

# Meaisínfhoghlaim agus Gramadach na Gaeilge

---

Kevin Scannell

Ollscoil Saint Louis / Acadamh na hOllscolaíochta Gaeilge

18 Feabhra 2020

# Cé mise?

- As Bostún ó dhúchas
- Thosaigh mé ag foghlaim na Gaeilge i Meiriceá sna 1990idí
- D'fhorbair mé Gaelspell sa bhliain 2000
- Go leor acmhainní eile le fiche bliain anuas
- Ar dheontas Fulbright faoi láthair i gCarna, Co. na Gaillimhe



Cadhan Aonair

► GaelSpell

► Gramadóir

► Teasáras

► Caighdeánaitheoir

► Gàidhlig-Gaeilge

► Gaelg-Gaeilge

► Treochtaí

► Droichead DIL

► Twitter

► Cartlann

► Gaelú

► Fleiscíniú

► Ríomhacadamh



## Cé muidne?

Bunaíodh Cadhan Aonair, LLC sa bhliain 2008 mar chomhlacht phríobháideach a sholáthraíonn seirbhísí teicneolaíochta do phobal na Gaeilge ar fud an domhain. An bunchuspóir atá againn ná saol ríomhaireachta go hiomlán trí mheán na Gaeilge a éascú trí bhogearraí, uirlisí profála, suímh Ghréasáin, APIanna, agus logánú.

## Ar an suíomh seo

Sa roghchlár ar chlé, gheobhaidh tú nascanna leis na táirgí go léir atá againn. Cuirimid gach uile rud ar fáil saor in aisce ón suíomh seo, chomh maith le cód foinseach agus sonraí oscailte. Sa chaoi seo, bíonn forbróirí eile in ann na hacmhainní seo a úsáid mar bhunús dá gcuid táirgí féin. Creidimid go láidir gurb é seo an bealach is fearr éiceachóras sláintiúil teicneolaíochta a dheimhniú don Ghaeilge ar líne.

## Beathaisnéisín

Is é Kevin Scannell, [ollamh le Ríomheolaíocht](#) in [Ollscoil Saint Louis](#), a bhunaigh Cadhan Aonair. Is Meiriceánach é Kevin, rugadh i mBostún, Massachusetts i 1970, de bhunadh Éireannach ar an dá thaobh. Rinne sé staidéar ar an matamaitic in [Institiúid Teicneolaíochta Massachusetts](#) (MIT), áit a mbronnadh BS air sa bhliain 1991, agus fuair sé céim dhochtúireachta in [Ollscoil California, Los Angeles](#) (sa matamaitic chomh maith). Chaith sé dhá bhliain i bpost iardhochtúireachta in [Ollscoil Rice](#) in Houston, Texas, sular tháinig sé go St. Louis i 1998.

Thosaigh Kevin ag foghlaim na Gaeilge sna 1990í agus go gairid tar éis sin, thosaigh sé ag obair ar bhogearraí éagsúla a chabhródh le foghlaimoirí eile agus le cainteoirí dúchais an teanga a úsáid ar an ríomhaire. Ba é an litreoir [GaelSpell](#) an chéad acmhainn ríomhaireachta don teanga a rinne sé, an chéad leagan foilsithe sa bhliain 2000.

# Buaicphointí d'aistritheoirí

- Uirlisí profála: GaelSpell agus An Gramadóir
- Teasáras mór (PDF nó mar chuid de LibreOffice)
- An Ríomhacadamh

# Google Translate

“... tubaiste amach is amach don Ghaeilge... córas neamhfhoirfe den chineál seo a bheith ar fáil” (mise, 2009)



# intergaelic.com



 Gàidhlig  Gaeilge

FOCLÓIR

**AISTRÍÚCHÁN**

ghlèidh an bùth na cèicean a bh'aca

 Aistrigh »

**ghlèidh an bùth na cèicean a bh'aca**

choinnigh an siopa na gcácaí a bhí acu

# Samhaltú Teanga

- Tugann **samhail teanga** dóchúlacht  $P(S)$  d'aon abairt  $S$  sa teanga
- Cuir i gcás  $S = \text{"is mise stoc na cille..."}$
- $P(S) = P(\text{is}|\wedge) P(\text{mise}|\wedge\text{is}) P(\text{stocl}|\wedge\text{is mise}) \dots$
- Is leor dóchúlacht a thabhairt d'fhocal, ag brath ar na focail a tháinig roimhe
- Tá *mórán* eolais ionchódaithe i samhail teanga (chuile rud?)
- Comhréir, gramadach, nathanna cainte, eolas faoin saol, ...

