

CORPAS FOGHLAIMEORA TEG AGUS AN PRÓIFÍLIÚ CUMAIS SA GHAELGE

In Ó Raghallaigh, E. (eag.) (2020) *Léachtaí Cholm Cille L: Téamaí agus Tionscadail Taighde*.¹

AOIFE NÍ GHLOINN

1 Réamhrá

Is minic a úsáidtear na téarmaí ‘bunleibhéal’, ‘meánleibhéal’ agus ‘ardleibhéal’ le cur síos a dhéanamh ar an gcumas teanga a bhíonn ag foghlaimoirí Gaeilge nó ar na cúrsaí a chuirtear ar fáil dóibh. Ach céard go díreach atá i gceist leis na leibhéil sin? An féidir iad a shainiú? Cén chaoi ar féidir iad a aithint agus a idirdhealú óna chéile? An ionann an tuiscint atá againn ar fad orthu? Seo cuid de na ceisteanna atá ag croílár an tionscadail a luaitear i dteideal an ailt seo, tionscadal atá cómhaoinithe ag an gComhairle um Oideachas Gaeltachta agus Gaelscolaíochta (COGG) agus ag Lárionad na Gaeilge: Taighde, Teagasc agus Tástáil, Ollscoil Mhá Nuad.² Is obair é seo atá fós idir lámha; mar sin, ní bheidh freagraí na gceisteanna thuas sa pháipéar seo, ach déanfar plé ar chúlra an tionscadail, an fáth ar cheart na ceisteanna seo a chur, agus na himpleachtaí a bhaineann leo d’fhoghlaim, do mhúineadh agus do mheasúnú na Gaeilge ag gach leibhéal den chóras oideachais. Sa dara cuid den pháipéar, déanfar cur síos ar an obair atá idir lámha agus ar ghnéithe den mhodheolaíocht atá in úsáid leis na ceisteanna seo a fhiosrú.

2 Cúlra agus comhthéacs

2.1 An Fráma Tagartha Comónta Eorpach (FTCE)

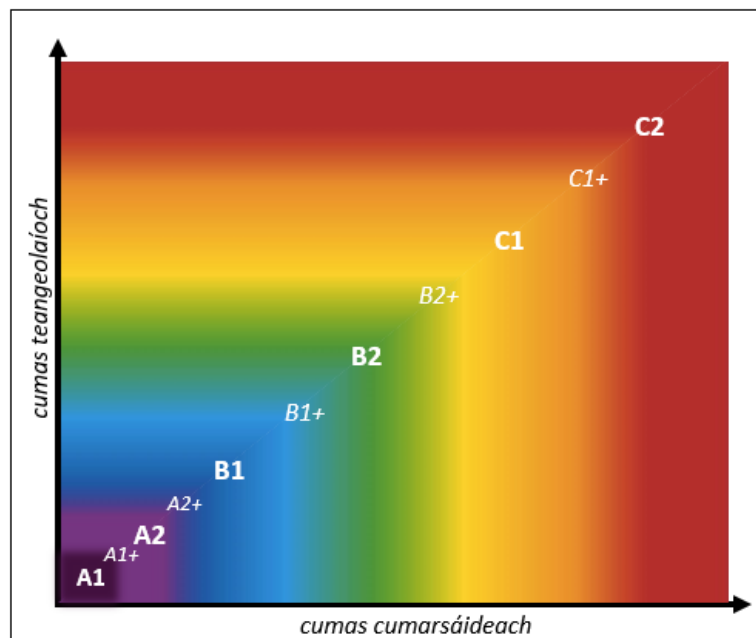
Is dócha gurb é an FTCE (Council of Europe 2001) an córas is mó a úsáidtear go hidirnáisiúnta sa lá atá inniu ann le cur síos a dhéanamh ar an gcumas teanga agus ar leibhéil éagsúla chumais. Ó foilsíodh an FTCE sa bhliain 2001, tá sé aistrithe go 40 teanga agus tá sé tar éis tionchar nach beag a imirt ar mhúineadh, ar fhoghlaim agus ar mheasúnú teangacha san Eoraip agus níos faide i gcéin (Figueras 2012, Deygers *et al.* 2017). Tá an FTCE bunaithe ar 30 bliain de thaighde ar an teangeolaíocht fheidhmeach; lena n-áirítear: teoiric na ngníomhartha cainte (Austin 1962, Searle 1969), siollabais na bhfeidhmeanna agus na nóisean (Wilkins 1976), taighde ar an inniúlacht chumarsáideach (Hymes 1972, Canale & Swain 1980, Bachman 1990, Bachman & Palmer 1996), an teangeolaíocht fheidhmiúil shistéamach (Halliday 1994) agus an fhoghlaim thascbhunaithe (Willis 1996, Skehan 1996). Ceann de na haidhmeanna a bhain leis an FTCE ná príomhléargais an taighde sin a thabhairt le chéile, chun an dea-chleachtas i bhfoghlaim, i múineadh agus i measúnú teangacha a chothú. Bhí sé mar sprioc aige, go háirithe, an foghlaimoir a chumasú le bheith ina úsáideoir teanga, trí thús áite a thabhairt do riachtanais chumarsáide an fhoghlaimora féin agus don fhoghlaim fhéinriartha. Mar is léir ó theideal na cáipéise, bhíothas ag iarraidh, freisin, téarmaíocht agus pointí comónta tagartha a thabhairt do lucht teagaisc, measúnaithe agus forbartha curaclam, rud a chuirfeadh leis an trédhearcacht agus leis an incomparáideacht a bhainfeadh le cúrsaí agus cáilíochtaí teanga ar fud na hEorpa, ó institiúid go hinstitiúid agus ó thír go tír (North 2014).

1 Tagairt iomlán: Ní Ghloinn, A. (2020). ‘Corpas Foghlaimora TEG agus an Próifiliú Cumais sa Ghaeilge.’ In E. Ó Raghallaigh (eag.), *Léachtaí Cholm Cille L: Téamaí agus Tionscadail Taighde*. Maigh Nuad: An Sagart, 119–156.

² Is tionscadal dochtúireachta é seo faoi stiúir an Dr Elaine Uí Dhonnchadha, Coláiste na Tríonóide, agus tá an Dr Anne O’Keeffe, Coláiste Mhuire gan Smál, Luimneach, ina stiúrthóir cunta ar an obair.

2.2 Leibhéal an FTCE

Ar mhaithe leis an sprioc dheireanach sin a bhaint amach, shonraigh an FTCE sé phríomhleibhéal cumais a bhíonn ag foghlaimeoirí teanga: Bunleibhéal nó ‘Bunúsáideoir’ (A1 agus A2), Meánleibhéal nó ‘Úsáideoir neamhspleách’ (B1 agus B2) agus Ardleibhéal nó ‘Úsáideoir oilte’ (C1 agus C2). Cé go bhfuil i bhfad níos mó i gceist leis an FTCE ná na leibhéal seo, is dócha gurb í sin an chuid den cháipéis is mó a tharraing aird lucht teagaisc agus tástála teangacha, agus an ghné is mó a bhfuil cur amach uirthi sa lá atá inniu ann (Figueras 2012, North 2014). Is bandaí an-leathan iad leibhéal an FTCE, agus is féidir iad a roinnt ina bhfoleibhéal; m.sh., is féidir leibhéal B2 a roinnt ina dhá fholeibhéal, B2 agus B2+, nó B2.1 agus B2.2, mar a thugtar orthu freisin (Council of Europe 2018: 36).



Íomhá 1: Leibhéal iltoiseacha an FTCE

Déantar mionchur síos ar na leibhéal sa FTCE i bhfoirm ráitis chumais (*can-do statements*) nó tuairiscíní (*descriptors*), a chuirtear le chéile i scálaí, agus a thugann cuntas ar na rudaí is féidir leis an bhfoghlaimeoir a dhéanamh leis an teanga ag gach leibhéal cumais ó A1 go C2. Mar atá le feiceáil in Íomhá 1, is cur síos iltoiseach atá i gceist; tá gné chothrománach agus gné ingearach ag baint leis.³ Ar an gcéad dul síos, déantar cur síos ar na leibhéal chumais ó thaobh réimse agus éifeacht na cumarsáide de (an ghné chothrománach). Aithnítear 67 gníomhaíocht agus straitéis chumarsáideach agus déantar cur síos ar an gcaighdeán cumarsáide a mbeifí ag súil leis ó fhoghlaimeoirí ag gach ceann de na sé leibhéal cumais, agus iad i mbun na ngníomhaíochtaí sin.⁴ Ag bun gach scála tuairiscíní (.i. leibhéal A1–A2), tá na gnéithe is bunúsaí de cibé gníomhaíocht atá i gceist — tascanna nithiúla, intuatha a bhraitheann ar fhoirmlí béil agus ar theanga

³ Is fiú a lua nach gá go bhforbródh an dá ghné seo den chumas teanga ag an ráta céanna i gcónaí, rud a fhágann nach líne dhíreach, thrasnánach a bheadh i bpróifíl forbartha gach foghlaimeora.

⁴ Tá cur síos ar 40 gníomhaíocht chumarsáideach ar fáil i gCaibidil 4 den FTCE (Council of Europe 2001). Anuas orthu siúd, tá 27 gníomhaíocht bhreise, a bhaineann leis an idirghabháil (*mediation*) agus leis an gcumarsáid ar líne den chuid is mó, ar fáil san imleabhar nuashonraithe den FTCE (Council of Europe 2018: 54–129).

réamhullmhaithe den chuid is mó. Ag barr gach scála (.i. leibhéal C1–C2), bíonn an foghlaimoir in ann ábhair chasta agus theibí a láimhseáil ar bhealach soiléir, éifeachtach agus solúbtha, réim na teanga a chur in oiriúint do chomhthéacsanna éagsúla sochtheangeolaíoch, béas a léiriú agus déileáil le comhthéacsanna casta cumarsáide, ar nós idirbheartaíocht a dhéanamh, déileáil le hachrann nó achrrann a sheachaint. Feictear an fhorbairt seo go soiléir sa scála tuairiscíní in Íomhá 2, a bhaineann leis an gcaint phoiblí. Dhá bhuntáiste a bhaineann le cur síos den chineál seo ná i) go dtugann sé tús áite d'éifeacht na cumarsáide mar ghné lárnach den chumas teanga, agus ii) go bhfuil sé ginearálta go leor gur féidir é a chur in oiriúint do réimse leathan teangacha, rud a fhágann gur féidir cúrsaí, scrúduithe agus cáilíochtaí atá bunaithe ar an gcóras seo a chur i gcomparáid lena chéile.

