhá ar an leathanach thall). Féach, mar shampla, an dán dar tús *A-nocht sgaoilid na sgola* (Bergin 1970, dán 38), a chum an mórfhile Tadhg Óg Ó hUiginn (a bhásaigh in 1448) nuair a d'imigh a dheartháir is a oide, Fearghal Ruadh, ar shlí na fírinne. Má iontráiltear an chéadlíne ar an leathanach cuardaigh, gheofar na torthaí seo a leanas: fad an dáin, an tréimhse is an cúige inar

cumadh é, na cúig lámhscríbhinn is sine ina bhfuil fáil air, ainm agus sloinne an té a chum is an té dár chum, meadaracht an dáin, eolas bibleagrafaíochta agus – ar na gnéithe is luachmhaire den bhunachar – liosta móitífeanna. Bua an-áisiúil atá ag an mbunachar ná gur mar hipearnasc a chuirtear gach eolas ar fáil. Cuir i gcás, na móitífeanna a luaitear leis an dán seo (‘education of poets’, ‘imagery for grief’ srl.) is féidir brú orthu sin agus gheofar liosta de dhánta eile ina bhfuil na móitífeanna céanna le feiceáil.

Search the Database

*Hint: to find words beginning "mac" use mac**

First Line:	<input type="text"/>	Area:	0. Any 1. Unknown 2. Ulster 3. Connacht 4. Leinster
Motif:	<input type="text"/>		
Christian Name	<input type="text"/>	Length:	0. Any 1. Short 2. Middling 3. Long 4. Very Long
	<input checked="" type="checkbox"/> Poet <input checked="" type="checkbox"/> Patron		
Surname	<input type="text"/>		
	<input checked="" type="checkbox"/> Poet <input checked="" type="checkbox"/> Patron		
Manuscript:	<input type="text"/>	Print:	<input type="text"/>
Apologue:	<input type="text"/>	Tracts:	<input type="text"/>
In Period:	12th C and earlier ▾ – Early 18th C ▾		
Meter Var:	0. Any 1. Unidentifiable 2. Deilbhidhe 3. Rann(aigheacht) Mhór 4. Rann(aigheacht) Bheag		
Class:	0. Any 30. Unknown=poem not yet read and catalogued 31. Miscellaneous 32. Religious 33. Historical		
Poem text:	<input type="text"/>		

Tá uasdátú leanúnach á dhéanamh ar an mbunachar seo ag Mícheál Hoyne san Institiúid Ard-Léinn, agus tá tuilleadh sonraí á

gcur leis an mbunachar aige. An t-uasdátú is mó a rinneadh ar an mbunachar ón am a seoladh é ná gur cuireadh téacs an dáin féin mar aguisín leis an iontráil do chuile dhán, nach mór.

An dara bunchloch: Is faoi stiúir Damian McManus i Roinn na Gaeilge, Coláiste na Tríonóide a tiomsaíodh téacsanna leictreonacha na ndánta. Fuarthas maoiniú ón gClár um Thaighde in Institiúidí Tríú Leibhéal (PRTL) chun an sprioc uailmhianach seo a bhaint amach, agus tosaíodh ar an obair féin thiar sna 1990idí. Rinneadh breis is 1,400 dán a bhí i gcló cheana féin a scanadh nó a chlósscríobh agus níos mó ná 500 dán nár foilsíodh riamh a thras-scríobh ó lámhscríbhinní. Foilsíodh an chuid is mó de na dánta a tras-scríobhadh ó lámhscríbhinní i bhfoirm leabhair freisin, in *A Bardic Miscellany: Five hundred bardic poems from manuscripts in Irish and British libraries* (eag. McManus agus Ó Raghallaigh 2010).

