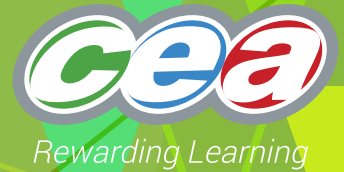


# COMHAD FÍRICÍ: TEICNEOLAÍOCHT DHIGITEACH GCE



## AONAD A2 1: CÓRAIS FAISNÉISE



### Teanga Nádúrtha agus Aithint Gutha

#### Torthaí Foghlama

##### Ba chóir do dhaltaí a bheith ábalta:

- cur síos ar an teicneolaíocht atá de dhíth ar chórais teanga nádúrtha agus aitheanta gutha.
- luacháil a dhéanamh ar úsáid chórais teanga nádúrtha agus aitheanta gutha.

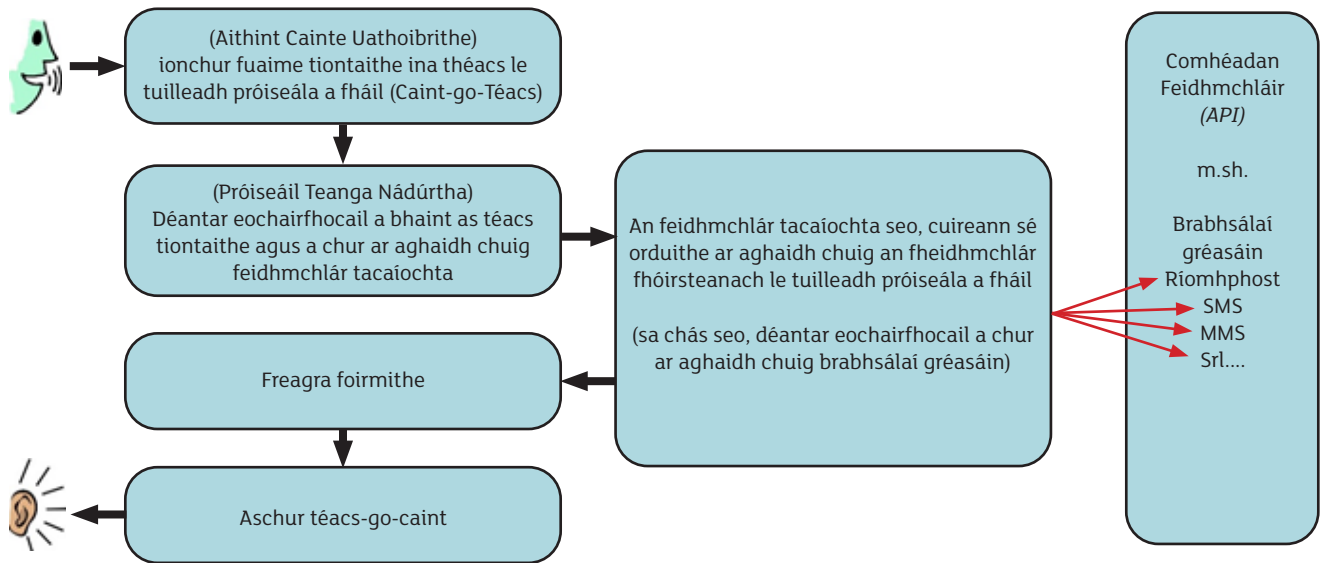
#### Inneachar i dTeanga Nádúrtha agus Aithint Gutha

- Córais Teanga Nádúrtha agus aitheanta gutha.
- An dóigh a mbaintear úsáid as próiseáil teanga nádúrtha sa lá atá inniu ann.
- Cad é mar a oibríonn próiseáil teanga nádúrtha?
  - Clibeáil Ranna Cainte.
  - Crainn pharsála.
  - Séimeantaic.
- Cad é mar a oibríonn córais aitheanta gutha?
- Luacháil a dhéanamh ar úsáid chórais aitheanta teanga nádúrtha agus aithint gutha.

