

Gramadóirí Gaeilge: an chéad ghlúin eile

Kevin Scannell
Cadhan Aonair
14 Samhain 2025

An Gramadóir

- Fiche (is a dó) bliain ag fás
- Clibeálaí + rialacha
- Go leor rialacha ann 
- Eisceachtaí ar eisceachtaí 
- Ceangailte le rialacha sa CO 
- Ní thuigeann sé comhréir 
- An-deacair rialacha a scríobh/athrú 
- Bunaithe ar rialacha => sobhriste 

Téacs le seiceáil:

Chonaic mé na fir móra.
Chonaic mé na caoirigh mhóra.
Bhí sé gan fheoirling.
Tharla sé gan fios dom.

Seol

Glan

1: Chonaic mé na **fir móra**.

Séimhiú ar iarraigdh

2: Chonaic mé na **caoirigh mhóra**.

Séimhiú gan ghá

3: Bhí sé **gan fheoirling**.

Séimhiú gan ghá

4: Tharla sé **gan fios** dom.

Séimhiú ar iarraigdh



Fintan

@fintan-galway.bsky.social

Gramadóir ag iarraidh fadhbanna a chruthú dom...

@kscanne.bsky.social 😊

1: Sruth Gaeilge den chéim sa Bhith-theicneolaíocht ar fáilanois ó Scoil na nEolaíochtaí Bitheacha agus Ceimiceacha, buíochas le teagasc ó Acadamh na hOllscolaíochta Gaeilge.
Unknown word: /bitseachá/?

July 29, 2025 at 10:32 AM ☺ Tig le chuile dhuine freagra a thabhairt · Aistrigh

1 athphostáil 6 mholadh

2020: Meaisínfhoghlaim amháin



Caoimhín Ó Scanail

@kscanne

...

Samhlaigh nach bhfuil Gaeilge ar bith agat. Míním duit go bhfuil rud darb ainm "séimhiú" ann, ina gcuirtear h isteachanois is arís. Tugaim 1000 leabhar Gaeilge duit (ní thuigeann tú iad - siombailí ar leathanaigh atá iontu!) An mbeifeá in ann na rialacha go léir a oibriú amach?

Bheifeá!

- Bhuel, bhí an ríomhaire in ann é a dhéanamh
- Thraenáil mé líonra néarach chun athruithe tosaigh a thuar
- Cruinneas an-ard (98.33%) 
- Ginearalú éifeachtach (focail nach raibh sa traenáil, gnéithe canúna) 
- Ní gá rialacha a scríobh (dhá lá á fhorbairt in áit dhá bhliain) 
- Ach, níl aon rud ceangailte le rialacha — gan míniú ar bith 
- D'fhoghlaim sé ó théacsanna ar líne nach gcloíonn leis an CO i gcónaí  ( ?)
- *Neural Models for Predicting Celtic Mutations*

An Gnáthscríbhneoir Féinmhuijníneach (GSFM)

- Ba mhaith liom an dá thrá a fhreastal
- Leagan amháin a chloíonn leis an gCaighdeán Oifigiúil
- Leagan amháin a ghlacann le foirmeacha a mbeadh GSFM sásta leo
- Téann an Gramadóir thar fóir ag iarraidh cloí leis an gCaighdeán Oifigiúil
- Ní ghlacann sé le go leor foirmeacha atá i mbéal an phobail:
- **muna bhfuil, na ceantracha, na comórtaisí, a thógaint, deireann** sé, srl.
- Chruthaigh mé corpas earráidí chun an GSFM a chur i ngníomh!
- 5k tweet, 100k focal, 3k botún, ceartaithe de réir nósanna an GSFM (i.m.t.)
- <https://github.com/kscanne/qbb/tree/main/datasets/errors>

Ní leor ChatGPT ná baol air!



2

The rule (from *An Caighdeán Oifigiúil*)



Section: *An Aidiacht* (Adjectives), §11.3 (2017 edition)

"Déantar séimhiú ar an aidiacht i ndiaidh ainmfhocail baininscneach iolra nach gcríochnaíonn ar ghuta."

("An adjective is lenited after a plural feminine noun that does not end in a vowel.")



3

Compare with other patterns

Noun	Gender	Plural	Adjective form	Example
<i>fir</i> (men)	masc.	plural	no lenition	<i>fir mhóra</i> ❌ → <i>fir móra</i> ✓

KPS:

If you were to lenite the word "phingin" (meaning "penny") in some other content (say, after the possessive "mo"), how would that be spelled?