Forbraíodh an corpas seo agus bunachar Simms ag an am céanna agus as lámh a chéile, agus réitíonn an dá acmhainn le chéile ar gach bealach. Ag an leibhéal is bunúsaí, mar shampla, liosta ó bhunachar Simms a bhí mar phríomhthreoir agus téacsanna an chorpais á dtiomsú, agus chuidigh an obair thiomsaithe le cur leis an liosta sin. (Gheofar cur síos níos iomláine air seo sa réamhrá le *A Bardic Miscellany*.)

An tríú bunchloch: Rud eile a rinneadh i Roinn na Gaeilge, Coláiste na Tríonóide ná gur ullmhaigh Damian McManus leagan leictreonach de na solaoídí ar fad ó thráchtas ghramádaí agus chomhréire na bhfilí (eag. Bergin 1916–1955, McKenna 1944). Seo inár ndiaidh, mar eolas don léitheoir, sampla d’iontráil sa tráchtas faoi ainmfhocail, *Irish Grammatical Tracts II*, §97. Ar dtús, déantar an ceannfhocal a dhíochlaonadh:

CEÓ .b. don cháigh, méd na cíach, na cíacha, dona cíachaib, méd na cíach, íar chácha. [Is é sin, le normalú ar an litriú: *Ceó* (‘mist’) baininsneach, tabharthach uatha *ciaigh*, ginideach uatha *ciach*, ainmneach iolra *ciacha*, tabharthach iolra *ciachaibh*, ginideach iolra *ciach*, áinsíoch iolra *ciacha*.]

Ansin, tugtar liosta d'ainmfhocail eile a leanann an patrún céanna:

Ceó, gleó, mí ... caí (ón eón), brú ... [Mar shampla, díochlaontar *caí* (ón eón) mar seo (le normalú ar an litriú): Ainmneach uatha *caoi* ('a cuckoo'), tabharthach uatha *cuaigh*, ginideach uatha *cuach*, ainmneach iolra *cuacha*, tabharthach iolra *cuachaibh*, ginideach iolra *cuach*, áinsíoch iolra *cuacha*.]

Ansin, tugtar solaoidí ó dhánta a léiríonn an úsáid cheart (nó, scaití, an úsáid mhícheart). I measc na solaoidí sa chás seo tá an leathrann seo (sampla comhardaidh 1768), a léiríonn foirm an ghinidigh uatha:

Do bhí adhbhar farbh fhúath libh / labradh na cuach do chloistin. [Is é sin (féach McManus 1997: 92), línte 5cd den dán a luadh thuas, *A-nocht sgaoilid na sgola*, a aistríonn Bergin (1970, 280), ag a bhfuil an ginideach iolra, *gcuach*, mar 'well might ye loathe to hear the utterance of the cuckoos'.]

Tá na solaoidí ar fáil anois mar chomhad téacs ag <https://www.tcd.ie/Irish/research/database.ga.php>. Tá tábhacht ar leith leo seo mar fhoinsé údarásach eolais faoin Nua-Ghaeilge Chlasaiceach (teanga chaighdeánach na bhfilí) agus tá siad le cur lenár gcorpas aontaithe ag <https://app.sketchengine.eu/#open> ar ball.

An ceathrú bunchloch: An ceathrú bunchloch ná na huirlisí ríomhtheangeolaíochta a bhí ann sular thosaigh muid ar an tionscnamh seo, agus beidh tuilleadh le rá againn fúthu sin ar ball.

Buntáistí

Cabhróidh cúpla cás-staidéar le léiriú a thabhairt ar na buntáistí a bhaineann le corpas leictreonach don taighdeoir.

Sular tosaíodh ar an obair seo, bhí an taighdeoir a bhí ag cuardach samplaí d'fhocail san fhilíocht i dtuilleamaí Fhoclóir an

Acadaimh (Quin 1976) agus innéacsanna focal i leabhair ar nós *Dioghluim Dána* (Mac Cionnaith 1938) agus, nuair nár leor iad sin, chaithfí léamh trí na dánta féin ar lorg samplaí.

