

# Meaisínfhoghlaim agus Gramadach na Gaeilge

---

Kevin Scannell  
Ollscoil Saint Louis  
4 Samhain 2022

# Meaisínfhoghlaim

- Modhanna ríomhaireachta a ligeann do ríomhairí **tascanna** nua a fhoghlaim
  - Aghaidh a aithint i bpictiúr
  - Turscar a scagadh
  - Cluichí a imirt (ficheall, Go, etc)
  - Íomhá x-gha a léamh
- Tá suim ar leith agamsa i dtascanna a bhaineann le **teanga**
  - Teangacha nó canúintí a aithint
  - Aistriúchán a dhéanamh
  - Botúin ghramadaí a aithint agus a cheartú
  - Comhrá a dhéanamh, ceistanna a fhreagairt
- I ngach uile chás, foghlaimíonn an ríomhaire ó shamplaí den tasc

## Sonraí traenála

M Bhíos ansan ar feadh uair an chloig.

C Meas tú bhfuil aon aimhreas uirthi nach bhfuil sé i Sasana?

U D'amharc sé thart agus thug sé spléachadh ar an chompás.

M Cad a chífinn ach í ag imeacht fé dhroichead.

C M'anam muise gur goideadh glaicín fhaochan uaimse.

U Bhí sé ag scairteadh leis an chaptaen.

M Bhí ochtó neomat caide feicthe againn roimhe sin.

C Ar fhaitíos nach mbeinn in am d'fhág mé Ros Muc le fáinniú an lae.

U Bhí cupla bomaite scíste ansin againn.

...

## Measúnú: Sonraí tástála

- ? Fá dheireadh, chonaic mé scaifte de chuid gasúraí na comharsan.
- ? Mhothaíos ar uairibh gurb é sin príomhthéama an scannáin.
- ? Ní fheileann an triomach do na fataí agus iad ag fás.

## Measúnú: Sonraí tástála

- U** Fá dheireadh, chonaic mé scaifte de chuid gasúraí na comharsan.
- M** Mhothaíos ar uairibh gurb é sin príomhthéama an scannáin.
- C** Ní fheileann an triomach do na fataí agus iad ag fás.

# Patrúin

- Leideanna sna sonraí a chabhraíonn leis an ríomhaire an lipéad ceart a roghnú
- Téarmaí: scaifte, fataí, caide, srl.
- Foirmeacha gramadaí: ar an **chompás**, Mhoth**aíos**
- Eolas ginearálta faoin domhan: Ros Muc

## Tasc nua: Svaihílis (sw) nó Suáinis (tn)?

sw Akitoa mafundisho katika mafungo hayo.

sw Wadau leo ni siku yangu ya Kuzaliwa.

sw Sanaa ya mozaiki iliendelezwa katika ustaarabu wa Bizanti

sw Hakika katika haya zipo ishara kwa watu wanao amini.

tn Mabalane ke motse o mo kgaolong ya Kgatleng mo Botswana.

tn A re o tshwanetse a duelwa ka madi e seng malatsi.

tn Fa o ntse o e tlhokometse e a tsala lesaka le tlale

tn Ke eng fa morafe o rata gore ngwana wa kgosi a hudusiwe?

- - - - -

? Wanapata fursa ya kutembelea nchi za nje pia, kujifunza.

? Mme boidiidi jo bogolo jwa bo bo mo reeditse ka boipelo.

## Tasc nua: Svaihílis (sw) nó Suáinis (tn)?

sw Akitoa mafundisho katika mafungo hayo.

sw Wadau leo ni siku yangu ya Kuzaliwa.

sw Sanaa ya mozaiki iliendelezwa katika ustaarabu wa Bizanti

sw Hakika katika haya zipo ishara kwa watu wanao amini.

tn Mabalane ke motse o mo kgaolong ya Kgatleng mo Botswana.

tn A re o tshwanetse a duelwa ka madi e seng malatsi.

tn Fa o ntse o e tlhokometse e a tsala lesaka le tlale

tn Ke eng fa morafe o rata gore ngwana wa kgosi a hudusiwe?

- - - - -

sw Wanapata fursa ya kutembelea nchi za nje pia, kujifunza.

tn Mme boidiidi jo bogolo jwa bo bo mo reeditse ka boipelo.



## Tasc a bhaineann le gramadach na Gaeilge



**Caoimhín Ó Scanail**

@kscanne



Samhlaigh nach bhfuil Gaeilge ar bith agat. Míním duit go bhfuil rud darb ainm "séimhiú" ann, ina gcuirtear h isteach anois 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?

# Tasc a bhaineann le gramadach na Gaeilge

N creidmheasa

N an

S cuntais

N nó

N na

U cuntas

N sin

N maidir

N le

H éileamh

N nó

N ceart

# Aidhm nua: Uirlis eagarthóireachta

- Seiceálaí litrithe: “Daoine **spéisiúila**”, “Bhí sé **tuilte** acu”
- Seiceálaí gramadaí: “ar an **bád**”, “dhá bhád **deas**”, “cúpla **báid**”
- Moltaí téarmaíochta: “**cás** ginideach”
- Moltaí úsáide: “Tá slaghdán **agam**”, “Bhí borradh **sa** turasóireacht”

# Cur chuige meaisínfhoghlama

- Chruthaigh mé “An Gramadóir” beagnach fiche bliain ó shin
- Tá sé measartha maith, ach teoranta ar go leor bealaí freisin
- Plean fadtéarmach: uirlis nua, bunaithe 100% ar an meaisínfhoghlaim
- Sonraí traenála ag teastáil!