	ADDRESSING AUDIENCES
C2	<i>Can present a complex topic confidently and articulately to an audience unfamiliar with it, structuring and adapting the talk flexibly to meet the audience's needs. Can handle difficult and even hostile questioning.</i>
C1	<i>Can give a clear, well-structured presentation of a complex subject, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples. Can handle interjections well, responding spontaneously and almost effortlessly.</i>
B2	<i>Can give a clear, systematically developed presentation, with highlighting of significant points, and relevant supporting detail. Can depart spontaneously from a prepared text and follow up interesting points raised by members of the audience, often showing remarkable fluency and ease of expression.</i>
	<i>Can give a clear, prepared presentation, giving reasons in support of or against a particular point of view and giving the advantages and disadvantages of various options. Can take a series of follow up questions with a degree of fluency and spontaneity which poses no strain for either him/herself or the audience.</i>
B1	<i>Can give a prepared straightforward presentation on a familiar topic within his/her field which is clear enough to be followed without difficulty most of the time, and in which the main points are explained with reasonable precision. Can take follow up questions, but may have to ask for repetition if the speech was rapid.</i>
A2	<i>Can give a short, rehearsed presentation on a topic pertinent to his/her everyday life, briefly give reasons and explanations for opinions, plans and actions. Can cope with a limited number of straightforward follow up questions.</i>
	<i>Can give a short, rehearsed, basic presentation on a familiar subject. Can answer straightforward follow up questions if he/she can ask for repetition and if some help with the formulation of his/her reply is possible.</i>
A1	<i>Can read a very short, rehearsed statement – e.g. to introduce a speaker, propose a toast.</i>

**Íomhá 2: Scála tuairiscíní a bhaineann le labhairt os comhair lucht éisteachta
(Council of Europe 2001: 60)**

Níl ansin ach toise amháin den chumas teanga, áfach. Ní léir ón gcur síos thuas cé na gnéithe den teanga nó na hinniúlachtaí teanga a theastódh chun na gníomhaíochtaí cumarsáideacha atá luaite a chur i gcrích. Seo an dara toise, mar sin, an ghné ingearach, theangeolaíoch. I gCaibidil 5 den FTCE,⁵ tugtar aghaidh ar na scileanna sin, ina measc, an ghramadach, an fhóineolaíocht, stór

⁵ Tá na scálaí tuairiscíní a bhaineann leis na hinniúlachtaí teanga ar fáil i gCaibidil 5 den FTCE (Council of Europe 2001) agus ar leathanach 131–144 den imleabhar nuashonraithe den FTCE (Council of Europe 2018).

focal, inniúlachtaí sochtheangeolaíocha agus pragmatacha. 13 scála tuairiscíní atá ann. Mar gheall nach gnéithe uilíocha iad seo, agus go mbraitheann siad ar an teanga áirithe atá i gceist, déantar an cur síos seo ar bhealach i bhfad níos ginearálta ná an cur síos a dhéantar ar na scileanna cumarsáide thuas, gan mionsonraí a thabhairt maidir leis na gnéithe den teanga a shealbhaítear ag gach ceann de na sé leibhéal cumais. Tá sé seo le feiceáil sa scála tuairiscíní thíos a bhaineann leis an gcrúineas gramadaí. Ar ndóigh, d'fhéadfaí an cheist a chur: céard atá i gceist le *'simple grammatical structures'* agus *'complex language'* i gcás na Gaeilge? Cén difríocht atá idir *'a relatively high degree of grammatical control'* a luaitear ag leibhéal B2 agus *'good grammatical control'* atá luaite le B2+/ B2.2? An ionann an tuiscint a bheadh againn ar fad ar na téarmaí sin? Murarbh ionann, bheadh an baol ann nach mbainfeadh gach úsáideoir an tuiscint chéanna as an ngné seo den FTCE. Bheadh impleachtaí tromchúiseacha ag baint leis sin dá mbeadh institiúidí agus eagraíochtaí éagsúla ag soláthar cúrsaí agus ag bronnadh teastas a maítear ina dtaobh go bhfuil siad bunaithe ar an FTCE, ach gan iad a bheith de réir a chéile ar chor ar bith.

	GRAMMATICAL ACCURACY
C2	<i>Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).</i>
C1	<i>Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.</i>
B2	<i>Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.</i>
	<i>Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.</i>
B1	<i>Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.</i>
	<i>Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.</i>
A2	<i>Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.</i>
A1	<i>Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.</i>

**Íomha 3: Scála tuairiscíní a bhaineann leis an gcrúineas gramadaí
(Council of Europe 2001: 114)**

2.3 Reference Level Descriptions/ Saintuairiscíní leibhéal

D'aithin údair an FTCE an baol a d'fhéadfadh a bheith ag baint leis an easpa mionsonraí seo ón gcéad lá riamh. Ní raibh sé i gceist acu go n-úsáidfí é mar chóras docht, saintreorach, ach mar cháipéis bheo, heorastúil, a spreagfadh plé agus machnamh i measc lucht a húsáide agus a mbeadh go leor saoirse agus solúbthachta ag baint léi lena cur i bhfeidhm i gcomhthéacsanna éagsúla (North 2014). Bhreathnaigh údair an FTCE ar an mbuncháipéis, mar sin, mar chreatlach lom a bheadh ina túsphointe comónta do gach teanga, agus mhol siad ón tús go rachadh speisialtóirí gach teanga i mbun oibre le feoil a chur ar na cnámha agus na tuairiscíní a mhionsonrú dá gcuid teangacha agus comhthéacsanna féin. Tugadh *Reference Level*

Descriptions (RLDs) nó ‘saintuairiscíní leibhéal’⁶ ar na tuairiscíní mionsonraithe seo, a thabharfadh cuntas níos cuimsithí ar na gnéithe den teanga (stór focal, struchtúir ghramadaí, srl.) agus ar na scileanna teanga (cruinneas, líofacht, srl.) a bhainfeadh le gach leibhéal cumais i gcás gach teanga ar leith (Council of Europe 2005, Saville & Hawkey 2010). Mar a mhíníonn Green (2010: 4):

RLDs are thus intended to mediate between the CEFR and specific contexts for its use. They offer meaningful illustrative learning objectives that more fully operationalise the CEFR descriptions for users by providing linguistic exponents.

Foilsíodh cáipéis bhreise (Council of Europe 2005) le treoir a thabhairt maidir leis na saintuairiscíní leibhéal a fhorbairt. Moladh go mbeadh bunús eolaíoch leo; go bhféadfaí taithí agus iomas seanbhunaithe lucht teagaisc na teanga a úsáid mar thúsphointe, ach go raibh gá, freisin, leis na saintuairiscíní a fhorbairt agus a bhailíochtú, bunaithe ar anailís eolaíoch ar chorpais teanga, go háirithe, corpais foghlaimeora (*learner corpora*, nó bunachair de theanga foghlaimeoirí), dá mbeadh a leithéid ar fáil.

2.4 Corpais foghlaimeora agus próifiliú cumais teanga

Tá an taighde corpais seo tosaithe anois i gcás go leor de theangacha na hEorpa. Is é an tionscadal is uailmhianáí agus is ceannródaíche atá curtha i gcrích sa réimse seo ná *English Profile*, a rinneadh in Ollscoil Cambridge.⁷ Rinne foireann taighde an tionscadail mionanailís ar Chorpas Foghlaimeora Cambridge (bunachar de 55 milliún focal, bunaithe ar scrúduithe scríofa na hinstiúide sin), chun teacht ar ‘*criterial features*’ nó sainchomharthaí leibhéal — na struchtúir ghramadaí agus stór focal is mó a thugann le fios cén leibhéal cumais atá ag foghlaimeoir Béarla agus a chabhraíonn linn idirdhealú a dhéanamh idir na leibhéil éagsúla (Hawkins & Filipović 2012). Tá saintuairiscíní taighde-bhunaithe curtha ar fáil freisin do bhreis agus 20 teanga eile; ina measc: an Ghearmáinis (Glaboniat *et al.* 2005); an Iodáilis (Spinelli & Parizzi 2010); an Ioruais (Carlsen 2013); agus an Ghearmáinis, an Iodáilis agus an tSeicis (Boyd *et al.* 2014).⁸

Cuireann torthaí na dtionscadal sin go mór leis an tuiscint atá againn ar leibhéil an FTCE i dteangacha éagsúla. Mar shampla, léirigh taighde Cambridge ar fhorbairt na gramadaí go méadaíonn meánfhad na habairte (.i. líon na bhfocal san abairt) i scrúduithe scríofa Cambridge ó 7.9 ag leibhéal B1, go 10.8 ag leibhéal B2, go 14.2 ag leibhéal C1, de réir mar a thagann méadú ar chumas na bhfoghlaimeoirí fochlásail agus struchtúir chasta eile a láimhseáil (Hawkins & Filipović 2012: 23). Patrún tábhachtach eile a léiríodh ná go bhfaigheann an foghlaimeoir greim ar go leor struchtúir ghramadaí ag na bunleibhéil, ach go dtagann forbairt léacsagramadúil (*lexicogrammatical*) agus phragmatach ar na struchtúir sin ina dhiaidh sin. Is é sin le rá go n-úsáideann an foghlaimeoir na struchtúir chéanna ag na leibhéil arda, ach le réimse níos leithne focal agus bríonna (O’Keeffe & Mark 2017). Ag leibhéal A1, cuir i gcás, is féidir le foghlaimeoirí Béarla an coinníollach, ‘*would*’, a úsáid leis an mbriathar ‘*like*’; ag leibhéal A2, úsáidtear réimse níos leithne briathra sa choinníollach; ag leibhéal B1, úsáidtear an coinníollach

⁶ ‘*Reference level descriptions*’ an téarma oifigiúil a úsáidtear i gcáipéisí Chomhairle na hEorpa, ach is minic a thugtar ‘*Reference level descriptors*’ orthu sa litríocht chomh maith. Tá an téarma Gaeilge atá cumtha agam anseo bunaithe ar an dara leagan sin, toisc an focal ‘tuairiscín’ a bheith níos gaire don bhrí atá leo ná ‘cur síos’ nó ‘cuntas’.

⁷ Féach <http://englishprofile.org/> [ceadaithe Eanáir 2020]

⁸ Tá liosta de na tionscadail seo ar fáil ag <https://www.coe.int/en/web/common-european-framework-reference-languages/reference-level-descriptions-rlds-developed-so-far> [ceadaithe Eanáir 2020].

foirfe, ‘*would have*’ + rangabháil chaite (m.sh., *would(n’t) have gone, met, said*); agus ag leibhéal B2, is féidir ‘*would*’ a úsáid le tagairt do ghníomhartha rialta san am atá caite, mar a úsáidtear an aimsir ghnáthchaite sa Ghaeilge (Cambridge University Press 2015). Ní hé ceartúsáid an struchtúir amháin a léiríonn an fhorbairt teanga ó leibhéal go leibhéal, mar sin, ach réimse na gcomhthéacsanna ina n-úsáidtear an struchtúr sin. Tá impleachtaí aige seo do chruinneas na teanga freisin. I gcás struchtúir áirithe ghramádaí, de réir mar a thagann méadú ar réimse a n-úsáide, tagann méadú freisin ar líon na n-earráidí a bhaineann leo. Fágann sé seo go n-úsáidtear iad ar bhealach teoranta, ach cruinn, ag leibhéil A1–A2, go dtagann méadú ar líon na n-earráidí a bhaineann leo ag leibhéil B1 nó B2, sula bhfaigheann an foghlaimoir smacht orthu arís ag leibhéil C1 nó C2. Léiríonn sé seo gur féidir breathnú ar earráidí áirithe Béarla mar chomharthaí forbartha nó mar leid go bhfuil leibhéal B1 nó B2 bainte amach ag an bhfoghlaimoir (Hawkins & Buttery 2010, Hawkins & Filipović 2012).

Cuireann an tuiscint bhreise seo go mór leis an trédhearcacht agus leis an leanúnachas a bhaineann le húsáid an FTCE i múineadh, measúnú agus teastasú teangacha éagsúla. Is réimse taighde é seo a bhfuil go leor torthaí praiticiúla eile ag baint leis freisin. Tá lucht deartha ábhar teagaisc agus acmhainní foghlama ag baint leas as na corpais foghlaimora agus na próifíilí cumais le cúrsaí, téacsleabhair, ábhair léitheoireachta, foclóirí agus áiseanna eile a fhorbairt a dhíríonn ar shainriachtanais foghlaimoirí ag leibhéil éagsúla chumais (McCarthy 2016). Tá tástálaithe teanga á n-úsáid le bailíocht a gcuid scrúduithe a dhearbhu agus a fheabhsú, agus le córais mharcála agus scrúduithe nua a fhorbairt (Callies & Götz 2015). Táthar á n-úsáid freisin le huirlisí ríomhaireachta a bhaineann le próiseáil teanga nádúrtha, ar nós ceartaitheoirí uathoibríocha gramadaí, a fhorbairt (Leacock *et al.* 2015).