# Samplaí

- Beidh cóisir againn Dé \_\_\_\_\_
- Nuair a bhí mé in Éirinn, thug mé cuairt ar Bhaile Átha \_\_\_\_\_
- Nuair a bhí mé i gContae na hIarmhí, thug mé cuairt ar Bhaile Átha \_\_\_\_\_
- Do bhíos ansan i m'áit dúchais i gContae \_\_\_\_\_
- Níor mhaith liom ceann beag, ba mhaith liom snideog \_\_\_\_\_



# Teicneolaíocht Teanga

- Úsáidtear samhlacha teanga i mbeagnach gach teicneolaíocht teanga
- Ríomhaistriúchán
- Aithint cainte
- Róbait chomhrá / ceisteanna a fhreagairt (IBM Watson, Siri, srl)
- Samhail teanga níos fearr => feidhmíocht níos fearr, go ginearálta

# Líonraí Néaracha

- Líonra néarach: creatlach mhatamaiticiúil a ligeann don ríomhaire “foghlaim”
- Dul chun cinn mór ar na teicneolaíochtaí go léir a luaigh mé
- An dúshlán is mó ná “rófhócas” ar Bhéarla sa réimse
- Go minic i bpáipéir acadúla, ní luaitear gurb é Béarla teanga an staidéir!
- Teicnící dírithe ar theangacha “cosúil le Béarla”:
  - Na billiúin focal ar fáil le haghaidh traenála
  - Easpa moirfeolaíochta spéisiúil
- Mo phlean i mbliana; na rudaí seo a chur in oiriúint don Ghaeilge

# Seiceálaí Gramadaí

- Ba mhaith liom díriú ar ghné amháin: na hathruithe tosaigh
- Cúig athrú tosaigh sa Ghaeilge: **séimhiú, urú, réimír t, réimír h, gan athrú**
- Déanann An Gramadóir jab measartha maith le botúin den sórt seo
- Úsáideann sé *rialacha!*
  - Foclóir teoranta, deacair é a choinneáil cothrom le dáta
  - Caithfear na rialacha agus eisceachtaí go léir a scríobh de lámh (cé acu Caighdeán?)
  - Déanann sé anailís ar ranna cainte; botúin san anailís sin => botúin sa seiceálaí
  - Rialacha áirithe ag brath ar chomhréir nó ar shéimeaintic
- Tugaigí faoi deara: **Inneall a thuarann athruithe tosaigh  $\subseteq$  samhail teanga**
- Sprioc: líonra néarach a úsáid chun seiceálaí gramadaí níos fearr a chruthú

# Teoiric na faisnéise

- Ní iompraíonn athruithe tosaigh sa Ghaeilge mórán eolais
- (Sainmhíniú beacht ar “eolas” anseo, de réir theoiric na faisnéise)
- Míniú níos simplí: dá scríosfainn na hathruithe tosaigh, bheadh cainteoir líofa in ann iad a chur ar ais i mbeagnach chuile chás:

Deirtear go iompraíodh sí gunnaí ina carr, iad faoi ceilt i mála plúir.

Ní raibh Gaoth Dobhair ann mar ainm dúiche ná paróiste ar tús, ach mar ainm ar an gaoth / abhainn ónar baisteadh an ceantar, an cainéal nó an inbhear farraige idir an paróiste agus na Rosa, ar a tugtar an Gaoth go dtí an lá inniú, agus an abhainn.

# Teoiric na faisnéise

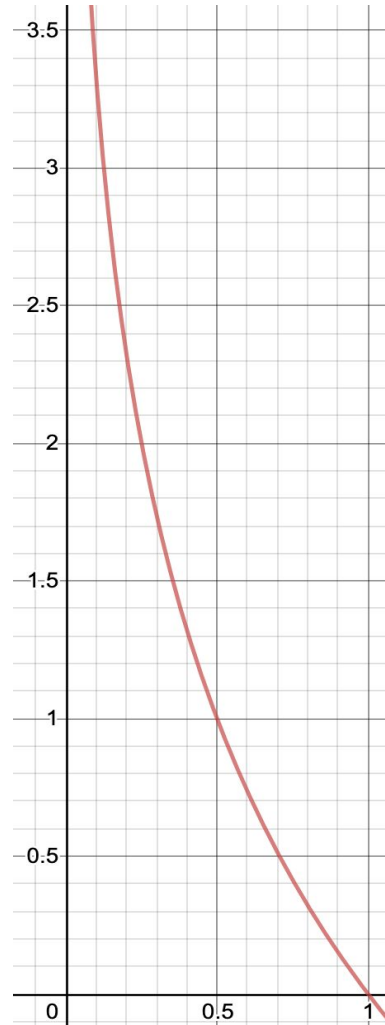
- Ní iompraíonn athruithe tosaigh sa Ghaeilge mórán eolais
- (Sainmhíniú beacht ar “eolas” anseo, de réir theoiric na faisnéise)
- Míniú níos simplí: dá scríosfainn na hathruithe tosaigh, bheadh cainteoir líofa in ann iad a chur ar ais i mbeagnach chuile chás:

Deirtear go n-iompraíodh sí gunnaí ina carr, iad faoi cheilt i mála plúir.