ChatGPT:

In Irish, when a noun begins with the consonant "p," it is lenited to "ph" after the possessive "mo" (which means "my" in Irish). So in the case of the noun "phingin" (which means "penny" in Irish), it would be lenited to "m'phhingin" after "mo," and the correct way to say "my penny" in Irish would be "m'phhingin."

KPS:

I see. If I want to say "without my penny", then that would be "gan mh'phhingin", yes?

ChatGPT:

Yes, that is correct. If you want to say "without my penny" in Irish, the correct way to say it would be "gan mh'phhingin." In Irish, the possessive "mo" (which means "my") undergoes lenition to "mh" when it is followed by a noun beginning with a vowel. Since the noun "phhingin" (which means "penny" in Irish) begins with the consonant "p," it undergoes lenition to "ph" after "mo."

Riachtanais

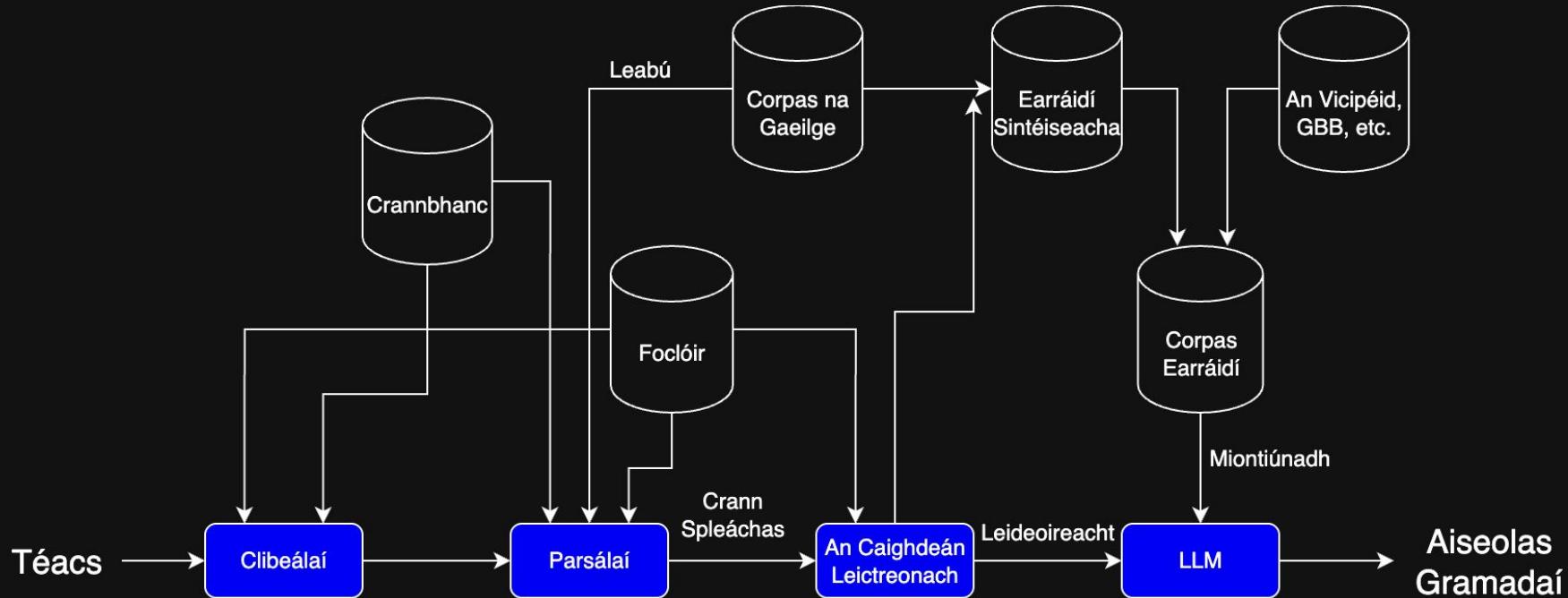
2003 2020 GPT 2025

Clúdach: déanann sé tuairisc ar chéatadán ard de na botúin atá ann	X	✓	X	
Cruinneas: tá céatadán ard de na tuairisci seo ceart	✓	✓	X	
Comhréir: aimsíonn sé botúin a bhaineann le struchtúr athchúrsach	X	X	X	
Mínithe: tugann sé míniú ar gach earráid agus tagairt don CO más gá	✓	X	X	
Iontaofacht: ní chumann sé rialacha bréagacha	✓	✓	X	
Líofacht: labhraíonn sé Gaeilge bhinn bhlasta é féin	✓	X	X	
Solúbthacht: ní bhriseann sé de bharr gnéithe nach rabhthas ag súil leo	X	✓	X	
Réchúiseacht: tacáíonn sé le nósanna atá i mbéal an phobail (GSFM)	X	✓	X	
Dífhabhtú: is féidir modh oibre an chórais a thuiscant agus a fheabhsú	✓	X	X	