3 An FTCE agus an Ghaeilge

3.1 Teastas Eorpach na Gaeilge

Ba é Teastas Eorpach na Gaeilge (TEG) an chéad sampla de chur i bhfeidhm an FTCE i gcás na Gaeilge. Is córas scrúduithe cumais é TEG atá forbartha ag Lárionad na Gaeilge: Taighde, Teagasc agus Tástáil, Ollscoil Mhá Nuad, agus atá bunaithe go ginearálta ar leibhéal A1–C1 den FTCE. Is é an t-aon chóras neamhspleách a dhéanann tástáil chaighdeánach agus teastasú ar chumas Gaeilge foghlaimoirí fásta. Tá siollabais agus ábhar teagaisc curtha ar fáil do gach leibhéal scrúdaithe, agus sonraítear liostaí téamaí, feidhmeanna cumarsáide, stór focal agus struchtúir ghramádaí mar chuid de na siollabais ag leibhéil A1–B2. Nuair a cuireadh tús le TEG sa bhliain 2004, ní raibh aon chorpas foghlaimora ar fáil a bhféadfaí na sonraíochtaí sin a bhunú air. Mar sin, bunaíodh na sonraíochtaí ar mhionanailís a rinneadh ar thuairiscíní an FTCE féin, ar shonraíochtaí a bhí curtha ar fáil don Bhéarla agus do theangacha eile cheana féin, ar liostaí d’fhocail ardmhínicíochta sa Ghaeilge, ar ábhar teagaisc a bhí ar fáil don Ghaeilge cheana féin agus ar thaithí na saineolaithe a bhí páirteach i bhforbairt na siollabas. Déanann Cumann Tástálaithe Teanga na hEorpa (*Association of Language Testers in Europe* nó ALTE) monatóireacht agus iniúchadh rialta ar scrúduithe TEG, lena chinntiú go bhfuil siad ag teacht leis an dea-chleachtas idirnáisiúnta maidir le dearadh agus riaradh na scrúduithe, oiliúint na

rátálaithe, caighdeánú na marcanna agus bailíochtú staitistiúil na dtorthaí. Tá Q-mharc ALTE bronnta ar TEG mar dhearbhu go bhfuil sé ag teacht leis na caighdeáin is airde sa tástáil teanga.⁹

Tá scrúduithe TEG ag leibhéal B1–C1 in úsáid anois ag roinnt eagraíochtaí agus institiúidí mar chuid den phróiseas iarratais i gcomhair cúrsaí acadúla, scéimeanna oiliúna agus deiseanna fostaíochta a bhaineann leis an nGaeilge. Ó 2019 i leith, cuir i gcás, caithfidh iarrthóirí ar an Máistir Gairmiúil san Oideachas (Bunmhúinteoireacht) ar a laghad 65% a bhaint amach i scrúdú béil TEG B1 sular féidir leo iarratas a dhéanamh ar an gcúrsa oiliúna féin. Tá Ollscoil na hÉireann, Coimisiún Fulbright, Na Naíonraí Gaelacha, an tSeirbhís um Cheapacháin Phoiblí agus eagraíochtaí eile tar éis úsáid a bhaint as scrúdú TEG B2 mar chuid de phróisis éagsúla iontrála agus earcaíochta. Tá TEG C1 luaite, chomh maith, mar riachtanas fágála do chúrsaí san aistriúchán agus san ateangaireacht atá cómhaoinithe ag an Údarás um Ard-Oideachas agus an Roinn Ealaíon, Oidhreacht agus Gaeltachta faoin Tionscnamh Ardscileanna Gaeilge.¹⁰

3.2 Samplaí eile d'úsáid an FTCE i gcás na Gaeilge

Ó forbraíodh TEG, tá úsáid an FTCE ag éirí i bhfad níos coitianta i múineadh na Gaeilge ag an dara agus an tríú leibhéal. Rinneadh athbhreithniú ar chúrsa Gaeilge na Sraithe Sóisearaí ag leibhéal an iar-bhunoideachais sa bhliain 2017 agus maítear go bhfuil na sonraíochtaí úra 'ailínithe, a bheag nó a mhór, leis na tuairiscíní atá le fáil i mbandaí A2–B1 ... den FTCE' i gcás scoileanna a fheidhmíonn trí mheán an Bhéarla (An Roinn Oideachais agus Scileanna 2017a: 3), agus le leibhéal B2 i gcás Gaelscoileanna agus scoileanna Gaeltachta (An Roinn Oideachais agus Scileanna 2017b: 4). Idir 2008 agus 2011, dearadh siollabas nua comónta don bhunchéim sa Ghaeilge, mar chuid de mheitheal náisiúnta idir-institiúideach.¹¹ Beartaíodh go mbeadh an siollabas bunaithe ar leibhéal B2 den FTCE, ag dul i dtreo C1 faoi dheireadh Bhliain 3 den chúrsa (Walsh & Nic Eoin 2010). Fágadh faoi na hinstiúidí aonair féin, áfach, athbhreithniú a dhéanamh ar a gcuid córas measúnaithe le go mbeidís ag teacht leis an siollabas nua agus leis an FTCE. Is cosúil gur céim é sin atá fós le cur i gcrích ar bhealach cuimsitheach ag an gcuid is mó de na hinstiúidí.

Bíonn dearbhú cumais de réir an FTCE ag teastáil anois ón té atá ag iarraidh clárú leis an gComhairle Mhúinteoireachta mar mhúinteoir Gaeilge ag leibhéal na hiarbhunscoile. Ó 2017 i leith, caithfidh múinteoirí nuacháilithe Gaeilge fianaise a chur ar fáil go bhfuil a gcuid Gaeilge ag leibhéal B2.2 (B2+), ar a laghad, sular féidir leo clárú leis an gComhairle.¹² Murab ionann agus cás na bunmhúinteoireachta atá luaite thuas, ní chuirtear aon iachall ar an dream seo scrúdú neamhspleách cumais, ar nós TEG, a dhéanamh. Is ar ranna Gaeilge na n-ollscoileanna a bhíonn an fhreagracht, mar sin, cumas Gaeilge a gcuid mac léinn a dhearbhu nó cáipéisíocht a chur ar fáil a léiríonn ionannas idir gráid áirithe sa bhunchéim Ghaeilge agus pointí áirithe ar scála an FTCE. Ar ndóigh, bheadh deacracht mhór ag baint leis seo sa chás nach mbeadh tástáil á déanamh ar chumas teanga fochéimithe ar bhealach a bheadh ag teacht leis an FTCE, nó sa chás nach mbeadh na hinstiúidí tríú leibhéal ar fad ar aon tuiscint maidir lena bhfuil i gceist le cumas teanga ag leibhéal B2.2.

⁹ Tá 17 gcritéar cáilíochta leagtha amach ag ALTE do thástálacha teanga, agus caithfidh córais tástála a bheith ag teacht le gach ceann de na critéir sin chun an Q-mharc a ghnóthú. Féach <https://www.alte.org/Setting-Standards> [ceadaithe Eanáir 2020].

¹⁰ Féach www.teg.ie [ceadaithe Eanáir 2020].

¹¹ Tá an siollabas seo ar fáil ag www.teagascnaGaeilge.ie [ceadaithe Eanáir 2020].

¹² Ní bhaineann an rialachán seo leis an nGaeilge amháin; tá an bunriachtanas céanna leagtha amach do mhúinteoirí teangacha eile. Féach <https://www.teachingcouncil.ie/en/Publications/Registration/Documents/Curricular-Subject-Requirements-after-January-2017.pdf> [ceadaithe Eanáir 2020].

3.3 Gá le sainiú a dhéanamh ar an FTCE don Ghaeilge

Mar gheall go bhfuil impleachtaí móra ag baint leis an measúnú ag leibhéil B1–C1 anois, tá tábhacht faoi leith ag baint leis an gcinnteacht, leis an trédhearcacht agus leis an leanúnachas a bhaineann le húsáid na lipéad seo. Tá géarghá, mar sin, le sainiú a dhéanamh ar na bandaí cumais sin i gcás na Gaeilge, chun na sainchomharthaí leibhéil a idirdhealaíonn B2 ó B1 agus C1, nó B2 ó B2.2, a thabhairt chun solais. Tá soiléiriú den chineál sin riachtanach le go mbeadh institiúidí agus eagraíochtaí éagsúla ar aon tuiscint maidir le leibhéil an FTCE agus le go mbeadh bailíocht ag baint leis na breithiúnais a thugann siad ar fad ar an gcumas teanga atá ag a gcuid mac léinn nó a gcuid fostaithe. Tá an tábhacht seo mínithe go soiléir ag Little (2011: 3):

... because the CEFR has brought new clarity to the world of international language tests, it is here to stay. Third parties are increasingly likely to expect universities to express their students' language learning achievement in terms of the CEFR levels; and if they do, it will not be enough to assert that graduates ... achieve this or that CEFR level. In a world that attaches great importance to quality assurance, universities will be expected to be able to justify their claims.

Tá deis ann anois tabhairt faoin sainiú seo a dhéanamh ar bhealach córasach, taighde-bhunaithe, ó tharla scrúduithe caighdeánaithe cumais, TEG, a bheith ar fáil don Ghaeilge, rud a fhágann go bhfuil stór mór samplaí ar fáil den chineál Gaeilge a bhíonn ag foghlaimoírí ag leibhéil éagsúla den FTCE. D'fhéadfaí corpas foghlaimoira de scrúduithe TEG a úsáid chun próifiliú a dhéanamh ar na sainchomharthaí a bhaineann le leibhéil éagsúla den FTCE sa Ghaeilge, rud a chabhródh le lucht a n-úsáide idirdhealú a dhéanamh eatarthu. Chuirfeadh sé seo go mór leis an trédhearcacht agus leis an leanúnachas a bhainfeadh le húsáid an FTCE i múineadh, measúnú, agus teastasú na Gaeilge, agus i ndruchtú polasaithe a bhaineann le múineadh agus foghlaim na Gaeilge ag gach leibhéil den chóras oideachais. Is é aidhm an tionscadail seo, 'Corpas Foghlaimoira TEG agus an Próifiliú Cumais sa Ghaeilge', ná tús a chur leis an obair sin, trí chorpas foghlaimoira 200,000 focal a thógáil, bunaithe ar scrúduithe scríofa agus cainte TEG ag leibhéil B1–C1, agus anótáil chuimsitheach a dhéanamh air ó thaobh na rann cainte agus na n-earráidí de. Déanfar próifiliú ansin ar an gcumas gramadaí a léirítear ag leibhéil B1–C1, .i. na foirmeacha agus na struchtúir chearta agus mhíchearta ghramadaí a úsáidtear ag gach leibhéil. Tríd an anailís sin, déanfar iarracht sainchomharthaí a aithint a chabhródh le lucht teagaisc agus tástála idirdhealú a dhéanamh idir na leibhéil sin i gcomhthéacs na Gaeilge.

4 Dearadh agus tógáil an chorpais

Is é an sainmhíniú a thugtar ar chorpais foghlaimoira ná '*electronic collections of natural or near-natural data produced by foreign or second language learners and assembled according to explicit design criteria*' (Granger *et al.* 2015: 1). Baineann na critéir dheartha sin le sampláil nó le roghnú na sonraí a bheidh sa chorpas, leis an bpróiseas tras-scríofa agus anaithnidithe, agus leis an anótáil nó clibeáil a dhéantar ar na téacsanna. Tabharfar cuntas ar na gnéithe sin den obair sa chuid eile den pháipéar seo.

4.1 Sampláil agus roghnú sonraí

Is é an chéad chéim a bhaineann le corpas a thógáil ná cinneadh a dhéanamh maidir leis na téacsanna a chuirfear san áireamh ann, ionas go mbeidh an anailís bunaithe ar thacar sonraí nó ar shampla atá cuí, cothrom agus ionadaíoch, agus a fhreagraíonn do cheisteanna an taighde. Toisc go mbaineann spriocanna an taighde seo go príomha le hanailís a dhéanamh ar an gcumas teanga, ní mór an leibhéil cumais a bhaineann le gach téacs a mheas ar bhealach bailí agus

córasach, agus tús áite a thabhairt don chritéar sin i ndearadh an chorpais. Is buntáiste mór é, mar sin, an t-ábhar a bhunú ar scrúduithe caighdeánaithe cumais, ar nós scrúduithe TEG, toisc go bhfuil rátáil chúramach déanta cheana féin ar leibhéal cumais gach téacs, agus dianphróiseas oiliúna agus monatóireachta i bhfeidhm lena chinntiú go bhfuil bailíocht agus leanúnachas ag baint leis an rátáil sin. Tá sé tábhachtach, chomh maith, dianrialú a dhéanamh ar athróa eile, ar nós chúlra agus aois an fhoghlaimora, ionas nach mbeidh aon rud eile, seachas an cumas teanga, ag imirt tionchar ar thorthaí na hanailíse. Buntáiste eile a bhaineann le corpas foghlaimora a bhunú ar scrúduithe TEG ná go bhfuil eolas úsáideach faoi aois, inscne, náisiúntacht, cúlra teanga, leibhéal oideachais agus gairm na n-iarrthóirí ar fáil ó na foirmeacha iarratais, chomh maith le heolas faoin inspreagadh atá acu cur isteach ar an scrúdú.¹³ Fágann sé sin go bhfuil meiteashonraí an-saibhir ag baint le hábhar an chorpais, meiteashonraí ar féidir iad a úsáid leis na hathróa a bhaineann leis na sonraí a rialú, lena chinntiú go bhfuil sampla ionadaíoch agus cothrom ar fáil sa chorpas.

