

Struchtúr na Gaeilge mar bhuntáiste sa ríomhtheangeolaíocht

Kevin Scannell
Ollscoil Saint Louis
8 Bealtaine 2017

Cruach na Teicneolaíochta

- Barrleibhéal: Uirlisí dírithe ar ghnáthúsáideoirí
- Meánleibhéal: Parsálaithe, clibeálaithe, srl.
- Bunleibhéal: Foclóirí, Corpais, teangeolaíocht...
- Inniu: rud éigin ag an bhfíor-bhunleibhéal
- I ngar don fhealsúnacht!
- Ábhar: Ríomhfhoghlaim, intleacht shaorga (AI)

Cluiche Cearrbhachais

- Abraimis go bhfuil 1,000,000 euro agat!
- Tabharfaidh mé tús abairte duit (mar shampla, “Ní bheidh a leithéid _____”)
- Caithfidh tú geall a chur ar an chéad fhocal eile
- Is gá an t-airgead go léir atá agat a chur i ngeall
- Is féidir leat an t-airgead a “leathnú amach” ar go leor focal éagsúla (agus ba chóir!)
- Ansin babhta eile dírithe ar an chéad fhocal eile

Babhta samplach

- “Ní bheidh a leithéid _____”

arís	€600k	
ann	€120k	
de	€80k	
siúd	€40k	
eile	€25k	
...	€135k (iomlán)	

Babhta samplach

- “Ní bheidh a leithéid _____”

arís	€600k	✓
ann	€120k	✗
de	€80k	✗
siúd	€40k	✗
eile	€25k	✗
...	€135k (iomlán)	✗

Samhail Teanga

- Ríomhchlár a imríonn an cluiche seo
- Sprioc: an méid is mó airgead a shaothrú
- (An méid is lú a chailleadh! “Eantrópacht”)
- Taobh thiar de beagnach gach teic teanga!
- Aistriúchán, aithint cainte, litreoir/gramadóir, ...
- Samhail teanga níos fearr => táirgí níos fearr
- Tá daoine i bhfad níos fearr ná an ríomhaire!
- Imrímis!

Sampla #1

“Beidh cóisir againn Dé _____”

Sampla #1

“Beidh cóisir againn Dé _____”

7003 Dé hAoine
5792 Dé Sathairn
5126 Dé Luain
4976 Dé Domhnaigh
4933 Dé Máirt
4920 Dé Céadaoin
928 Dé go
577 Dé agus
508 Dé a
353 Dé Satharn
320 Dé ar
223 Dé Danann

(“cóisir” áit éigin san abairt)

16 Dé Sathairn
7 Dé Domhnaigh
6 Dé hAoine
4 Dé Máirt
2 Dé Céadaoin
1 Dé Satharn

Sampla #2a

“Nuair a bhí mé in Éirinn, thug mé cuairt ar
Baile Átha _____”

Sampla #2b

“Nuair a bhí mé i gContae na hIarmhí, thug mé cuairt ar Baile Átha _____”

Sampla #3

“Do bhíos ansan i m’áit dúchais i gContae _____”

Sampla #4

“Chomh líonmhar le gaineamh na _____”



Aonghus Ó hAlmhain

@aonghusoha

Á Leanúint



Chomh líonmhar le gaineamh na mara, nó drithlí tine nó drúcht maidin Bealtaine nó réanna neimhe - atá mo chuid túíteanna. 5000 acu anois

ATWEET

1

CROÍ

1



16:04 - 31 Lún 2011



Sampla #4

“Bhí mé ag siúl timpeall na cathrach agus
chonaic mé snideog _____”

(geall níos mó ar “mór” nó “mhór”?)

Sampla #5

“Tá Kim ag smaoineamh ar a _____”

(geall níos mó ar “pobal” nó “phobal”?)

Sampla #6

“Bhí an dochtúir ina _____”

(geall níos mó ar “oifig” nó “hoifig”?)

Promhadh Turing

- Sainmhíniú ar Intleacht Shaorga
- Bunaithe ar chumas sa chomhrá
- An féidir le ríomhaire bob a bhualadh orainn?
- Samhail teanga => róbat comhrá, go héasca
- Samhail teanga inchomparáide le duine => intleacht shaorga!

Staid na Teicníochta

- N-gramanna
- Líonraí néaracha
- Leabharlanna neamhspleách ar theanga
- Traenáilte ar chorpas gnáth-théacs
- Google (MT)
- Microsoft, Facebook, etc.



Facebook Makes A.I. Program Available in 294 Languages

Dyani Sabin | Facebook | May 2, 2017

For a computer programmer working in English, building a program that sorts email spam — using the power of artificial intelligence, you see — is straightforward enough. But doing it in Wolof — a language spoken in Senegal —

Bruscar isteach, bruscar amach

- Traenáilte ar théacs ón Ghréasán
- Facebook, m.sh., traenáilte ón Vicipéid

	Scannell	Twitter	Vicipéid
a	1	3	2
an	2	1	1
ar	3	2	6
agus	4	10	4
na	5	5	3
mé	47	17	550
catagóir	5962	20459	8

Sainoiriúint don Ghaeilge

- Neamhspleách ar theanga, traenáilte ar chorpas gnáth-théacs => inscálaithe
- Ach ní féidir leas a bhaint as gnéithe speisialta
- Mar shampla, séimhiú agus urú i nGaeilge
- Samhail teanga “fachtóirithe”

Ceachtanna

- Ná bíodh imní oraibh roimh an tsingilteacht!
- Ná cuirigí muinín in aon bhosca dubh, fiú má tá “Gaeilge” scríofa air!
- Tógáil ar na buanna atá againn
 - Foclóirí den chéad scoth idir shean agus nua
 - Corpas mór ilghnéitheach: leabhair, nuachtáin, sóisialta
 - Eolas teangeolaíochta nach bhfuil ag na comhlachtaí teic
 - Cultúr oscailte