Cuir i gcás, is mar seo ba chóir an t-alt + *uile* + ainmfhocal a úsáid de réir *Irish Grammatical Tracts* II, §20 (leis an eolas san ord a chonaic muid thuas – paraidím, focail eile a leanann an patrún céanna, solaoídí ó dhánta):

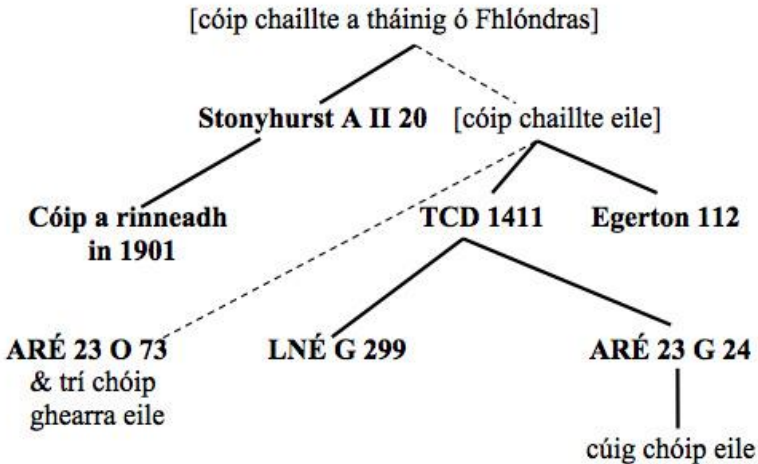
IN UILE FHER, don uile fher, méd na huile (*var.* nuile) fher, na huile fher, dona huilibh fhear, meic (*var.* mic) na n-uile bhfer, marbh na huile fher (*var.* fhera).

In uile fher, in uile dhúl, in uile dhúileadh, inann ghabhaid (*var.* mar sin).

Mó iná feidm duine don Dúileamh . deilbh na n-uile dhúiledh dhó
A trénlámh an uili fhear . a éngrádh Muiiri Maighdean .I.

In alt ar an ábhar seo (Mac Cárthaigh 1998), léiríodh go bhfuil lochtanna éagsúla ar an bparaidím thuas agus, chomh maith leis sin, nach mbeadh glacadh ag gach scoil filíochta leis gur chóir *.l.* (is é sin *lochtach*) a chur leis an dara solaoid. Rinneadh sin ar bhonn 93 sampla ó dhánta, ar bailíodh an chuid is mó acu trí léamh trí chnuasaigh fhoilsithe filíochta, dán ar dhán. Thóg an cuardach sin cúpla mí. Cúpla nóiméad a thógfadh sé anois, le corpas leictreonach anótáilte agus bogearraí cuardaigh saincheaptha, agus d’aimseofaí níos mó ná 93 sampla.

Cás-staidéar eile a léiríonn an tábhacht le corpas incheardaithe mar uirlis d’uirlisí an taighdeora teanga ná deacracht meadarachta i líne den dán *Do toirneadh ceannas chlann gCuinn*, a cumadh thart ar an mbliain 1655 agus atá le foilsiú in *Ériu* 69 (Mac Cárthaigh, le teacht). Tá an dán ar marthain i gcúig lámhscríbhinn déag:



Seo rann 32 (le normalú ar an litriú) mar atá sna trí chóip neamhspleácha lámhscríbhinne ag a bhfuil an rann seo a tháinig anuas chugainn (Stonyhurst, TCD agus Egerton, iad uile ón 18ú haois):

Do hinnleadh géirghleó goimhe
ar lár Mhoighe Mucraimhe
sul ráinig a n-eadráin de,
teagmháil dá dtáinig turbhaidh.