Ní raibh Gaoth Dobhair ann mar ainm dúiche ná paróiste ar dtús, ach mar ainm ar an ghaoth / abhainn ónar baisteadh an ceantar, an cainéal nó an t-inbhear farraige idir an pharóiste agus na Rosa, ar a dtugtar an Gaoth go dtí an lá inniú, agus an abhainn.

# Cad is eantrópacht ann?

- An cluiche céanna, ach anois tá €1.00 agat chun *geall* a chur ar na hathruithe tosaigh ar gach focal
- I gcás “...iad faoi ceilt”, d’fhéadfá geall €0.99 a chur ar shéimhiú, agus €0.0025 ar na cinn eile
- I gcás “idir an paróiste”, d’fhéadfá €0.75 a chur ar “gan athrú”, €0.24 ar shéimhiú, agus €0.003333 ar na cinn eile
- Ar gach focal, **cailleann tú**  $-\log_2$  den gheall a chuir tú ar an bhfreagra ceart (féach an graf ar dheis)
- Eantrópacht = an méid a chailleann tú ar gach focal, ar an meán (níos mó => níos deacra tuar ceart a dhéanamh)



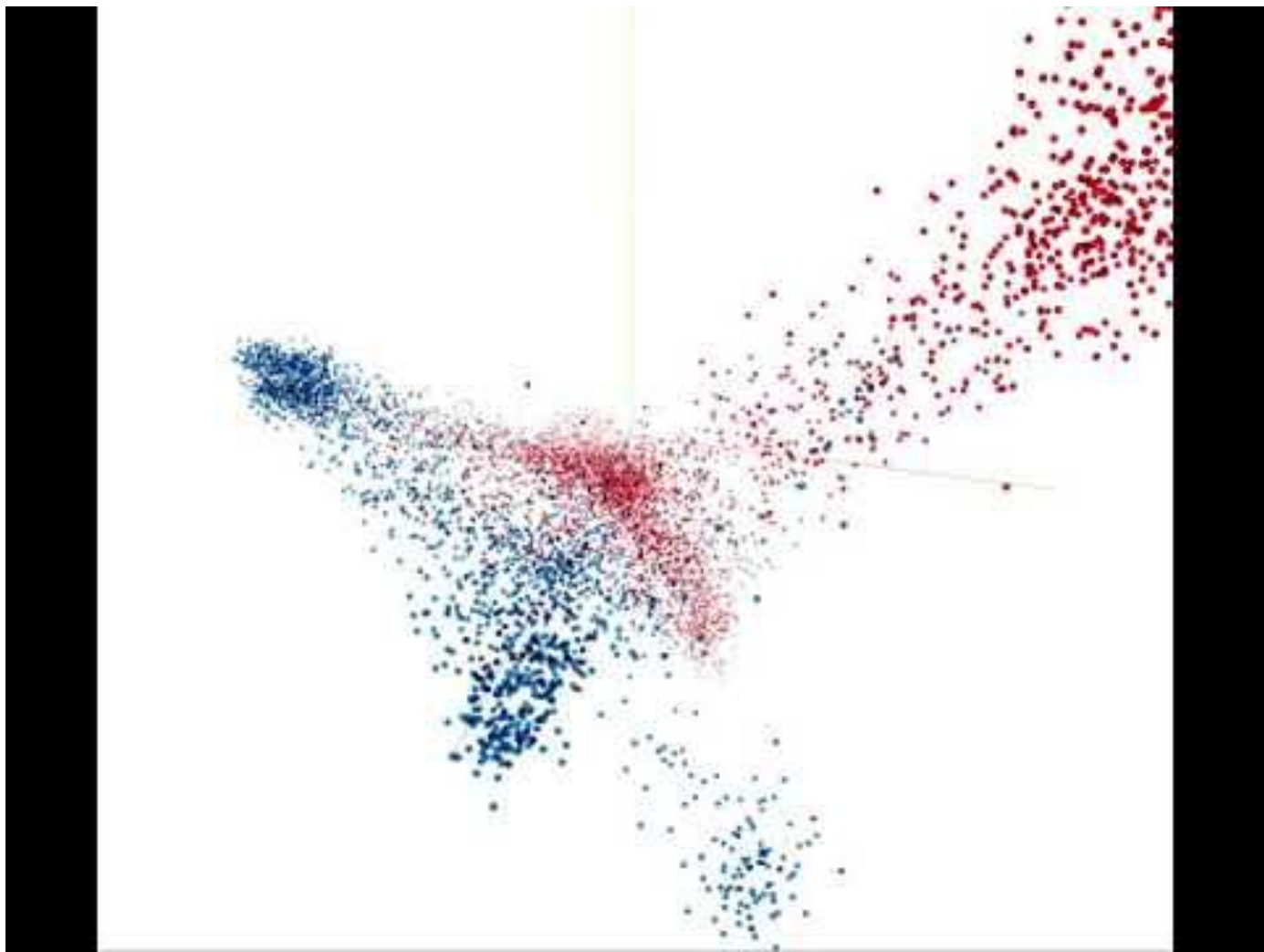
# Céard go díreach a rinne mé?