Mar a míníodh thuas, beartaíodh an corpas a bhunú ar na leibhéil B1–C1, toisc gurb iad sin na lipéid chumais is mó atá á n-úsáid i múineadh agus measúnú na Gaeilge ag an dara agus an tríú leibhéal den chóras oideachais. Fágann sé sin go bhfuil tábhacht faoi leith ag baint leis na leibhéil sin chumais a shainiú agus a idirdhealú óna chéile. Léiríonn na meiteashonraí go mbíonn réimse aoiseanna agus cúlraí ag iarrthóirí ag leibhéil B1–C1, agus go mbíonn cúiseanna éagsúla acu tabhairt faoi na scrúduithe, idir chúiseanna pearsanta agus spriocanna gairmiúla. Beartaíodh gur cheart tús áite a thabhairt do shonraí ó na hiarrthóirí is mó a bhain le sprioc an taighde féin. Bhí dhá ghrúpa i gceist ansin: (i) daltaí iarbhunscoile 16–18 mbliana d’aois a thug faoi scrúdú B1 nó B2 ar scoil, agus iad i mBliain 4 nó 5 sa mheánscoil; (ii) iarrthóirí 18–45 bliain d’aois a rinne scrúduithe B1–C1 ar mhaithe le sprioc ghairmiúil a bhaint amach, ar nós bunriachtanais do ghairm na múinteoireachta nó an aistriúcháin a shásamh. Rinneadh cinneadh thart ar dhá thrian de shonraí an chorpais ag leibhéil B1 agus B2 a roghnú ón ábhar scrúdaithe a bhain leis an dá ghrúpa sin. Roghnaíodh na sonraí eile ó ábhar scrúdaithe daoine fásta eile a rinne TEG ar mhaithe le spéis nó sprioc phearsanta. Ar an gcaoi sin, rinneadh iarracht a chinntiú go mbeadh an sampla iomlán (.i. sonraí an chorpais ar fad) cothrom agus ionadaíoch ar na príomhghrúpaí iarrthóirí a thugann faoi na scrúduithe. Ós rud é nach ndéantar scrúdú TEG C1 in aon mheánscoil, baineann sonraí an chorpais ag an leibhéal sin le hiarrthóirí fásta amháin, agus sprioc ghairmiúil ag thart ar a leath acu agus iad ag tabhairt faoin scrúdú.

Beartaíodh athróa áirithe a chur as an áireamh sa chorpas, ionas nach mbeadh na sonraí ró-ilchineálach. Tá sé feicthe i staidéir eile den chineál seo, cuir i gcás, go mbíonn tionchar suntasach ag máthairtheanga na bhfoghlaimoirí ar shealbhú an dara teanga (Murakami & Alexopoulou 2016). Is minic a chruthaítear fochorpais, mar sin, d’ábhar a bhaineann le máthairtheangacha éagsúla chun iad a chur i gcomparáid lena chéile. Sa chás seo, bhí líon na n-iarrthóirí nárbh iad an Béarla ná an Ghaeilge a dteanga dhúchais róbheag le go bhféadfaí tionchar na máthairtheanga ar na sonraí a mheas. Bhí líon áirithe iarrthóirí as Meiriceá, as Ceanada nó as an mBreatain, a raibh an Béarla mar chéad teanga acu, ach a raibh cúlra oideachais ar leith acu ó thaobh fhoghlaim na Gaeilge de, rud a d’fhágfadh go bhféadfad tréithe faoi leith a bheith ag baint lena gcuid teanga. Rinneadh cinneadh, mar sin, na sonraí sin ar fad a chur as an áireamh.

D’ardaigh sé seo ceist maidir leis na hiarrthóirí (thart ar 15% acu ag leibhéil B1–C1) a mhaígh gurb í an Ghaeilge, nó an Ghaeilge agus an Béarla araon, a dteanga dhúchais. Ar cheart

¹³ Is i bhfoirm anaithnid atá sonraí TEG, idir na scrúduithe agus na meiteashonraí, in úsáid sa tionscadal seo. Tá tuilleadh eolais faoi chúrsaí eitice agus faoin anaithnidiú ar fáil in alt 4.2 den pháipéar seo.

iad sin a chur san áireamh nuair is corpas ‘foghlaimoora’ a bhí i gceist, nó an gcuirfidís na sonraí as a riocht? Cé gur téarma é ‘cainteoir dúchais’ a úsáidtear go han-mhinic sa ghnáthchaint, is téarma doiléir agus débhríoch é, a bhfuil fadhbanna ag baint lena shainmhíniú (Paikeday 1985, Davies 2003, O’Rourke & Ramallo 2011). Tá an scéal níos casta arís i gcás mionteanga de bharr fheiniméan an tsealbhaithe neamhiomlán nó easnamhaigh, atá coitianta go leor i measc cainteoirí dúchais mionteangacha, fiú agus an teanga acu ón gliabhán nó mar theanga phobail (Montrul 2013, Ó Curnáin 2009, Lenoach 2014). Leis an scéal a fhiosrú, rinneadh scagadh ar na marcanna a ghnóthaigh iarrthóirí ar scrúdú TEG B2 sa tréimhse 2014–16 a mhaígh go raibh an Ghaeilge acu ó dhúchas. Chonacthas gur ghnóthaigh an dream seo réimse an-leathan marcanna, idir 42% (teip) agus 100% i scrúdú cainte TEG B2 agus idir 38% agus 99% sa scrúdú iomlán. Ba léir ó na meiteashonraí eile gurbh as ceantair taobh amuigh de na Gaeltachtaí traidisiúnta go leor de na hiarrthóirí seo, ach gur bhain cuid mhaith acu le grúpaí faoi leith a thug faoi na scrúduithe i scoileanna lán-Ghaeilge. Léirigh sé seo gur athróg chasta í teanga dhúchais an iarrthóra, mar gheall ar an éagsúlacht tuiscintí a d’fhéadfadh a bheith ag na hiarrthóirí féin maidir le ciall na dtéarmaí ‘céad teanga’ nó ‘teanga dhúchais’. Bhí iarrthóirí eile i measc an tsampla chéanna arbh as ceantair Ghaeltachta iad, ach a raibh deacrachtaí faoi leith acu le struchtúir na teanga, rud a thug le fios go raibh an Ghaeilge sealbhaithe ar bhealach neamhiomlán acu. Léirigh sé seo ar fad go bhfuil deacracht ag baint le hidirdhealú dubh agus bán a dhéanamh idir cainteoir dúchais agus foghlaimoora i gcás na Gaeilge agus é sin a úsáid mar chritéar roghnaithe sonraí sa taighde. Beartaíodh, mar sin, gan iarrthóirí a mhaígh go raibh an Ghaeilge mar chéad teanga acu a chur as an áireamh sa chorpas, ach breathnú ar an athróg sin arís mar chuid den anailís.

Nuair a bhí na critéir thuas curtha i bhfeidhm, rinneadh iarracht an rogha dheireanach sonraí a dhéanamh cothrom ó thaobh inscne an iarrthóra de, nuair ab fhéidir sin a dhéanamh, agus ó thaobh bhliain an scrúdaithe de, le go mbeadh réimse ceisteanna scrúdaithe ar fáil ag gach leibhéal cumais. Tá achoimre sa tábla thíos ar líon na bhfocal atá sa chorpas ag an bpointe seo, agus ar chineál na dtascanna scríofa agus labhartha atá i gceist ag gach leibhéal cumais. Tá tascanna eagarthóireachta, aistriúcháin agus ceisteanna eile gramadaí mar chuid de pháipéir scríofa TEG ag leibhéal B2 agus C1, ach is iad na tascanna oscailte, cumarsáideacha amháin a cuireadh sa chorpas. Is tascanna treoraithe cuid de na tascanna scríofa, sa chaoi is go n-iarrtar ar an iarrthóir pointí áirithe a lua sa téacs. Cé go bhfuil difríochtaí idir na cineálacha tascanna scrúdaithe a dhéantar ag leibhéil éagsúla, tá roinnt cosúlachtaí eatarthu freisin; m.sh., bíonn litreacha agus comhfhreagrach le scríobh ag gach leibhéal agus bíonn agallamh mar chuid de gach scrúdú cainte. Cabhróidh na cosúlachtaí sin le hinchomparáideacht na dtéacsanna ag gach leibhéal.

ACHOIMRE AR SHONRAÍ AN CHORPAIS

Leibhéal	Scríobh		Caint		Iomlán
	<i>Tascanna</i>	<i>Líon na bhfocal</i>	<i>Tascanna</i>	<i>Líon na bhfocal</i>	
B1	Ríomhphost Litir neamhfoirmiúil (treoraithe)	12,757	Agallamh (5–6 nóiméad) Sraith pictiúr	41,138	53,895
B2	Litir fhoirmiúil (treoraithe) Aiste/ Alt/ Léirmheas	25,502	Agallamh (6–7 nóiméad) Plé ar phictiúr	42,188	67,690
C1	Téacs feidhmeach (litir, tuairisc, léirmheas) Aiste/ Alt	23,715	Cur i láthair (5 nóiméad) Agallamh (10 nóiméad) Plé ar phictiúr	52,401	76,116
Iomlán		61,974		135,727	197,701

4.2 Tras-scríobh agus anaithnidiú na dtéacsanna

Bhí obair mhór agus go leor dúshlán ag baint leis an ábhar seo ar fad a thras-scríobh agus a dhigitiú. Bhí an t-ábhar scríofa le scanadh ar dtús, agus na téacsanna le tras-scríobh ó na comhaid PDF. Bhí an t-ábhar labhartha ar fad le tras-scríobh ó thaifeadtaí fuaime. Ní mór a aithint nach próiseas oibiachtúil ar fad atá sa tras-scríobh; ní bhíonn súile ná cluasa aon tras-scríobhaí gan locht, agus bíonn cinntí le déanamh go minic maidir leis an rud is dóichí atá scríofa nó ráite ag an iarrthóir i gcásanna nach bhfuil an pheannaireacht nó an fuaimniú iomlán soiléir. Dúshlán amháin a tháinig chun cinn go minic le linn na hoibre sin ná an trasnú idir an mhoirfeolaíocht agus an fhóineolaíocht sa Ghaeilge. Is tríd an bhfóineolaíocht a chuirtear an mhoirfeolaíocht i bhfeidhm go minic sa chaint, rud a fhágann go mbíonn cumas fóineolaíoch ag teastáil chun an cumas gramadaí a léiriú, agus go mbíonn sé deacair an dá rud a scaradh óna chéile.¹⁴ Mar gheall air sin, bhí sé tábhachtach cleachtas áirithe a chur i bhfeidhm lena chinntiú go mbeadh na tras-scríbhinní chomh cruinn agus chomh leanúnach agus ab fhéidir. Ar dtús, forbraíodh treoirlínte an-mhionsonraithe do thras-scríobh an ábhair ar fad. Leagadh síos coinbhinsiúin chaighdeánacha tras-scríofa le déileáil le gnéithe den teanga, ar nós sosanna líonta agus neamhlíonta,¹⁵ abairtí nó focail nach gcríochnaítear, athrá, athfhoclú, féincheartú, leaganacha giorraithe, canúnacha agus iasachta, focail nó frásaí nach bhfuil soiléir sa scríobh nó sa chaint, agus araile. Rinneadh iarracht, a oiread agus ab fhéidir, cloí le coinbhinsiúin tras-scríofa atá aitheanta go hidirnáisiúnta (MacWhinney 2000), nó atá ag teacht le nósanna tras-scríofa atá in úsáid i gcorpais eile Ghaeilge.¹⁶ Ba iad an príomhthaighdeoir agus beirt chúntóirí a raibh ardchaighdeán Gaeilge acu a rinne an tras-scríobh. Rinneadh gach tras-scríbhinn a sheiceáil ar a laghad dhá uair; sheiceáil na cúntóirí tras-scríbhinní a chéile agus sheiceáil an príomhthaighdeoir gach tras-scríbhinn arís, le bheith cinnte go raibh siad chomh cruinn agus chomh dílis agus ab fhéidir don bhunábhar. Ar

¹⁴ Mura mbíonn an ‘r caol’ ag foghlaimeoir ag deireadh an fhrása ‘ag léamh an leabhair’, mar shampla, bíonn sé deacair a rá an botún gramadaí nó fuaimnithe atá i gceist, toisc go bhfuil an dá rud fite fuaite ina chéile.