An deacracht ná *de* sa tríú líne. Ní dhéanann sé uaim le *eadráin* ná rinn agus airdrinn le *turbhaidh*. Is é a éilíonn an mheadarach anseo ná focal ina bhfuil (*fh* +) *ai* / *oi* / *ui* + *l* / *n* / *r* / *mh* / *bh* / *gh* / *dh*. Tugann cuardach sciobtha sa chorpas na samplaí seo d'*eadráin* + forainm réamhfhoclach:

do shaoil gan eadráin oraibh
teagmháil daoibh re Danoraibh (Mac Airt 1944, línte 4116–7)

ar son a eadrána air
ar ccor theagmhála um Theamraigh (McKenna 1951, línte 1079–80)

re fear a eadrána air,
fear na tegm(h)ála ar tosaigh (Ó Macháin 1986: 70).

Iontu seo ar fad tá úsáid an fhorainm réamhfhoclaigh *oraibh / air* deimhnithe ag an meadaracht (ag uaim leis an bhfocal roimhe agus ag rinn agus airdrinn leis an bhfocal ag deireadh na líne dár gcionn). Shásódh coigeartú ó *de go air* riachtanais na meadarachta inár ndán-na, agus thioctadh sé leis seo ó *DIL* (Quin 1976) s.v. *etráin II*:

An interfering on behalf of someone; a defending, protecting (usually with ar, for of pers. or thing warded off)

Ach an coigeartú sin a dhéanamh, is féidir an rann a thuiscint mar seo:

Before they could be held back from it, lines were drawn up for sharp venomous battle at Magh Mucraimhe, an encounter which led to disaster.

Thángthas ar na samplaí thuas ón gcorpas trí chuardach simplí amháin a dhéanamh don fhocal *eadráin*. (Munar léir láithreach cén réamhfhocal is minice a fhaightear leis sin, d'fhéadfaí cuardach ní ba chasta a dhéanamh a chuimseodh focail eile timpeall air.) Toisc go bhfuil anótáil (clibeáil) déanta ar shamplaí de *eadráin* sa corpas, ní raibh gá le cuardach ar leith chun teacht ar an nginideach uatha *eadrána* ná chun athrú tosaigh a chur san áireamh. Míneofar thíos cén chaoi a ndearnadh an anótáil seo agus cé na prionsabail ar a bhfuil sí bunaithe.

Bunachar agus corpas le chéile

An tslí is éasca chun é seo a léiriú ná sampla a thabhairt. Ag seo an chéad rann den chéad dán in McManus agus Ó Raghallaigh 2010 (le moltaí na bhfear eagair arna gcur i bhfeidhm ar an téacs):

A bhean do bhi san Bheththil(l)
dar bhfaire-ne ar bhorr[i]thín
 gé fuairis fer nuadhe [a-]nocht
 mo truaighe a bean do bochtacht.

(D'fhéadfaí sin a aistriú go Béarla mar 'O woman who was in Bethlehem, watching over us, helping us, although you attained a new husband tonight, I pity your poverty'.)

Ag seo thíos an leagan amach inmheánach atá in úsáid againn don chéad leathrann de sin in inneall cuardaigh SketchEngine (<https://www.sketchengine.eu/>). Ag barr an tábla tá meiteashonraí don dán uile (rogha sonraí ó bhunachar Simms) agus tosaíonn an t-eolas faoi théacs an dáin féin ina ndiaidh sin, ar líne 2 den tábla:

1	<code><doc POEM_ID="P10" Collection="ABM" FIRSTLINE="A bhean do bhi san Bheithil" AREA="Ulster" METRE="Deibhidhe" PATRON_CNAME="BVM" PATRON_SNAME="saint" PERIOD="17th" POET_CNAME="Giollaphadraig (Og)" MS="RIA10" POET_SNAME="Mac an Bhaire" checked="N" ></code>			
2	<code><lg type="verse" number="1"></code>			
3	<code><l></code>			
4	A	A	a-U	Uv
5	bhean	bhean	bean-N	Ncfsv
6	do	do	do-Q	Q
7	bhi	bhí	bí-V	Vmis
8	san	san	i-S	Spa-s
9	Bheththil (l)	Bheithil	Beithil-N	Npfsc
10	<code></l></code>			
11	<code><l></code>			
12	dar	dár	dár-S	Spp
13	bhfaire-ne	bhfaire-ne	faire-N	Nvfsc
14	ar	ár	ár-D	Dp1-p
15	bhorr[i]thín	bhfóirithint	fóirithint-N	Nvfsc
16	<code></l></code>			

