

# Meaisínfhoghlaim, Gramadach na Gaeilge, agus an Caighdeán Oifigiúil

---

Kevin Scannell  
Cadhan Aonair  
29 Samhain 2024

# Cad is teanga ann?

- Ceist sách casta!
- Beidh mé ag plé nósanna gramadaí agus litrithe amháin
- Teanga scríofa amháin freisin
- Roinnt freagraí...
  - **CO**: Gramadach atá leagtha amach sa Chaighdeán Oifigiúil (do rogha leagan)
  - **NEID**: Gramadach atá leagtha amach i bhfoclóir Uí Mhianáin
  - **GSFM**: Gramadach a chleachtann an “gnáthscríbhneoir féinmhuiníneach”
  - **CNG**: Gramadach a fhaightear i gCorpas Náisiúnta na Gaeilge
  - **WEB**: Gramadach a fhaightear ó chorpas ollmhór a bhailítear ón Idirlíon
  - **LLM**: Gramadach a fhaightear ó shamhail teanga a thraenáiltear ar chorpas
- Mo chuid taighde i mbeagán focal: comparáidí idir na dearcthaí seo

# Cén fáth?

- Tá suim agam in **uirlisí profála** a chabhraíonn le daoine an teanga a scríobh
- Scríobh mé seiceálaí gramadaí in 2003 atá ag dul in aois
- I mbun oibre ar uirlis nua a chloíonn leis an **CO** (nó **NEID** más fearr leat)
- Ceann eile a thacóidh leis an **GSFM** freisin
- Bainim leas as **LLManna** san obair seo
- Ceisteanna faoi **infhoghlamthacht** ghramadach na Gaeilge in LLManna

# WEB (or CNG) vs. LLM: Infhoghlamacht

- Is féidir samhail teanga a thraenáil ó chorpas ar bith
- Torthaí taighde: difríocht an-bheag idir corpas traenála agus samhail
- Fiú le samhlacha “seanfraiseanta” (2020!); difríocht níos lú anois
- Is é sin le rá: is féidir leis an tsamhail cuid mhaith den ghramadach a fhoghlaim
- Níos mó ná sin: meastachán matamaiticiúil ar an difríocht
- Agus, na rudaí a chuireann mearbhall ar an tsamhail a aithint
- e.g. gnéithe canúna, anaphora (ina gcarr/charr/carr), códmheascadh

# WEB vs. CNG

- Dea-scéal agus drochscéal gur scáthán ar an gcorpas é an tsamhail teanga
- Go leor fadhbanna le Gaeilge a bhailítear ón Idirlíon
  - Ábhar aistrithe le Google Translate suas le 10% den Ghaeilge ar an Idirlíon
  - Go leor droch-Ghaeilge (foghlaimoirí den chuid is mó)
  - Níos mó cainteoirí L2 ná cainteoirí dúchais, níos mó fir ná mná
  - An iomarca cáipéisí oifigiúla seachas Gaeilge nádúrtha
- Úsáideann Google, OpenAI, etc. **WEB**, nó rud éigin cosúil leis
- Imríonn seo drochthionchar ar na huirlisí a dhéanann siad (agus a dhéanfaidh)
- Réitíonn **CNG** na fadhbanna seo, agus go leor eile
- Is scáthán maith ar an **GSFM** é
- Tá saineolas teanga fíorthábhachtach (ach is cuma leis na comhlachtaí teic)

## CO vs. NEID

- Difríochtaí beaga, leagtha amach go cúramach san fhoclóir féin
- m.sh. “traein paisinéirí” (**CO**) vs. “traein phaisinéirí” (**NEID**)
- nó “imreoir peile Meiriceánaí” vs. “imreoir peil Mheiriceánach”
- Iarracht, is dócha, **NEID** a tharraingt i dtreo nósanna an **GSFM**

# CO/NEID vs. CNG

- Bealach amháin ar bhain mé leas as **CNG**
- Rinne mé suirbhé cuimsitheach ar úsáid an **CO** sa chorpas, riail ar riail
- Bunaithe ar chód Python darb ainm “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 ['dlí', 'sí']:
    return [Constraint('!Len', '10.2.7.k: Do not lenite “dlí” or “sí” after a feminine noun')]

if self['lemma'] in ['bliain', 'coicís', 'mí', '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', 'dís']:
        return [Constraint('!Len', '10.2.7.m.e1: Lenite nouns after “beirt” or “dís” 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 shéimhítear an dara hainmfhocal i ndiaidh na bhfocal *aicme, comhairle, corparáid, cuideachta, earnáil, feidhmeannacht, foireann, gníomhaireacht, institiúid, oifig, rannóg, roinn, scéim, seirbhís, e.g., aicme ceannais; comhairle contae; corparáid baincéireachta; earnáil gnó; foireann bainistíochta; institiúid breisoidéachais; oifig preasa; rannóg pearsanra; roinn cosanta.*

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

(k) Ní shéimhítear na hainmfhocail *dlí* ná *sí*, i ndiaidh ainmfhocal baininsneach, e.g., *feidhm dlí; bó sí*:

Cuirfidh an rialtas cóip den **ionstraim dlí** ar fáil.  
Deirtear gur comhartha an mhí-áidh an **long sí** a fheiceáil.

(l) Ní shéimhítear an dara hainmfhocal más ceann de na haonaid tomhais bheachta *bliain, coicís, mí* nó *seachtain* an dara hainmfhocal, e.g., *saoire míosa; tréimhse bliana; moil bliana*:

Cuireadh tús leis an **bhféile seachtaine** le hócáid i dteach an Mhéara.  
Bhí **scíth coicíse** de dhíth uirthi tar éis na hoibre.

(m) Ní shéimhítear an dara hainmfhocal má tá an dara hainmfhocal, atá sa ghinideach, cáilithe le hainmfhocal eile nó 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* nó *dís* an chéad ainmfhocal, e.g., *beirt bhan mhisiúla* nó nuair is é *beirt* an dara hainmfhocal, e.g., *páistí bheirt bhan*.

# Na rialacha a bhristear is minice in CNG

- Rialacha a bhaineann leis na hathruithe tosaigh (**CO** §10), de réir %
  - 10.3.1.d.ii: “trí leabhar **fh**ada”, “deich scannán **ch**áiliúla”
  - 10.2.8.f: “beartais **c**osanta sonraí”
  - 10.9.1.b: “an **t**-ochtú cluiche”
  - 10.11.2.b: “Theip uirthi trí **h**uaire”
  - 10.2.7.m: “obair **t**ionscadail aistriúcháin”
- Maidir le “traein phaisinéirí”; tá thart ar 10% de na cásanna seo séimhithe



# Cén tairbhe dúinn an cur chuige seo?

- Níl na torthaí seo fíorthábhachtach
- Is éard atá ann i ndáiríre ná ardán lenar féidir ceisteanna mar seo a fhreagairt
- Anailís oibiachtúil, mar shampla, ar na difríochtaí idir **CO** agus **NEID**
- Eolas úsáideach maidir le haon leasú ar an **CO** amach anseo
- Eolas a chabhródh le múinteoirí teanga
  - Na pointí gramadaí is deacra
  - Na botúin is coitianta

# Go raibh maith agaibh!

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