#### Córais teanga nádúrtha agus aitheanta gutha

Is craobh den intleacht shaorga (IS) (AI) í próiseáil teanga nádúrtha (NLP) atá ag baint le cumas ríomhaire bheith ábalta idirghníomhaíocht dhaonna a thuiscint ina foirm nádúrtha labhartha nó scríofa. Is féidir í a shainmhíniú mar an cumas atá ag meaisín anailís a dhéanamh ar chaint an duine, í a thuiscint agus í a ghiniúint. Is é an cuspóir deiridh atá le próiseáil teanga nádúrtha ná na hidirghníomhaíochtaí idir daoine agus ríomhairí a dhéanamh chomh cosúil leis na cinn idir daoine agus is féidir.

Is féidir le haithint cainte uathobrithe (ASR) bheith mar chuid de phróiseas na próiseála teanga nádúrtha (ach ní gá dó a bheith amhlaidh). Roinnt feidhmchlár (m.sh. Siri) bainfidh siad úsáid as aithint cainte le fuaimeanna a thiontú ina dtéacs sula dtarlaíonn próiseáil teanga nádúrtha. An léaráid atá taispeánta ina dhiaidh seo, léiríonn sí an dóigh a n-úsáideann Siri aithint cainte uathobrithe agus próiseáil teanga nádúrtha le tacú le hiarratas úsáideora sonraí a fháil ón idirlíon: –



### An dóigh a n-úsáidtear próiseáil teanga nádúrtha sa lá atá inniu ann

Seo cuid de na feidhmeanna is mó a bhaintear as próiseáil teanga nádúrtha sa lá atá inniu ann:

- **Scagairí turscair** – a lán eagraíochtaí ar nós Gmail, úsáideann siad próiseáil teanga nádúrtha mar an chéad líne chosanta in éadan an turscair. Baineann scagairí turscair úsáid as próiseáil teanga nádúrtha le hiarracht a dhéanamh ar chiall a bhaint as sreanga téacs le cuidiú le ríomhphost nár iarradh a shainaithint agus lena chosc ó dhul isteach in inbhosca cliant ina bhfeidhmchláir ríomhphoist.
- **Ceisteanna a fhreagairt** – cuireann innill chuardaigh saibhreas faisnéise ar fáil dúinn ach bíonn siad ag brath ar úsáideoirí bheith thar a bheith sainiúil le heochairfhocail a úsáidtear le tacú le cuardaigh ghréasáin-bhunaithe. Comhlachtaí ar nós Google, díróinn siad ar úsáid próiseála teanga nádúrtha le cuidiú le próiseáil na gceisteanna teanga nádúrtha a chuireann an t-úsáideoir ionas gur féidir an chiall (eochairfhocail) a asbhaint agus na freagraí cuí a chur ar fáil (uaireanta i bhfoirm teanga nádúrtha).
- **Faisnéis a asbhaint** – bíonn a lán eagraíochtaí sa mhargadh airgeadais ag baint úsáid as *trádáil algartamach* mar bhealach le hinfheistíochtaí a bhainistiú. Tá infheistíochtaí airgeadais rialaithe go príomha ag teicneolaíocht a dhéanfaidh ailt nuachta a luacháil agus faisnéis ábhartha a asbhaint le patrúin i stocmhargáí a luacháil roimh chinneadh a dhéanamh cé acu ba chóir do chliaint stoic ina bpunanna a cheannach, a dhíol nó a choinneáil.

- **Faisnéis a achoimriú** – is fadhb mhór ró-ualach faisnéise do chuid mhór úsáideoirí digiteacha sa lá atá inniu ann, go háirithe trí mheáin shóisialta feidhmchláir ina bhfuil tuille ollmhór faisnéise ag teacht ann agus fógraí ó úsáideoirí eile. Soláthróirí meán sóisialta ar nós Facebook, úsáideann siad próiseáil teanga nádúrtha le hanailís a dhéanamh ar fhaisnéis faoi úsáideoirí trí mheáin shóisialta le cuidiú leo a gcuid roghnachais a chinneadh agus le cuidiú leo a fháil amach cad iad na hailt agus na fógraí ar chóir iad a chur ar fáil níos faide suas ina bhfótha nuachta. Cuid de na comhlachtaí seo, bhí siad sa nuacht ar na mallaibh cionn is go bhfuil a gcuid feidhmchláir ríomhchláraithe acu le micreafón an úsáideora a rochtain ar a ngairis mhóibíleacha agus tá siad ag úsáid na faisnéise a fuarthas ar an dóigh seo le faisnéis a bhailiú le hanailísiú.