# Comhpháirt #1: Samhail teanga

ceadúnas      An dtabharfaidh siad \_\_\_\_ craolacháin do Raidió Rí Rá?  
chairdeas      Cuireann an scéal béim ar \_\_\_\_ agus ar chomhimirt an tsaoil.  
duine          Is féidir an méid a deir \_\_\_\_ a scríobh focal ar fhocal.  
mbeadh      Gheall Pat Rabitte dom go \_\_\_\_ ag an bpobal cead úsáide.  
leis          Tá deacrachtaí móra againn \_\_\_\_ na tíortha móra Eorpacha.  
a              Ag an am seo a scríobh sé \_\_\_\_ phríomhdhráma, Revizor.  
d'airigh      Ba é an t-uaigneas céanna a \_\_\_\_ Liam ina chroí féin.  
scrúdaithe    Ná déan dearmad den uimhir \_\_\_\_ a cuireadh chugat.  
annamh      In ainneoin sin uile is \_\_\_\_ a chloistear í san Oireachtas.

- - - - -  
?              Rinne an t-údarás áitiúil \_\_\_\_ eile sa chéad seo chaite.

# Comhpháirt #1: Samhail teanga

ceadúnas	An dtabharfaidh siad ____ craolacháin do Raidió Rí Rá?
chairdeas	Cuireann an scéal béim ar ____ agus ar chomhimirt an tsaoil.
duine	Is féidir an méid a deir ____ a scríobh focal ar fhocal.
mbeadh	Gheall Pat Rabitte dom go ____ ag an bpobal cead úsáide.
leis	Tá deacrachtaí móra againn ____ na tíortha móra Eorpacha.
a	Ag an am seo a scríobh sé ____ phríomhdhráma, Revizor.
d'airigh	Ba é an t-uaigneas céanna a ____ Liam ina chroí féin.
scrúdaithe	Ná déan dearmad den uimhir ____ a cuireadh chugat.
annamh	In ainneoin sin uile is ____ a chloistear í san Oireachtas.
- - - - -	
athruithe	Rinne an t-údarás áitiúil ____ eile sa chéad seo chaite.

# Sonraí traenála gan teorainn (?)

- Is féidir aon téacs Gaeilge a úsáid mar shonraí traenála
- Tá mé ag bailiú téacs ón Idirlíon, leabhair, irisí, srl. le 20+ bliain anuas
- Na céadta milliún focal
- Ach... is beagán é seo i gcomparáid leis an mBéarla
- Úsáideann na samhlacha teanga is déanaí níos mó ná 270 **billiún** focal
- B'fhéidir 100x níos mó ná an méid Gaeilge scríofa nó clóite **riamh**

# Cothrom na Féinne, Caighdeán, Údarás

- Caithfear na téacsanna seo a bhailiú agus a roghnú
- Tá go leor droch-Ghaeilge ar an Idirlíon; cad ba chóir dom cur san áireamh?
- Cothrom na Féinne? inscne, canúint, cainteoir dúchais/L2, Gaeltacht/cathair
- Imríonn na roghanna seo tionchar ar uirlisí a úsáideann na sonraí traenála seo
- Níor chóir domsa na roghanna seo a dhéanamh (ach déanaim chuile lá)
- Déanann Google an rud céanna chuile lá, agus is cuma leo faoin teanga
- Go leor fíor-dhroch-Ghaeilge sna sonraí traenála a úsáideann siad



## Comhpháirt #2: Samhail litrithe

Dá	Dá
mbeadh	mbeadh
ballaí	ballaí
teach	teach
tábhairne	tabhairne
Húdaí	Húdí
Bhig	Bhig
inann	inann
labhairt	labhairt

# “An Cainteoir Féinmhuiníneach”

- Caithfear cinneadh a dhéanamh — cad iad na botúin?
- Mo chur chuige: Aon rud nach mbeadh “cainteoir féinmhuiníneach” sásta leis
- Téann an Gramadóir thar fóir ag iarraidh cloí leis an gCaighdeán Oifigiúil
- Leanann sé na foclóirí (FGB, NEID, téarma.ie) go huile is go hiomlán
- Dá bharr sin, tá go leor foirmeacha coitianta i mbéal an phobail marcáilte:
- **muna** bhfuil, na **ceantracha**, na **comórtaisí**, a **thógaint**, **deireann** sé, srl.
- Ní botúin iad seo de réir na samhla nua seo
- Corpas traenála: 5k tweetanna, 100k focal, ~3k botún (ceartaithe de lámh)
- <https://github.com/kscanne/gbb/tree/main/datasets/errors>

# Na píosaí a chur le chéile

“*Táirgeadh* post dó.” → “*Tairgeadh* post dó.”

Samhail litrithe	Samhail teanga	Rangú deireanach
táirgeadh	tairgeadh	tairgeadh
tairgeadh	táirgeadh	táirgeadh
tháirgeadh	atáirgeadh	tháirgeadh
atáirgeadh	tháirgeadh	atáirgeadh
dtáirgeadh	dtáirgeadh	dtáirgeadh

# Tuilleadh sonraí: 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→mhúscail] foireann an Enterprise é de thaisme”
- Sonraí traenála nua: 115,000 ceartúchán as 11.4 milliún focal

# Obair idir lámha

- Tuilleadh sonraí trí “fhéinbhuanú” — aschur na huirlise féin a úsáid le traenáil
  - sampla den algartam EM (expectation maximization)
- Obair ar mhodhanna “inmhínithe”
  - Ba chóir don uirlis míniú a thabhairt in éineacht leis an gceartúchán
- An taighde seo a chur ar fáil don phobal i bhfoirm éigin!

# Go raibh maith agaibh!

- <https://cadhan.com/>
- <https://github.com/kscanne>
- <https://indigenoustweets.com/>
- <https://chuala.me/>