Riachtanais

2003 2020 GPT 2025

Clúdach: déanann sé tuairisc ar chéatadán ard de na botúin atá ann				
Cruinneas: tá céatadán ard de na tuairisci seo ceart				
Comhréir: aimsíonn sé botúin a bhaineann le struchtúr athchúrsach				
Mínithe: tugann sé míniú ar gach earráid agus tagairt don CO más gá				
Iontaofacht: ní chumann sé rialacha bréagacha				
Líofacht: labhraíonn sé Gaeilge bhinn bhlasta é féin				
Solúbthacht: ní bhriseann sé de bharr gnéithe nach rabhthas ag súil leo				
Réchúiseacht: tacáíonn sé le nósanna atá i mbéal an phobail (GSFM)				
Dífhabhtú: is féidir modh oibre an chórais a thuiscant agus a fheabhsú				



DEMO

Foinse Eolais: Crannbhanc

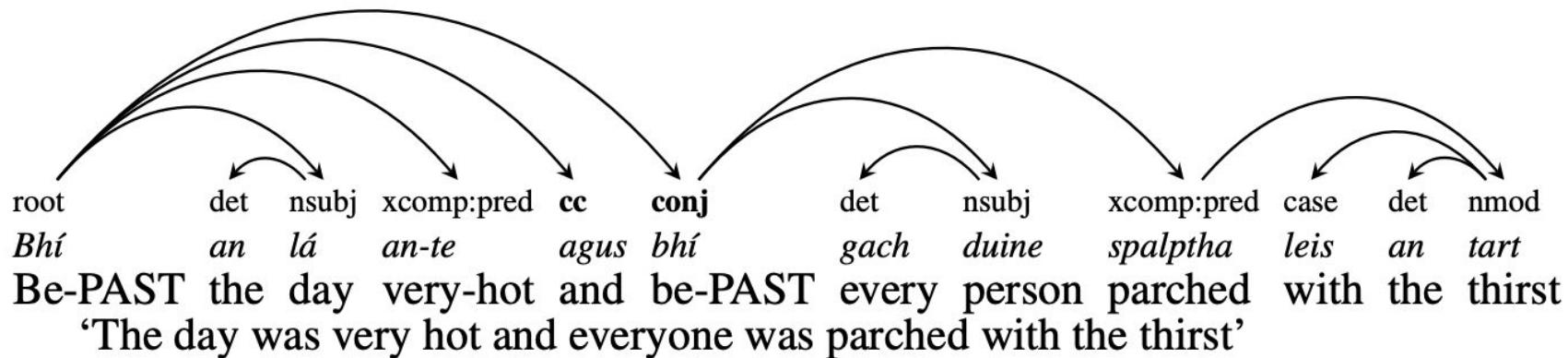


Figure 4 in Lynn, Teresa and Foster, Jennifer (2016) *Universal dependencies for Irish*. In: Second Celtic Language Technology Workshop. (CLTW 2016), 4 July 2016, Paris, France.

Foinse Eolais: An Caighdeán Leictreonach

```
if hd['lemma'] in gadata.feminineGroups:  
    return [Constraint('!Len', '10.2.7.j: Do not lenite a genitive noun after various feminine noun for groups or organizations')]  
  
if self['lemma'] in ['dli', 'si']:  
    return [Constraint('!Len', '10.2.7.k: Do not lenite "dli" or "si" after a feminine noun')]  
  
if self['lemma'] in ['bliain', 'coicis', 'mi', 'seachtain']:  
    return [Constraint('!Len', '10.2.7.l: Do not lenite various units of time after a feminine noun')]  
  
if self.isQualifiedNoun():  
    if hd['lemma'] in ['beirt', 'dis']:  
        return [Constraint('Len', '10.2.7.m.e1: Lenite nouns after "beirt" or "dis" even if the genitive noun is modified by an adjective or another noun')]  
        return [Constraint('!Len', '10.2.7.m: Do not lenite after a feminine noun if the genitive noun is modified by an adjective or another noun')]
```