Tar éis na meiteashonraí ar an gcéad líne den tábla, léiríonn an chlib <lg> ('line group') ar líne 2 den tábla go bhfuil rann 1 den dán le teacht, léiríonn <l> ('line') ar líne 3 den tábla go bhfuil an chéad líne de rann 1 le teacht, agus léiríonn </l> ('line end') ar líne 10 den tábla go bhfuil deireadh leis an gcéad líne den rann.

Seachas na clibeanna sin, tá líne ar leith sa tábla ag gach téacschomhartha ('token', .i. focal nó comhartha poncaíochta) sa dán. Feicfear go bhfuil eolas breise curtha anseo le gach téacschomhartha: litriú caighdeánach (i gCaighdeán Oifigiúil an lae inniu), leama caighdeánach agus roinn chainte an leagain chaighdeánaigh. Ar líne 15, mar shampla, tá an téacschomhartha *bhorr[i]thín* (díreach mar atá sa téacs foilsithe); ansin sa chéad cholún eile tá an leagan caighdeánach (i gCaighdeán Oifigiúil an lae inniu) *bhfóirithint*; ansin tá an leama caighdeánach *fóirithint* agus an roinn chainte ('N' sa chás seo, .i. 'noun'); agus, ar deireadh, tá roinn chainte mhionsonraithe 'Nvpsc' (.i. 'noun, verbal, feminine, singular, common case', féach <https://www.scss.tcd.ie/~uidhonne/irish.utf8.htm>).

An buntáiste is mó leis an gcur chuige seo don taighdeoir i ngort fhilíocht na scol ná gur féidir cuardach a dhéanamh faoi leamaí (mar shampla, *bí i*, seachas *do bhí san* nó *a-tá sna* nó *ní fhuil insan* nó na céadta cruth eile ina bhfaightear na leamaí seo le chéile) agus gur féidir meiteashonraí a úsáid sa chuardach (cuir i gcás samplaí de *bí + i* sa 17ú haois a lorg nó *bí + i* i ndánta mhuintir Uiginn, abramis, a fhiosrú).

Píblíne phróiseála

Bun agus barr na próiseála atá taobh thiar den tábla thuas ná clibeáil ranna cainte (nodaireacht ghramadaí a chur le gach focal) agus leamú (gach focal a bheith inchoardaithe faoina bhunfhoirm, gan athrú tosaigh, infhilleadh ná a leithéid). Bíonn clibeáil agus leamú casta i gcónaí ach tá castacht bhreise leo i gcás fhilíocht na scol – a ghlacadh go minic le foirmeacha éagsúla d'aon fhocal

amháin (mar shampla, *abhsdal*, *absdal*, *asbal*, *asbhal*, *eabhsdal*, *eabsdal*, *easbal* agus *easbhal* le haghaidh *aspal* an lae inniu), a bhí an-solúbtha maidir le litriú (mar shampla, d'fhéadfaí *o* nó *u* a scríobh in áit *a* sa dara siolla de *abhsdal*, *absdal* srl.) agus a tháinig anuas chugainn i bhfoinsí ina bhfaightear an-éagsúlacht ortagrafaíochta (mar shampla, seans gur mar *epstal* a gheofaí *eabsdal* i lámhscríbhinn).

Chun aghaidh a thabhairt ar na dúshlán seo, agus de cheal na n-acmhainní a ligfeadh dúinn tosú as an nua ar fad, shocraigh muid gur i gCaighdeán Oifigiúil an lae inniu a bheadh na leamaí. Shocraigh muid freisin go dtosóimis le huirlisí ríomhtheangeolaíochta a bhí ann cheana féin is a forbraíodh ar dtús don Ghaeilge iarchlasaiceach – ag cur leo de réir mar ba ghá chun iad a chur in oiriúint do theanga fhilíocht na scol. Ag seo ár bpíblíne phróiseála le haghaidh clibeáil ranna cainte: Téacsanna > Téacschomharthóir > Caighdeánaitheoir leasaithe > Anailíseoir moirfeolaíochta > Imdhealaitheoir > Corpas anótáilte.