¹⁵ Is ionann ‘sos líonta’ agus fuaim nach focal é (m.sh., eh, em, um) a úsáidtear le bearnaí a líonadh sa chaint. Is ionann ‘sos neamhlíonta’ agus briseadh beag sa chaint nach líontar le fuaim den chineál sin.

¹⁶ Féach nósanna tras-scríofa an chorpais labhartha *Comhrá*, atá ar fáil ag: <https://www.tcd.ie/slscs/research/projects/current/gala.php#treoir> [ceadaithe Eanáir 2020].

mhaithe le trédhearcacht agus leanúnachas na hoibre, coinníodh miontaifead ar an obair ar fad agus ar aon cheisteanna nó dúshlán a tháinig aníos le linn na hoibre, ar na cinntí ar fad a rinneadh agus ar an réasúnaíocht a bhí taobh thiar díobh.

Ar ndóigh, tá cúrsaí eitice agus cosaint sonraí an-tábhachtach sa teangeolaíocht chorpais agus sa tástáil teanga araon (Wynne 2005). Tá ráiteas ar fhoirm iarratais scrúduithe TEG a chuireann in iúl d'iarrthóirí go bhféadfaí cuid dá gcuid “sonraí pearsanta a úsáid, i bhformáid anaithnid, ar mhaithe le cúiseanna taighde agus staitistiúla”.¹⁷ Mar gheall air sin, tá polasaí cosanta sonraí ag an tionscadal seo ina ngealltar go ndéanfar sonraí an chorpais ar fad a anaithnidiú — is é sin, aon tagairtí d'ainmneacha nó d'eolas pearsanta na n-iarrthóirí, ar nós áiteanna cónaithe nó oibre, a bhaint amach as sonraí an chorpais. Bíonn dúshlán breise ag baint leis an bpróiseas seo i gcomhthéacs mionteanga ar nós na Gaeilge, áfach, mar gheall go bhfuil pobal na teanga chomh beag sin go bhféadfadh sé tarlú go mbeadh daoine inaitheanta óna gcuid cainte uaireanta, fiú dá mbainfí na hainmneacha agus an gnátheolas pearsanta amach as na téacsanna. Tá impleachtaí ag baint leis seo d'úsáid an chorpais amach anseo. Is minic a chuirtear corpais den chineál seo ar fáil don phobal, ach is dócha nach mbeifear in ann é sin a dhéanamh i gcás an chorpais seo, go háirithe i gcás na gcomhad fuaime. An réiteach atá faighte air seo, ná go mbeidh an corpas féin, gan na comhaid fuaime, ar fáil ar ardán faoi leith ar líne, agus go mbeidh lucht taighde in ann rochtain a fháil air agus cuardaigh a dhéanamh ar ghnéithe den teanga a bhfuil suim acu iontu, tríd an ardán sin, ach cead speisialta a lorg. Ní bheifear in ann scrúduithe iomlána a fheiceáil ná na comhaid fuaime a chloisteáil agus beidh comhaontú cosanta sonraí le síniú roimh ré ag an té a bheidh ag iarraidh rochtain a fháil ar an gcorpas.

4.3 Anótáil

Is ionann anótáil nó clibeáil a dhéanamh ar chorpas agus cóid nó clibeanna a chur leis na focail, struchtúir nó téacsanna atá ann, ionas go mbeifear in ann cuardach níos cuimsithí, níos casta agus níos cliste a dhéanamh ar na sonraí, agus an luach is fearr a bhaint as an gcorpas dá réir (van Rooy 2015). Braitheann an cineál anótála a theastaíonn ar cheisteanna taighde an tionscadail, ar ndóigh. Tá dhá chineál clibeála a dhéantar go minic ar chorpais foghlaimeora ar mhaithe le hanailís a dhéanamh ar an ngramadach: clibeáil ar earráidí agus clibeáil ar ranna cainte agus ar leamaí.¹⁸

4.3.1 Clibeáil ar earráidí

Tá roinnt céimeanna ag baint le clibeáil a dhéanamh ar earráidí i gcorpas. Ar dtús, caithfear na hearráidí ar fad a aithint agus leagan ceartaithe nó ‘spriocfhoirm’ a chur isteach ina ndiaidh. D'fhéadfadh an próiseas sin a bheith sách suibiachtúil, ar ndóigh, agus ardaíonn sé go leor ceisteanna. Céard is ‘earráid’ ann, mar shampla? Céard é an leagan ‘ceart’? Cén stádas atá ag leaganacha canúnacha, leaganacha malartacha, nó leaganacha nach mbaineann leis an nGaeilge chaighdeánach ach atá an-choitianta sa chaint bheo? Leis an tsuibiachtúlacht a laghdú oiread agus is féidir, agus leis an trédhearcacht agus leis an leanúnachas a chinntiú, ní mór sainmhíniú soiléir a thabhairt ar an gcoincheap ‘earráid’ i gcomhthéacs an chorpais áirithe seo. Is trí thagairt a dhéanamh do theanga an chainteora dúchais a dhéantar sin go minic. Maíonn Lennon (1991: 182), mar shampla, gurb ionann earráid agus “*a linguistic form or combination of forms which in the same context and under similar conditions of production would, in all likelihood, not be*

¹⁷ Tá foirm iarratais TEG ar fáil ag: <https://secure.teg.ie/foirm/iarratas.htm> [ceadaithe Eanáir 2020].

¹⁸ Is ionann leama agus bunfhoirm an fhocail nó an ceannfhocal a bheadh le feiceáil i bhfoclóir.

produced by the speakers' native speaker counterparts.” Tá deacracht mhór ag baint le sainmhíniú den chineál sin i gcás na Gaeilge, ar ndóigh, toisc a dheacra a bhíonn sé sainmhíniú a thabhairt ar an gcainteoir dúchais féin, agus toisc go bhfuil réimse leathan cumais agus cruinnis le fáil i dteanga na gcainteoirí sin, mar atá luaite thuas agus pléite go mion ag údair eile (Ó Curnáin 2009, Lenoach 2014). Fágann sé sin gur dócha nach slat chuí tomhais i gcónaí é an ‘cainteoir dúchais’ ar an gcruinneas i gcás na Gaeilge. D’fhéadfaí earráidí a shainmhíniú trí thagairt a dhéanamh don Chaighdeán Oifigiúil chomh maith, ar ndóigh. Bíonn deacracht ag baint leis sin freisin, áfach, toisc go bhfuil dhá leagan den Chaighdeán Oifigiúil ann (Rannóg an Aistriúcháin 1958, Seirbhís Thithe an Oireachtais 2017), agus roinnt difríochtaí eatarthu maidir leis na foirmeacha atá ceadaithe iontu. I gcás rialacha gramadaí atá athraithe nó simplithe sa leagan is déanaí den Chaighdeán, ní féidir a rá gur earráid atá i gceist, má chloíonn scríbhneoir nó cainteoir leis an nós seanbhunaithe.¹⁹ Is caighdeán scríofa atá sa Chaighdeán Oifigiúil freisin, rud a fhágann go bhfuil roinnt mhaith leaganacha agus foirmeacha a bhfuil glacadh forleathan leo sa chaint bheo, nach bhfuil aitheanta sa Chaighdeán. Ní treoir fheiliúnach i gcónaí é, mar sin, i gcás na Gaeilge labhartha (Ó Baoill 2000, Ní Ghearáin 2019). Deacracht eile a bhaineann leis an gCaighdeán Oifigiúil ná go bpléann sé, den chuid is mó, leis an litriú agus leis an moirfeolaíocht, seachas leis an gcomhréir agus ceartúsáid focal (Mac Mathúna 2008). Fágann sé sin go bhfuil bearnaí móra ann agus muid ag iarraidh cinntí a dhéanamh maidir leis an rud atá cruinn nó míchruinn i dteanga an fhoghlaimora. Tá réitigh éagsúla ar an bhfadhb seo le feiceáil sa taighde atá déanta go dtí seo ar earráidí na Gaeilge. Rinne Ó Duibhir (2009) sainmhíniú ar earráidí bunaithe ar bhreithiúnais triúr cainteoirí dúchais. Glacadh leis go raibh ráiteas míchruinn sa chás go raibh an triúr cainteoirí dúchais ar aon tuairim go raibh earráid ann. Sheachain Péterváry *et al.* (2014) an téarma ‘earráid’, ach thug siad na téarmaí ‘seachfhoirm’, ‘leagan míchuí’ agus ‘seach-leagan’, ar fhoirmeacha agus ar struchtúir a bhí míchruinn de réir leaganacha éagsúla den Ghaeilge, .i. an Gaeilge thraidisiúnta, iarthraidisiúnta, chaighdeánach, laghdaithe, srl. Cé gur bealaí an-phointeáilte iad an dá chur chuige sin le sainmhíniú a dhéanamh ar earráidí, is dócha go mbeadh dúshlán ag baint lena gcur i bhfeidhm ar bhealach leanúnach i gcorpas mór foghlaimora. Ag deireadh an lae, tá gá le cothromaíocht a aimsiú idir an rud atá pointeáilte agus an rud atá praiticiúil. Mar a deir McEnery *et al.* (2006: 73): ‘*Corpus-building is of necessity a marriage of perfection and pragmatism...*’. Is é an cur chuige a beartaíodh don tionscadal seo, mar sin, ná ceithre phríomhfhoinsé thagartha a úsáid mar threoir maidir le cruinneas na Gaeilge; *An Caighdeán Oifigiúil* (1958, 2017), *Foclóir Gaeilge–Béarla* (Ó Dónaill 1977) agus *Graiméar Gaeilge na mBráithre Críostaí* (1999). Níor cuireadh clib earráide le haon fhoirm nó struchtúr atá aitheanta ag ceann de na foinsí sin. I gcás leaganacha canúnacha, ar nós ‘sa’ móide urú, níor marcáladh mar earráidí iad má bhí an córas céanna in úsáid go leanúnach sa téacs. I gcás foirmeacha agus struchtúir eile chomhréire nach bhfuil clúdaithe sna foinsí thuas agus a rabhtas idir dhá chomhairle ar cheart iad a mharcáil mar earráidí, ceadaíodh corpais eile, ar nós *Nua-Chorpas na hÉireann* (NCI), le húsáid na bhfoirmeacha nó na struchtúr sin a fhiosrú (Kilgarriff *et al.* 2006).²⁰ Sa chás go raibh foirm nó struchtúr áirithe in úsáid go forleathan sa chorpas sin, i réimse leathan téacsanna agus ag cainteoirí agus scríbhneoirí oilte Gaeilge, níor

¹⁹ Féach, mar shampla, na rialacha a bhaineann le foirm an ainmnigh in ionad an ghinidigh, 3.3.5 ar lch 48, nó córas na n-uimhreacha, 9.3.10 ar lch 152, áit a dtugtar aitheantas do nós seo an ‘roghnachais’ idir foirmeacha seanbhunaithe agus na foirmeacha simplithe (Seirbhís Thithe an Oireachtais 2017).