- (j) Ní shéimhítéar an dara hainmfocal i ndiaidh na bhfocal *aicme, comhairle, corporáid, cuideachta, earnáil, feidhmeannacht, foireann, gníomhaireacht, institiúid, oifig, rannóg, roinn, scéim, seirbhís*, e.g., *aicme ceannais; comhairle contae; corporáid baincéireachta; earnáil gnó; foireann bainistiochta; institiúid breisoideachais; oifig preasa; rannóg pearsanna; roinn cosanta.*

Ní foláir dóibh an **scéim marcála** a choigeartú.
Cuireann an gnólacht **seirbhís comhairleoireachta** ar fáil.

- (k) Ní shéimhítéar na hainmfhocail *dli* ná *si*, i ndiaidh ainmfocal baininsneach, e.g., *feidhm dli; bó si*.
Cuirfidh an rialtas cóip den **ionstraim dli** ar fáil.
Deirtear gur comhartha an mhí-áidh an **long sí** a fhéiceáil.

- (l) Ní shéimhítéar an dara hainmfocal más ceann de na haonad tomhais bheactha *bliain, coicis, mi nó seachtain* an dara hainmfocal, e.g., *saorí miosá; tréimhse bliana; moill bliana*:

Cuireadh tú leis an **bhféile seachtaíne** le hócáid i dteach an Mhára.
Bhí **scíth coicise** de dhíth uirthi tar éis na hoibre.

- (m) Ní shéimhítéar an dara hainmfocal má tá an dara hainmfocal, atá sa ghinideach, cáilithe le hainmfocal eile ná le haidiacht, e.g., *oíche gaoithe móire; bileog páipéir dúigh; gloine beorach fuaraithe*:

Ní raibh **deoir bainne úir** fágtha sa teach.
Cuireadh an **obair tionscadail aistriúcháin** i gcrích inné.

- Eisceacht: nuair is é *beirt* ná *dís* an chéad ainmfocal, e.g., *beirt bhan mhisiúla* ná nuair is é *beirt* an dara hainmfocal, e.g., *páistí bheit bhan*.

Foinse Eolais: Earráidí sintéiseacha

- Cuir an parsálaí i bhfeidhm ar chorpas mó�
- Déan neamhaird d'fhocail a mbaineann fadhb ghamadaí leo cheana
- (Meascán: botúin a rinne an parsálaí agus botúin ghamadaí sa chorpas)
- Ceanglaíonn an Caighdeán Leictreonach (CL) “srianta” le gach focal
- Sáraigh na srianta seo d'aon turas:
 - Abairt cheart: *Tá an bhean ag canadh*
 - Cuireann an CL srian i bhfeidhm ar *bhean* — “séimhiú ag teastáil + CO 10.2.1”
 - Cúpla bealach chun an srian seo a shárú:
Tá an bean/S+10.2.1 ag canadh
Tá an mbean/S+10.2.1 ag canadh

Foinse Eolais: Ceartúcháin ar Vicipéid na Gaeilge

- Ar an bhfód ó Dheireadh Fómhair 2003; tá **57,713** alt ann faoi láthair
- Níos mó ná **1.1 milliún** athrú curtha i bhfeidhm ó 2003
- Scríobh mé ríomhchlár a dhéanann anailís ar na hathruithe seo
- Bhain mé amach na “ceartúcháin” go léir agus a gcomhthéacsanna:
 - “Is é bunachar sonraí téarmaíochta an Aontais [Eorpach→Eorpaigh] é.”
 - “...ceann de na toghcháin ba chonspóidí riamh [sa Mheiriceá→i Meiriceá]”
 - “Go luath [i ndiaidh→ina dhiaidh] sin, bunaíodh...”
 - “...an dóigh ar [thá inig→tháinig] an Chríostaíocht go hÉirinn”
 - “...tá oifigí nua-aimseartha [an chomhairle→na comhairle] suite ar phríomhshráid an bhaile”
 - “...go dtí gur [mhuscail→mhúscaill] foireann an Enterprise é de thaisme”
- Sonraí traenála nua: 115,000 ceartúchán as 11.4 milliún focal

Fós le déanamh

- Feabhas a chur ar an LLM Gaeilge
- Leagan GSFM a chur i gcrích
- Chuile rud a chur ar líne

Go raibh maith agaibh!

- <https://cadhan.com/>
- <https://github.com/kscanne>