Téacschomharthóir: An chéad chéim ná téacs gach dáin a bhriseadh ina théacschomharthaí, is é sin ina fhocail agus ina chomharthaí poncaíochta, mar atá le feiceáil sa chéad cholún sa tábla thuas. Dúshlán breise a bhaineann le téacsanna fhilíocht na scol ná an chiall speisialta atá le lúibíní in eagráin leathdhioplómaitiúla de dhánta, a bhfuil breis is 500 acu sa chorpas. Mar shampla, ciallaíonn na lúibíní cruinne in *Bheththil(l)* sa chéad líne den rann thuas go bhfuil an dara *l* scríofa sa lámhscríbhinn ach nár theastaigh sé ón bhfile féin agus gur chóir é a bhaint amach, agus ciallaíonn na lúibíní cearnacha in *[a-]nocht* sa tríú líne den rann nach bhfuil *a-* sa lámhscríbhinn ach gur theastaigh sé ón bhfile agus gur chóir é a chur isteach. Dá scarfaí an phoncaíocht ó na focail sna cásanna seo, mar a dhéanann téacschomharthóirí de ghnáth, chuirfí na focail as a riocht ar fad.

Caighdeánaitheoir leasaithe: Is é seo a chuireann ar fáil an dara colún sa tábla thuas (ina n-aistrítear a bhfuil sa bhuntéacs go Caighdeán an lae inniu). An bhnuirlis phróiseála atá in úsáid againn anseo ná an Caighdeánaitheoir a chruthaigh Kevin

Scannell (2009), a forbraíodh le téacsanna ó Thobar na Gaedhilge a chaighdeánú do Nua-Chorpas na hÉireann (Kilgarriff, Rundell agus Uí Dhonnchadha 2007) agus arís, ina dhiaidh sin, do théacsanna Chorpas Stairiúil na Gaeilge 1600–1926 (Uí Dhonnchadha, Scannell *et al.* 2014, agus féach <http://corpas.ria.ie/>). Tá trí straitéis chaighdeánaithe i gceist: (a) rialacha caighdeánaithe lámhdhéanta, (b) liosta péirí ina gceanglaítear leaganacha stairiúla d’fhocail le leamaí i gCaighdeán an lae inniu, agus (c) rialacha caighdeánaithe a fhoghlaimítear go huathoibríoch ó chorpais chomhthreomhara (ina bhfuil téacs neamhchaighdeánach agus leagan caighdeánach den téacs céanna taobh le taobh).

Is féidir (a) agus (b) a chur in oiriúint do chorpas ar leith. Ar rialacha lámhdhéanta nua atá cruthaithe againn do chorpas fhilíocht na scol tá, mar shampla, rialacha ar nós *sg-* ag tús focail = *sc-*, *cc-* = *gc-* agus *tt-* = *dt-*. Díreach mar a chuir coiste eagarthóireachta Fhoclóir Stairiúil na Gaeilge breis agus 13,000 péire focal nua ar fáil don tréimhse 1600–1926, tá breis is 5,000 péire nua cruthaithe againne go dtí seo do chorpas fhilíocht na scol, ar nós *fer / fear*, *nuadhe / nua*, *baoghal / baol* srl. Táimid ag leanúint ar aghaidh leis an obair seo, le cabhair ó Cholm Ó Cuaig, Mícheál Hoyne, Philip Mac a’ Ghoill, Damian McManus, Deirdre Nic Chárthaigh agus Eoghan Ó Raghallaigh. An modh oibre atá againn ná tosú ag barr liosta in ord minicíochta de na focail sa chorpas nach n-aithníonn an Caighdeánaitheoir leasaithe ná an tAnailíseoir moirfeolaíochta.