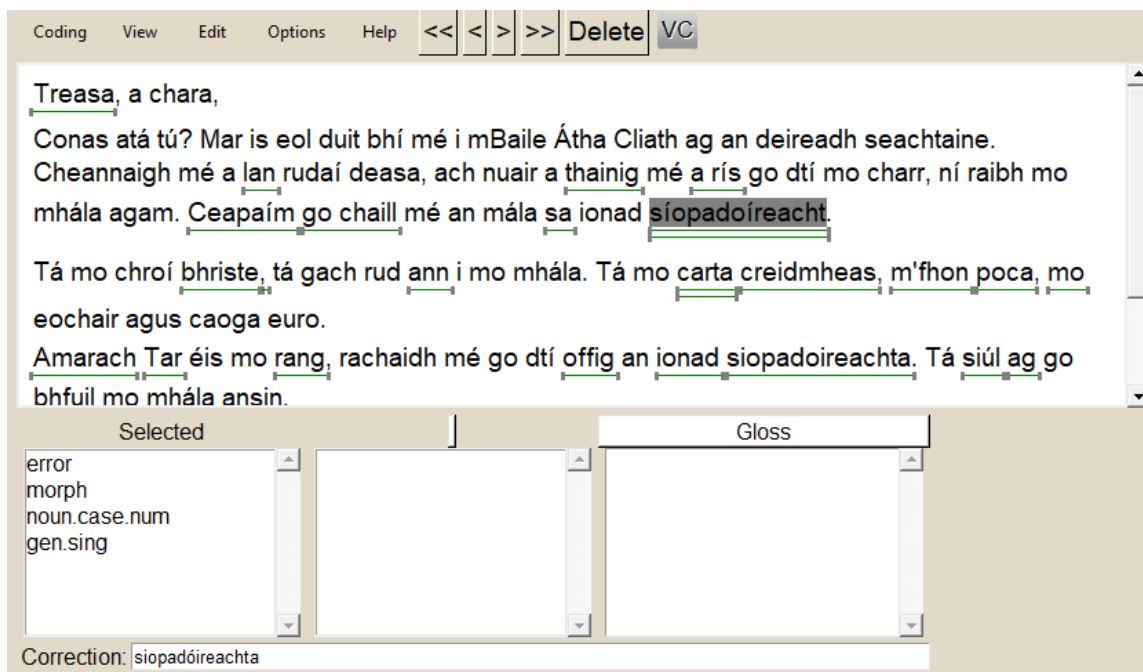
²⁰ Is corpas 30 milliún focal é NCI atá clibeáilte ó thaobh na rann cainte agus na leamaí de, rud a chuireann ar chumas an úsáideora cuardaigh chasta ar dhéanamh ar struchtúir chomhréire, seachas ar fhocail faoi leith. Tá sé ar fáil ag: <https://focloir.sketchengine.co.uk/run.cgi/index> [ceadaithe Eanáir 2020].

cuireadh clib earráide leis. Chuidigh beirt chúntóirí, a raibh ardchaighdeán Gaeilge acu, leis an obair chlibeála. Rinne beirt (duine de na cúntóirí agus an príomhthaipeoir) gach téacs a chlibeáil, agus rinneadh aon chásanna easaontais nó amhrais a fhiosrú, bunaithe ar fhianaise chorpais, mar atá mínithe thuas.

Is é an chéad chéim eile sa phróiseas clibeála ná na hearráidí a rangú, ionas go mbeifear in ann cuardach agus anailís a dhéanamh ar chineálacha éagsúla earráidí sa chorpas. Leis sin a dhéanamh, ní mór scéim clibeála a réiteach, a leagann amach liosta iomlán na gcatagóirí lena bhféadfaí na hearráidí a rangú (James 1998, Ellis & Barkhuizen 2005). Is céim an-tábhachtach í seo, toisc go mbíonn an léargas a fhaightear ón anailís ag brath go mór ar an mbealach a ndéantar na hearráidí a rangú. Molann Granger (2003) go mbeadh an scéim earráidí cuimsitheach agus mionsonraithe go leor le hidirdhealú fiúntach a dhéanamh idir cineálacha éagsúla earráidí, ach go mbeadh sí teoranta go leor le go mbeifí in ann í a láimhseáil agus a chur i bhfeidhm ar bhealach cruinn agus leanúnach. Tá sé tábhachtach a chinntiú, freisin, go bhfuil na catagóirí ar fad soiléir agus neamhspleách ar a chéile, agus nach bhfuil catagóirí sa scéim atá ag teacht salach ar a chéile. Tá scéim mhionsonraithe earráidí forbartha don tionscadal seo. Is córas ordlathach é, a bhfuil trí mhórchatagóir ann (an ortagrafaíocht, an mhoirfeolaíocht, agus léacsas agus comhréir) agus fochatagóirí taobh istigh de na mórchatagóirí sin. Faoi ortagrafaíocht, mar shampla, tá 6 fochatagóir a bhaineann leis an litriú, síntí fada, spásanna, fleiscíní agus ceannlitreacha. Tá 11 fochatagóir agus 36 mionchatagóir ann a bhaineann leis an moirfeolaíocht. Is earráidí iad sin a bhaineann le foirm ghramadúil an fhocail, mar shampla, na hathruithe tosaigh, foirmeacha infhillte ainmfhocal, aidiachtaí, briathra, agus araile. Ar deireadh, tá na hearráidí a bhaineann leis an léacsas (botúin ina n-úsáidtear an focal mícheart) agus an chomhréir (ord na bhfocal san abairt) curtha le chéile i mórchatagóir amháin, agus 18 bhfochatagóir luaite leo. Fágann sé seo go bhfuil 69 catagóir ar an iomlán i gceist. Tá breis is 13,000 earráid clibeáilte sa chorpas de réir na gcatagóirí sin ag an bpointe seo.

Úsáideadh an clár clibeála UAM Corpus Tool, leagan 3.3 (O'Donnell 2008) leis an obair chlibeála a éascú.²¹ Is uirlis é sin inar féidir scéimeanna clibeála a chruthú, bunaithe ar cibé teanga agus ábhar taighde atá i gceist, agus anótáil a dhéanamh ar théacsanna corpais de réir na scéimeanna sin. Tá roinnt feidhmeanna ag an gclár díreach ar mhaithe leis an gclibeáil earráidí; mar shampla, is féidir ceartúchán gach earráide a chódú mar chuid den obair chlibeála, is féidir roinnt earráidí a chlibeáil ar aon fhocal nó ar aon fhrása amháin sa téacs, agus is féidir an comhaontú idir beirt anótálaithe a ríomh. Tá cuid de na feidhmeanna seo le feiceáil in Íomhá 4, a léiríonn comhéadan clibeála an chláir.

²¹ Is féidir UAM Corpus Tool a íoslódáil ó <http://www.corpustool.com/index.html> [ceadaithe Eanáir 2020].



Íomhá 4: Comhéadan clibeála earráidí an chláir UAM Corpus Tool 3.3 (O'Donnell 2008)²²

Tá uirlisí bunúsacha anailíse ar fáil mar chuid de UAM freisin, a léiríonn líon agus liostaí na n-earráidí atá clibeáilte i ngach catagóir, rud atá fíorúsáideach le súil a choinneáil ar dhul chun cinn na hoibre agus le leanúnachas na clibeála a chinntiú. Déantar na comhaid anótála a stóráil neamhspleách ar na buntéacsanna (*stand-off annotation*), rud a fhágann gur féidir leasuithe beaga a dhéanamh sna téacsanna nó sa scéim clibeála, de réir mar is gá le linn na hoibre, gan cur isteach ar na buntéacsanna ná ar an obair atá déanta cheana féin. Is féidir na comhaid atá clibeáilte a easpórtáil i bhformáid XML, ionas go mbeifear in ann iad a léamh i ríomhchláir eile nó a aistriú go formáidí eile, más gá.

4.3.2 Clibeáil ranna cainte

Tugann an chlibeáil earráidí léargas ar na deacrachtaí is mó a bhíonn ag foghlaimoirí ag leibhéal éagsúla, ach le léargas níos iomláine a fháil ar chumas gramadaí na bhfoghlaimoirí, caithfear dul níos faide ná an anailís ar earráidí — caithfear anailís a dhéanamh ar na gnéithe den ghramadach a bhíonn i gceart acu, na struchtúir a n-éiríonn leo máistreacht nó smacht a fháil orthu. Leis seo a dhéanamh ar bhealach cuimsitheach, córasach, ní mór a bheith in ann cuardach agus scagadh a dhéanamh ar fhoirmeacha moirfeolaíochta agus ar struchtúir chomhréire sa chorpas ag leibhéal níos doimhne ná leibhéal an fhocail. Caithfear a bheith in ann cuardach a dhéanamh ar na sonraí ó thaobh na ranna cainte agus na leamaí de, agus teaghráin (*strings*) focal a aimsiú, chomh maith le focail aonair. Is féidir clibeáil a dhéanamh ar ranna cainte go huathoibríoch taobh istigh de go leor uirlisí corpais. San uirlis Sketch Engine, mar shampla, tá clibeáil uathoibríoch ar ranna cainte ar fáil do 36 teanga.²³ Sa chóras CHILDES, a úsáidtear le

²² San íomhá seo de chomhéadan clibeála UAM Corpus Tool, léiríonn líne faoi fhocal go bhfuil earráid ag baint leis an bhfocal sin. Tá dhá earráid (earráid litrithe agus earráid a bhaineann leis an nginideach uatha) ag baint leis an bhfocal atá aibhsithe. Léiríonn an fhuinneog atá ag bun na híomhá ceann amháin de na catagóirí earráide atá roghnaithe don fhocal áirithe sin (an ginideach uatha), agus an leagan ceartaithe den fhocal ('siopadóireachta').

²³ Féach <https://www.sketchengine.eu/corpora-and-languages/> [ceadaithe Eanáir 2020].

hanailís a dhéanamh ar theanga páistí agus foghlaimeoirí, tá clibeáil uathoibríoch ar ranna cainte ar fáil do 11 teanga.²⁴ Sna cásanna sin, ní theastaíonn saineolas teicniúil leis an gclibeáil a dhéanamh. Mar a deir MacWhinney (2008: 172): ‘*Running MOR on a CHAT file is easy. In the simplest case, it involves nothing much more than a one-line command.*’ Ar an drochuair, níl sé chomh héasca céanna i gcás na Gaeilge, toisc nach bhfuil sí i measc na dteangacha a bhfuil na huirlisí sin ar fáil go forleathan dóibh. Is míbhuntáiste é seo do lucht taighde na Gaeilge, mar a deir Ó Duibhir (2009: 47–8): ‘*Software such as CHILDES contains parts of speech tagging and error tagging for many world languages. Researchers of lesser-used languages do not enjoy these advantages however.*’ Dá bharr seo, bhraith na hanailísí atá déanta ar Ghaeilge foghlaimeoirí go dtí seo ar chuardaigh ar shonraí loma corpais (m.sh., Ó Duibhir 2009, 2018), ar struchtúir ghramadaí a aimsiú agus a chomhaireamh de láimh (m.sh., Péterváry *et al.* 2014) nó ar an anailís ar earráidí amháin (m.sh., Ó Domhnalláin & Ó Baoill 1978). Cé go raibh fiúntas mór ag baint leis an taighde sin ar fad, bhí teorainn ag baint le cumhacht na hanailíse, le méid na gcorpas agus leis an leas fadtéarmach a d’fhéadfaí a bhaint as na sonraí, toisc nach raibh na hacmhainní ar fáil chun clibeáil chuimsitheach, uathoibríoch a dhéanamh orthu.