- Thóg mé líonra néarach, traenáilte ar chorpas mór Gaeilge (50 milliún focal)
- Níl aon eolas ar bith ag an líonra seachas na téacsanna sa chorpas
- Níl foclóir, ná rialacha gramadaí, ná clibeanna ranna cainte, srl. aige
- Tar éis traenála, tá sé in ann athruithe tosaigh a thuar le hardchruinneas

# Torthaí

- Tástáil le 10000 focal:  $P > 0.5$  don athrú tosaigh ceart 98.63% den am
- Tá na 137 “botún” freagrach as 77% den eantrópacht (airgead cailte!)
- 61 as 137: botúin ghramadaí sa chorpas tástála!!
- 23 as 137: leaganacha cearta nach gcloíonn leis an gcaighdeán (m.sh. *dhom*)
- Baineann 16 le haidiacht shealbhach sa tríú pearsa (*a, ina, faoina, ...*)
- Baineann 9 gcinn le cúrsaí canúna (séimhiú vs. urú sa tuiseal tabharthach)
- Roinnt samplaí eile...





# Claonadh Inscne

AIMSIGH TEANGA

BÉARLA

HAVÁIS



GAEILGE

BÉARLA

PORTAINGÉILIS

My PhD advisor Teresa is a full professor. ×



42/5000



Is ollamh iomlán é mo chomhairleoir PhD Teresa.



# Claonadh inscne

- Struchtúr tábhachtach ina bhfuil inscne marcáilte go soiléir:  
*tá sé/sí ina mhúinteoir/múinteoir*
- Bhí 446 gairm sa struchtúr seo níos mó ná seacht n-uaire sa chorpas
- Tá an t-athrú tosaigh firinscneach níos coitianta do 434 (sic!)
- Eisceachtaí: altra, aoi-léachtóir, comhláithreoir, comhordaitheoir, comhstiúrthóir, cuiditheoir, damhsóir, fídléir, gnáthurlabhra, mainicín, striapach, tréidlia
- Focail “is firinscní”: ardcheannasaí, coirnéal, ginearál, giúistís, iascaire, marcach, misinéir, óglach, peileadóir, píobaire, printíseach, seanchaí, tosaí

•saighdiúir

•rincoir

•ambasadóir

•ginearál

•giúistís

•damhsóir

•dídeanaí

•caomhnóir

•tuismitheoir

•coirpeach

•polaiteoir

•cathaoirleach

•teifeach

•óstach

•feighlí

•file

•maisitheoir

•cúramóir

•brídeog

•dearthóir

•gruagaire

•mainicín

•leabharlannaí

•dochtúir

•fiaclóir

•eacnamaí

•mainlí

•síceolaí

•múinteoir

•striapach

•ailtire

•liachleachtóir

•altra

•staráí

•réalteolaí

•innealtóir

•ollamh

•ceimiceoir

•matamaiticeoir

•fealsamh

# Claonadh inscne sa seiceálaí gramadaí

An t-athrú tuartha

An t-athrú  
atá ann  
i ndáiríre

	Firinscneach	Baininscneach	Iolra	Cruinneas	Clúdach	F-Scór
Firinscneach	189	7	2	0.9594	0.9545	0.9570
Baininscneach	8	29	2	0.8056	0.7436	0.7733

# Go raibh maith agaibh!

- <https://cs.slu.edu/~scannell/>
- <https://cadhan.com/>
- <http://crubadan.org/>
- <http://indigenoustweets.com/>
- <http://chuala.me/>
- <http://intergaelic.com/>
- <http://corpas.ria.ie/>
- <https://github.com/kscanne/>