Anailíseoir moirfeolaíochta: Déanann sé seo anailís ar aschur an Chaighdeánaitheora (is é sin, ar a bhfuil sa dara colún sa tábla thuas, ina bhfuil leagan caighdeánaithe den téacs) agus tosaíonn ar an obair a chuirfidh ar fáil an t-eolas sa tríú colún sa tábla (leama agus roinn chainte) agus sa cheathrú colún (roinn chainte mhionsonraithe). Sa chás go bhféadfadh níos mó ná freagra amháin a bheith i gceist, liostálann sé na roghanna uile. Mar shampla, is é seo a sholáthraíonn sé i gcás na bhfocal ag tús an dáin sa tábla thuas:

```
"<A>"
    "an" Art Sg Def ^A^
    "a" Det Poss 3P Sg Masc ^A^
    "a" Det Poss 3P Sg Fem ^A^
    "a" Det Poss 3P Pl ^A^
    "a" Part Voc ^A^
    "a" Part Vb Rel ^A^
    "a" Part Deg ^A^
    "a" Part Inf ^A^
    "a" Part Nm ^A^
"<bhean>"
    "bean" Noun Fem Com Sg DefArt ^bhean^
    "bean" Noun Fem Com Sg Len ^bhean^
    "bean" Noun Fem Voc Sg Len ^bhean^
"<do>"
    "do" Prep Simp ^do^
    "do" Det Poss 2P Sg ^do^
    "do" Part Vb ^do^
"<bhí>"
    "bí" Verb VI PastInd Len ^bhi^
    "bí" Noun Fem Com Sg DefArt ^bhi^
```

Imdhealaitheoir: Úsáideann sé seo rialacha comhréire chun an anailís cheart a oireann don chomhthéacs a roghnú ón liosta a sholáthraíonn an tAnailíseoir. Táimid ag baint leasa anseo as clibeálaí ranna cainte don Ghaeilge (leimeoir, anailíseoir moirfeolaíochta agus imdhealaitheoir) a dearadh don teanga chomhaimseartha (Uí Dhonnchadha 2008) agus a cuireadh i bhfeidhm cheana ar Nua-Chorpas na hÉireann (Kilgarriff, Rundell agus Uí Dhonnchadha 2007):

```
"<A>"
    "a" Part Voc ^A^
"<bhean>"
    "bean" Noun Fem Voc Sg Len ^bhean^
"<do>"
    "do" Part Vb ^do^
"<bhí>"
    "bí" Verb VI PastInd Len ^bhi^
```

Toradh na píblíne próiseála seo uile ná corpas anótáilte san fhormaid atá le feiceáil sa tábla thuas.

Tátal

Tá torthaí ár gcuid oibre go dtí seo ar fáil anois don phobal sa chorpas anótáilte inchuardaithe a gheofar ag <https://app.sketchengine.eu/#open>. Is é atá ansin faoi láthair ná na dánta as *A Bardic Miscellany* (McManus agus Ó Raghallaigh 2010). Tá na solaoídí ó na tráchtas ghradaí le cur leo sin ar ball beag agus, de réir a chéile, na dánta eile atá le léamh faoi láthair ag <https://bardic.celt.dias.ie/>.

Obair idir lámha is ea an tionscnamh seo. Is maith a thuigimid go bhfuil ceartú fós le déanamh ar théacs na ndánta féin agus ar an gclibeáil. Aithnímid freisin go mbaineann míbhuntáistí leis an gcur chuige atá againn. Tuigimid go háirithe go gceileann Caighdeán Oifigiúil an lae inniu (lena bhfuil na dánta á gceangal againn) roinnt de ghnéithe sainiúla theanga fhilíocht na scol – rud a fhágann, mar shampla, gur mar *fear*, *bean*, *fír* agus *mná* faoi seach atá na tabharthaigh *fíor*, *mnaoi*, *fearaibh* agus *mnáibh* á léamh sa SketchEngine faoi láthair – ach tá súil againn go réiteoidimid na deacrachtaí seo amach anseo.