Tá córas clibeála forbartha ag Uí Dhonnchadha (2009) don Ghaeilge, córas ar baineadh úsáid as le hanótáil a dhéanamh ar chorpas 30 milliún focal Gaeilge, *Nua-Chorpas na hÉireann* (Kilgarriff *et al.* 2006), ar an gcorpas labhartha *Comhrá*,²⁵ agus ar chorpais a bhaineann leis an nGaeilge stairiúil.²⁶ Is córas uathoibríoch, riail-bhunaithe atá ann, a bhfuil ráta cruinnis idir 94 agus 95% ag baint leis nuair a úsáidtear é ar an nGaeilge chaighdeánach. Bíonn dúshlán breise ag baint le teanga foghlaimeoirí a chlibeáil, ar ndóigh, toisc go gcuireann botúin litrithe agus ghramadaí isteach ar chóras riail-bhunaithe an chlibeálaí. Tá roinnt modhanna éagsúla aimsithe ag lucht na gcorpas foghlaimeora leis an dúshlán sin a shárú agus clibeálaithe caighdeánacha a chur in oiriúint do theanga foghlaimeoirí (de Haan 2000, Dickinson & Meurers 2003, van Rooy & Schäfer 2003, Díaz-Negrillo *et al.* 2010).

I gcás Chorpas Foghlaimeora TEG, rinneadh clibeálaí ranna cainte Uí Dhonnchadha a thástáil ar shampla 4000 focal de théacsanna scríofa an chorpais, a raibh beagnach 400 botún iontu. Léirigh anailís ar na sonraí sin go raibh ráta cruinnis de 84.5% ag baint leis na clibeanna; is é sin, ráta cruinnis 10% níos ísle ná an ráta cruinnis a bhíonn ag an gclibeálaí céanna nuair a bhíonn teanga chaighdeánach i gceist. Bhain dhá thrian de na clibeanna a bhí míchruinn le focail a raibh botúin litrithe iontu nó le focail a tháinig díreach roimh nó i ndiaidh botún litrithe nó gramadaí sa téacs, rud a thug le fios go raibh na hearráidí i dteanga na bhfoghlaimeoirí ag teacht salach ar chóras riail-bhunaithe an chlibeálaí agus ag cur isteach ar chruinneas na gclibeanna dá réir. Leis an bhfadhb seo a réiteach, leasaíodh an modh clibeála ranna cainte chun é a dhéanamh in dhá chéim; ar dtús, socraíodh an clibeálaí le neamhaird a dhéanamh de na hearráidí ar fad a bhí sna téacsanna, agus clibeáil a dhéanamh ar na gnéithe de na téacsanna a bhí cruinn, na ceartúcháin san áireamh; ina dhiaidh sin, rinneadh clibeáil ar na focail a raibh earráidí ag baint leo. Sa chás nach raibh an clibeálaí in ann an focal a aithint (i gcás botún litrithe, mar shampla), cuireadh clib faoi leith leis, chun é sin a chur in iúl.

²⁴ Féach <https://talkbank.org/morgrams/> [ceadaithe Eanáir 2020].

²⁵ Féach <https://www.scss.tcd.ie/~uidhonne/comhra/index.utf8.html> [ceadaithe Eanáir 2020].

²⁶ Féach Mac Cárthaigh & Uí Dhonnchadha san imleabhar seo.

4.3.3 Uirlis anailíse

Roghnaíodh an clár Sketch Engine mar uirlis óstála agus anailíse don chorpas seo.²⁷ Is uirlis chorpais é seo inar féidir corpas a stóráil ar líne agus anailís a dhéanamh air. Is féidir cuardaigh shimplí a dhéanamh ar fhoirmeacha nó ar earráidí faoi leith, nó cuardaigh chasta a dhéanamh ar theaghráin leamaí agus ranna cainte, le hanailís a dhéanamh ar an gcomhréir (Kilgarriff *et al.* 2014). Is féidir na torthaí ar fad a scagadh de réir meiteashonraí, ar nós leibhéal cumais, scil, tasc, próifíl an iarrthóra, agus araile. Léiríonn Íomhá 5 an leagan amach atá ar shonraí an chorpais féin in Sketch Engine. Is cuardach ar earráidí a bhaineann le húsáid réamhfhocal atá le feiceáil san íomhá agus tá na torthaí sórtáilte de réir leibhéil an FTCE. Tá an earráid agus an ceartúchán le feiceáil i lár an scáileáin agus clibeanna ranna cainte curtha leo. Tá an leibhéal cumais a bhain leis an té a scríobh gach abairt le feiceáil ag taobh an scáileáin.²⁸ Buntáiste mór a bhaineann le Sketch Engine ná gur ardán ar líne atá ann, rud a fhágann go mbeifear in ann rochtain a thabhairt do thaighdeoirí eile úsáid a bhaint as an gcorpas ar bhealach sábháilte. Tá tábhacht ag baint leis an bpointe áirithe sin, toisc go gcuirfidh sé le trédhearcacht an taighde féin. Beidh taighdeoirí eile in ann breathnú, ní hamháin ar thorthaí na hanailíse, ach ar na sonraí atá taobh thiar díobh, agus an obair ar fad a mheas ar bhealach níos criticiúla dá bharr.

B1	me </err><corr> mé </corr><err>	agat /Pr2-s </err><corr> leat /Pr2-s </corr>	sa teach <err> tait
B1	</corr><corr> Téann </corr><err> ar	/Wssi Wpdi Wpsi </err><corr> chuig /Sp </corr>	rang teanga gach
B1	r><err/><corr> triúr </corr><err>	isteach /Rd </err><corr> i /Sp </corr>	mo rang . Tá siad
B1	r> caite </corr> chuaigh mé <err>	go dtí /Spc </err><corr> go /Cs </corr><err>	sasana </err>
B2	<corr> scamhógach </corr><err>	agat /Pr2-s </err><corr> ort /Pr2-s </corr>	. Ní <err> shé </err>
B2	don chuid eile <err>	do /Sp </err><corr> de /Pr3ms Sp </corr>	do shaol ? Is é do
B2	smacht agat <err>	faoi /Pr3ms </err><corr> air /Nmm-2s </corr>	sa <err> domhain
B2	<corr> thuismitheoirí </corr><err>	a /Ud </err><corr> i /Sp </corr>	láthair na
B2	<corr> rachaidh </corr> mé <err>	go dtí /Spc </err><corr> go /Sp </corr>	Chicago ar feadh
B2	anois , mar is eol <err>	agat /Pr2-s </err><corr> duit /Pr2-s </corr>	, agus <err> tá </err>
B2	</err><corr> mhaith </corr><err>	agam /Pr1-s </err><corr> dom /Pr1-s </corr>	a bheith ag obair
B2	a lán grá agam <err>	sa /Spa-s </err><corr> don /Spa-s </corr>	teanga seo mar <err>
B2	tá grá nua agam <err>	sa /Spa-s </err><corr> don /Spa-s </corr>	teanga seo cé go
C1	</err><corr> fhágáil </corr><err>	leatsa /Pr2-s-e </err><corr> fútsa /Pr2-s-e </corr>	. Táim idir dhá

Íomhá 5: Cuardach Sketch Engine ar earráidí a bhaineann le húsáid réamhfhocal sa chorpas

5 Conclúid

D'eascair an tionscadal seo as ceisteanna maidir le leibhéil chumais an FTCE, agus na sainchomharthaí a bhaineann leo i gcás na Gaeilge. Is é an chéad aidhm atá leis an taighde ná anailís a dhéanamh ar an gcumas gramadaí ag leibhéil B1–C1 den FTCE, toisc gurb iad sin na leibhéil is mó atá in úsáid faoi láthair le cumas teanga múinteoirí, aistritheoirí agus fostaithe eile a dhearbhuí. Mar gheall ar na himpleachtaí móra a bhaineann leis an measúnú agus leis an teastasú ag na leibhéil sin anois, tá freagracht ar eagraíochtaí agus institiúidí a bheith cinnte go bhfuil bailíocht, leanúnachas agus trédhearcacht ag baint leis an mbealach a n-úsáideann siad

²⁷ Féach <https://www.sketchengine.eu/>.

²⁸ Tá mé an-bhuíoch den Dr Elaine Uí Dhonnchadha as an obair mhór atá déanta aici leis an gclibeálaí ranna cainte a chur in oiriúint d'ábhar an chorpais seo, agus leis an aschur ón gclibeálaí earráidí UAM Corpus Tool a aistriú go formáid atá comhoiriúnach do Sketch Engine.

lipéid an FTCE. Tá súil agam go mbeidh an obair seo ina cuidiú dóibh siúd atá ag iarraidh é sin a dhéanamh. Níl sa tionscadal áirithe seo, áfach, ach tús na hoibre. Buntáiste mór a bhaineann le corpas atá dea-dheartha agus anótáilte ar bhealach córasach, ná go mbíonn luach fadtéarmach ag baint leis. Nuair a bheidh an tionscadal seo curtha i gcrích, beifear in ann an corpas a úsáid le hanailís a dhéanamh ar ghnéithe eile de theanga na bhfoghlaimoirí, nó le hábhar teagaisc agus acmhainní foghlama a dhearadh atá dírithe ar shainriachtanais foghlaimoirí ag leibhéal éagsúla chumais, ar bhealach taighde-bhunaithe.

Má thosaigh an obair seo le ceisteanna, is iomaí ceist eile atá tar éis teacht chun cinn ó cuireadh tús léi — céard is ‘cainteoir dúchais’ ann agus céard is ‘foghlaimoir’ ann i gcás na Gaeilge? Céard í earráid agus céard atá i gceist le cruinneas Gaeilge, go háirithe i gcás na teanga labhartha? Céard é an cumas gramadaí, nó, go deimhin, céard í an ghramadach féin, agus cá bhfuil an teorainn idir an ghramadach, an fhóineolaíocht agus an léacsas i gcás na Gaeilge? Agus mé ag iarraidh dul i ngleic leis na ceisteanna seo ar fad, tá fírinne an ráitis seo ó Hickey (1992: 3) feicthe agam ag chuile chéim den phróiseas:

Setting out to study the acquisition of Irish was somewhat akin to deciding to build a house, being architect, surveyor and builder, and then discovering that the bricks also had to be made. The most basic concepts in language acquisition had to be re-thought in the context of studying the acquisition of Irish.

Go leor de na coincheapa a ndéantar talamh slán díobh i gcás teangacha agus comhthéacsanna eile, caithfear athmhachnamh a dhéanamh orthu i gcás na Gaeilge. Go leor de na huirlisí atá ar fáil do theangacha eile chun an obair chorpais a éascú, caithfear iad a fhorbairt ó bhonn nó a chur in oiriúint don Ghaeilge. Teastaíonn cur chuige criticiúil agus solúbtha mar sin, agus teastaíonn comhoibriú idir teangeolaithe feidhmeacha agus ríomhtheangeolaithe chun na huirlisí seo a chur ar fáil. Tá súil agam go gcuirfidh an tionscadal seo bunchloch mhodheolaíochta, nó roinnt ‘brící’, ar fáil, a mbeifear in ann tógáil orthu amach anseo, le corpais eile foghlaimora a fhorbairt don Ghaeilge.