Beidh fáilte is fiche againn roimh cheartúcháin agus roimh mholtaí ag bardic@tcd.ie.

TAGAIRTÍ

- Bergin, O. (ed.) (1916–1955) ‘Irish Grammatical Tracts’.
Forlónadh le: *Ériu* 8, *Ériu* 9, *Ériu* 10, *Ériu* 14, *Ériu* 17.
- Bergin, O. (ed.) (1970) *Irish Bardic Poetry*. Dublin: Dublin Institute for Advanced Studies.
- Kilgarriff, A., Rundell, M. agus Uí Dhonnchadha, E. (2007) ‘Efficient corpus development for lexicography: building the New Corpus for Ireland’. In: *Language Resources and Evaluation Journal* 40 (2): 127–152.
- Mac Airt, S. (ed.) (1944) *Leabhar Branach: The Book of the O’Byrnes*. Dublin: Dublin Institute for Advanced Studies.
- Mac Cáirthaigh, E. (1998) ‘Article + uile + noun and IGT ii §20’. In: *Ériu* 49: 45–70.
- Mac Cáirthaigh, E. (le teacht) ‘Gofraidh Óg Mac an Bhaird cecinit: 4. *Do toirneadh ceannas chlann gCuinn*’. In: *Ériu* 69.
- Mac Cionnaith, L. (eag.) (1938) *Dioghlúim Dána*. Baile Átha Cliath: Oifig an tSoláthair.
- McKenna, L. (ed.) (1944) *Bardic Syntactical Tracts*. Dublin: Dublin Institute for Advanced Studies.
- McKenna, L. (ed.) (1951) *The Book of O’Hara: Leabhar Í Eadhra*. Dublin: Dublin Institute for Advanced Studies.
- McManus, D. (1997) ‘The Irish Grammatical and Syntactical Tracts: A concordance of duplicated and identified citations’. In: *Ériu* 48: 83–101.
- McManus, D. agus Ó Raghallaigh, E. (ed.) (2010) *A Bardic Miscellany: Five hundred bardic poems from manuscripts in Irish and British libraries*. Baile Átha Cliath: Roinn na Gaeilge, Coláiste na Tríonóide.
- Ó Macháin, P. (1986) ‘Additional readings to some poems by Tadhg Dall Ó hUiginn’. In: *Celtica* 18: 69–76.
- Quin, E.G. (ed.) (1976) *Dictionary of the Irish language, based mainly on Old and Middle Irish materials* (dlútheagrán). Dublin: Royal Irish Academy.

- Scannell, K. (2009) ‘Standardization of corpus texts for the New English–Irish Dictionary’. Páipéar a tugadh ag comhdháil bhliantúil an NAACL, Nua Eabhrac, 22 Bealtaine 2009.
- Uí Dhonnchadha, E. (2008) ‘Part-of-speech tagging and partial parsing for Irish using finite-state transducers and constraint grammar’. Tráchtas PhD neamhfhoilsithe. Dublin City University.
- Uí Dhonnchadha, E., Scannell, K., Ó hUiginn, R., Ní Mhearraí, E., Nic Mhaoláin, M., Ó Raghallaigh, B., Toner, G., Mac Mathúna, S., D’Auria, D., Ní Ghallchobhair, E. agus O’Leary, N. (2014) ‘Corpas na Gaeilge 1882–1926: Integrating historical and modern Irish texts.’ In: Bjarnadóttir, K. (ed.), *LREC 2014 Workshop LRT4HDA: Language Resources and Technologies for Processing and Linking Historical Documents and Archives – Deploying Linked Open Data in Cultural Heritage, Reykjavik, Iceland, May 2014*. Reykjavik: 12–18.