TAGAIRTÍ

- An Roinn Oideachais agus Scileanna (2017a) *Sonraíocht Ghaeilge na Sraithe Sóisearaí: Scoileanna a fheidhmíonn trí mheán an Bhéarla (T2)*. Ar fáil ag: https://www.curriculumonline.ie/getmedia/ac4ea67a-3318-4900-a156-9e8cb9cafaac/JC-GAE-T2_sonraiocht-deiridh.pdf [ceadaithe Eanáir 2020].
- An Roinn Oideachais agus Scileanna (2017b) *Sonraíocht Ghaeilge na Sraithe Sóisearaí: Scoileanna a fheidhmíonn trí mheán na Gaeilge (T1)*. Ar fáil ag https://www.curriculumonline.ie/getmedia/394a74f1-ba27-4572-9f1b-93696859fc03/JC-GAE-T1_sonraiocht-deiridh.pdf [ceadaithe Eanáir 2020].
- Austin, J.L. (1962) *How to Do Things with Words*. Oxford: Clarendon Press.
- Bachman, L.F. & Palmer, A.S. (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Boyd, A., Hana, J., Nicolas, L., Meurers, D., Wisniewski, K., Abel, A., Schöne, K., Štindlová, B. & Vettori, C. (2014) ‘The MERLIN corpus: Learner language and the CEFR’. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno,

- A., Odijk, J., Piperidis, S. (eds) *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC '14)*, European Language Resources Association (ELRA): Reykjavik, Iceland: 1281–1288.
- Callies, M. & Götz, S. (eds) (2015) *Learner Corpora in Language Testing and Assessment: Prospects and Challenges*. Amsterdam: John Benjamins Publishing Company.
- Cambridge University Press (2015) *English Grammar Profile*. Ar fáil ag: <https://www.englishprofile.org/english-grammar-profile/egp-online> [ceadaithe Eanáir 2020].
- Canale, M. & Swain, M. (1980) 'Theoretical bases of communicative approaches to second language teaching and testing'. In: *Applied Linguistics* 1: 1–47.
- Carlsen, C. (eag.) (2013) *Norsk Profil. Det europeiske rammeverket spesifisert for norsk*. Et første steg. Novus, Oslo.
- Council of Europe (2001) *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2005) *Reference Level Descriptions for National and Regional Languages: Guide for Production of RLDs*. Ar fáil ag: <https://rm.coe.int/090000168077c574> [ceadaithe Eanáir 2020].
- Council of Europe (2018) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Ar fáil ag: <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989> [ceadaithe Eanáir 2020].
- Davies, A. (2003) *The Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters.
- de Haan, P. (2000) 'Tagging non-native English with the TOSCA-ICLE tagger'. In: C. Mair & M. Hundt (eds) *Corpus Linguistics and Linguistic Theory*. Amsterdam: Rodopi: 69–79.
- Deygers, B., Carlsen, C.H., Saville, N. & Van Gorp, K. (2017) 'The use of the CEFR in higher education: a brief introduction to this special issue'. In: *Language Assessment Quarterly* 15 (1): 1–2.
- Díaz-Negrillo, A., Meurers, D., Valera, S. & Wunseh, H. (2010) 'Towards interlanguage POS annotation for effective learner corpora in SLA and FLT'. In: *Language Forum* 36 (1–2): 139–154.
- Dickinson, M. & Meurers, W.D. (2003) 'Detecting errors in part-of-speech annotation'. In: *Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics EACL* 3 (1): 107–114.
- Ellis, R. & Barkhuizen, G. (2005) *Analysing Learner Language*. Oxford: Oxford University Press.
- Figueras, N. (2012) 'The impact of the CEFR'. In: *ELT Journal*, 66 (4): 477–485.
- Glaboniat, M., Müller, M., Schmitz, H., Rusch, P. & Wertenschlag, L. (2005). *Profil Deutsch*. Berlin: Langenscheidt.
- Granger, S. (2003) 'Error-tagged learner corpora and CALL: A promising synergy'. In: *CALICO Journal* 20 (3): 465–480.
- Granger, S., Gilquin, G. & Meunier, F. (2015) 'Introduction: learner corpus research – past, present and future'. In: Granger, S., Gilquin, G. & Meunier, F. (eds) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press: 1–6.
- Green, A. (2010) 'Requirements for Reference Level Descriptions for English'. In: *English Profile Journal* 1: 1–19.

- Halliday, M.A.K. (1994) *An Introduction to Functional Grammar (2nd ed.)*. London: Edward Arnold.
- Harrison, J. & Barker, F. (eds) (2015) *English Profile in Practice*. English Profile Studies Vol. 5. Cambridge: Cambridge University Press.
- Hawkins, J.A. & Buttery, P. (2010) 'Critical features in learner corpora: Theory and illustrations'. In: *English Profile Journal* 1 (1): 1–23.
- Hawkins, J.A. & Filipović, L. (2012) *Critical Features in L2 English. Specifying the Reference Levels of the Common European Framework*. English Profile Studies Vol. 1. Cambridge: Cambridge University Press.
- Hickey, T. (1992) 'The acquisition of Irish as a first language: What do we know?' In: Ó Baoill, D. (eag.) *Insealbhú na Gaeilge mar Chéad Teanga*. Baile Átha Cliath: Irish Association for Applied Linguistics: 1–25.
- Hymes, D.H. (1972) 'On communicative competence'. In: Pride, J.B. & Holmes, J. (eds) *Sociolinguistics*. Baltimore, USA: Penguin Education, Penguin Books Ltd: 269–293.
- James, C. (1998) *Errors in Language Learning and Use: Exploring Error Analysis*. London: Longman.
- Kilgarriff, A., Rundell, M. & Uí Dhonnchadha, E. (2006) 'Efficient corpus development for lexicography: Building the New Corpus for Ireland'. In: *Language Resources and Evaluation* 40 (2): 127–152.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014) 'The Sketch Engine: Ten years on'. In: *Lexicography* 1 (1): 7–36.
- Leacock, C., Chodorow, M. & Tetreault, J. (2015) 'Automatic grammar and spell-checking for language learners'. In: Granger, S., Gilquin, G. & Meunier, F. (eds) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press: 567–586.
- Lenoach, C. (2014) 'Sealbhú neamhiomlán na Gaeilge mar chéad teanga sa dátheangachas dealaitheach'. Tráchtas dochtúireachta. Ollscoil na hÉireann Gaillimh. Ar fáil ag: <https://www.cogg.ie/wp-content/uploads/phd-ciaran-lenoach.pdf> [ceadaithe Samhain 2019].
- Lennon, P. (1991) 'Error: Some problems of definition, identification, and distinction'. In: *Applied Linguistics* 12 (2): 180–196.
- Little, D. (2011) 'The CEFR, the European Language Portfolio, and language learning in higher education'. In: *Language Learning in Higher Education* 1 (1): 1–21.
- Mac Mathúna, Liam. (2008) 'Linguistic change and standardization'. In: Nic Phóidín, C. & Ó Cearnaigh, S. (eag.) *A New View of the Irish Language*. Baile Átha Cliath: Cois Life: 76–92.
- MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2008) 'Enriching CHILDES for morphosyntactic analysis'. In: Behrens, H. (ed.) *Corpora in Language Acquisition Research*. John Benjamins: Amsterdam: 165–198.
- McCarthy, M.J. (2016) 'Putting the CEFR to good use: designing grammars based on learner-corpus evidence'. In: *Language Teaching* 49 (1): 99–115.
- McEnery, T., Xiao, R. & Tono, Y. (2006) *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- Montrul, S. (2013) 'Incomplete L1 acquisition'. In: Herschensohn, J.R. & Young-Scholten, M. (eds) *The Cambridge Handbook of Second Language Acquisition*. Cambridge: Cambridge University Press: 353–371.
- Murakami, A. & Alexopoulou, T. (2016) 'L1 Influence on the Acquisition Order of English

- Grammatical Morphemes'. In: *Studies in Second Language Acquisition* 38 (3): 365–401.
- Na Bráithre Críostaí (1999) *Graiméar Gaeilge na mBráithre Críostaí*. Baile Átha Cliath: An Gúm. Ar fáil ag:
http://www.scriobh.ie/ScriobhIe/Media/Graimear%20Gaeilge%20na%20mBraithre%20Críostaí_Eag1999.pdf [ceadaithe Eanáir 2020].
- North, B. (2014) *The CEFR in Practice*. English Profile Studies Vol. 4. Cambridge: Cambridge University Press.
- Ní Ghearáin, H. (2019) 'Caighdeán, caighdeánú agus cumhacht'. In: Ó hIfearnáin, T. (eag.). *An tSochtheangeolaíocht: Taighde agus Gníomh*. Baile Átha Cliath: Cois Life: 155–184.
- Ó Baoill, D. (2000) *Athchaighdeánú na Nua-Ghaeilge*. Baile Átha Cliath: Coiscéim.
- Ó Curnáin, B. (2009) 'Mionteangú na Gaeilge'. In: B. Ó Catháin (eag.) *Sochtheangeolaíocht na Gaeilge, Léachtaí Cholm Cille 39*. Maigh Nuad: An Sagart: 90–153.
- Ó Domhnalláin, T. & Ó Baoill, D. (1978) *Earráidí Scríofa Gaeilge Cuid 1: Earráidí Briathra*. Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.
- Ó Dónaill, N. (1977) *Foclóir Gaeilge–Béarla*. Baile Átha Cliath: An Gúm.
- O'Donnell, M. (2008) 'The UAM CorpusTool: Software for corpus annotation and exploration'. In: Bretones Callejas, C.M. et al. (eds) *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*. Almería: Universidad de Almería: 1433–1447. Ar fáil ag:
<http://www.wagsoft.com/Papers/AESLA08.pdf> [ceadaithe Eanáir 2020].
- Ó Duibhir, P. (2009) 'The spoken Irish of sixth-class pupils in Irish immersion schools'. Unpublished Doctoral Dissertation. Trinity College Dublin.
- Ó Duibhir, P. (2018) *Immersion Education: Lessons from a Minority Language Context*. Bristol: Multilingual Matters.
- O'Keeffe, A. & Mark, G. (2017) 'The English Grammar Profile of learner competence'. In: *International Journal of Corpus Linguistics* 22 (4): 457–489.
- O'Rourke, B. & Ramallo, F. F. (2011). 'The native–non-native dichotomy in minority language contexts: Comparisons between Irish and Galician'. In: *Language Problems and Language Planning* 35 (2): 139–159.
- Paikeday, T. (1985) *The Native Speaker is Dead!: An Informal Discussion of a Linguistic Myth with Noam Chomsky and Other Linguists, Philosophers, Psychologists, and Lexicographers*. Toronto / New York: Paikeday Publishing.
- Péterváry, T., Ó Curnáin, B., Ó Giollagáin, C. & Sheahan, J. (2014) *Iníúchadh ar an gCumas Dáttheangach: An Sealbhú Teanga i Measc Ghlúin Óg na Gaeltachta*. Baile Átha Cliath: An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta.
- Rannóg an Aistriúcháin (1958) *Gramadach na Gaeilge: An Caighdeán Oifigiúil*. Baile Átha Cliath: Oifig an tSoláthair.
- Saville, N. & Hawkey, R. (2010) 'The English Profile Programme – the first three years'. In: *English Profile Journal* 1 (1): 1–14.
- Saville, N. & Milanović, M. (2012). 'Series editors' note'. In: Hawkins, J.A. & Filipović, L. (eds) *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. English Profile Studies Vol. 1. Cambridge: Cambridge University Press: x–xv.
- Searle, J.R. (1969) *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Seirbhís Thithe an Oireachtais (2017) *Gramadach na Gaeilge: An Caighdeán Oifigiúil: An*

- Treoir le haghaidh Scríbhneoireacht na Gaeilge*. Baile Átha Cliath: Seirbhís Thithe an Oireachtais.
- Skehan, P. (1996) 'A framework for the implementation of task-based instruction'. In: *Applied Linguistics* 17 (1): 38–62.
- Spinelli, B. & Parizzi, F. (2010) *Profilo della lingua italiana*. Firenze: La Nuova Italia.
- Uí Dhonnchadha, E. (2009) 'Part-of-speech tagging and partial parsing for Irish using finite-state transducers and constraint grammar'. Tráchtas dochtúireachta. Dublin City University. Ar fáil ag: http://doras.dcu.ie/2349/1/PhD_Elaine_Final.pdf [ceadaithe Eanáir 2020].
- van Rooy, B. (2015) 'Annotating learner corpora'. In: Granger, S., Gilquin, G. & Meunier, F. (eds) *The Cambridge handbook of learner corpus research*: 79–106.
- van Rooy, B. & Schäfer, L. (2003) 'An evaluation of three POS taggers for the tagging of the Tswana Learner English Corpus'. In: Archer, D., Rayson, P., Wilson, A. & McEnery, T. (eds) *Proceedings of the Corpus Linguistics 2003 Conference*, Lancaster University, 28–31 March 2003, UCREL Technical Paper 16. Lancaster University: 835–44.
- Walsh, J. & Nic Eoin, M. (2010) 'Siollabas nua Gaeilge don chéad bhliain ollscoile'. In: *Teagasc na Gaeilge* 9: 13–24.
- Wilkins, D. (1976) *Notional Syllabuses*. Oxford: Oxford University Press.
- Willis, J. (1996) *A Framework for Task-Based Learning*. Harlow: Longman.
- Wynne, M. (ed.) (2005) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books.