### Cad é mar a oibríonn próiseáil teanga nádúrtha?

Is féidir le ríomhairí fadhbanna a bheith acu le próiseáil teanga nádúrtha cionn is go ndéanann siad iarracht ciall na bhfocal ar leith a thuiscint, in áit na habairte nó an fhrása iomláin. Is féidir leis bheith níos deacra abairtí agus frásaí iomlána a thuiscint mar gheall ar an chiall dhúbailte a bhíonn ag cuid mhór focal (m.sh. to “pass” the milk / to “pass” an exam) agus mar sin de, caithfidh feidhmchláir atá ag iarraidh ionchur teanga nádúrtha ar ríomhaire a thuiscint, caithfidh siad tuiscint a bheith acu fosta ar an chomhthéacs ina bhfuil an téacs á chur i láthair.

Le tacú le tuiscint a fháil ar chastachta teanga nádúrtha ní mór dóibh siúd atá ag imscrúdu próiseáil teanga nádúrtha, tuiscint a bheith acu ar na téarmaí seo a leanas: –

- Moirfeolaíocht – an dóigh a bhfoirmítear focail agus a ngaol le focail eile (baineann sí mar shampla leis an dóigh a dtig le comhthéacs ciall focail a athrú).
- Comhréir – an dóigh a gcuirtear focail agus abairtí le chéile.
- Séimeantaic – ciall na bhfocal agus na ngrúpaí focal.
- Pragmataic – comhthéacs na bhfriotal labhartha.
- Fóineolaíocht – an fhuaim atá ag baint le teanga labhartha (an dóigh a bhfuaimnítear na focail agus na frásaí nuair a deirtear iad).

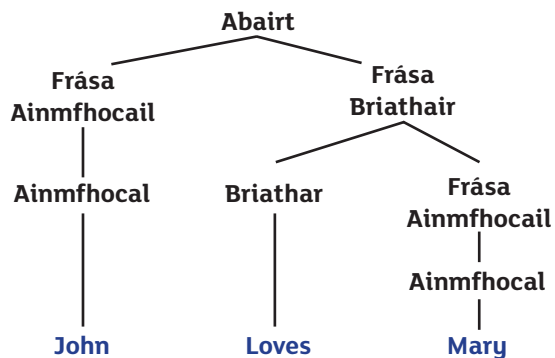
### Clíbeáil Ranna Cainte (PoS)

An chéad chéim i bpróiseáil na dteangacha nádúrtha, baineann sí leis an mhoirfeolaíocht (feidhmeanna na bhfocal ar leith a shainmhíniú; go háirithe má bhíonn débhríocht ann – m.sh. sampla an fhocail ‘pass’ thuas). An chuid is mó de na feidhmchláir nua-aimseartha, cuirfidh siad algartam féinfhoghlamtha i bhfeidhm a chlibeálfaidh focail a bhfuil a lán ciall acu.

Na feidhmchláir seo, i dtús báire, cinnfidh siad an chiall is coitianta don fhocal chlibeáilte agus ansin bainfidh siad úsáid as seo le hiarracht a dhéanamh feidhmeanna / cialla na bhfocal eile thart timpeall air a thuiscint.

### Crainn pharsála / léaráidí parsála

An chéad chéim eile sa phróiseas ná eolas a fhaightear ón chomhréir a úsáid le hiarracht a dhéanamh struchtúr na habairte a thuiscint. Déanfaidh an t-algartam an abairt a bhriseadh anuas arís agus arís eile ina frásaí ainmfhocail agus briathair le cuidiú leis an abairt a thuiscint. Is é toradh phróiseas na parsála ná crann parsála a tháirgeadh cosúil leis an cheann atá taispeánta thíos.



### Crainn pharsála / léaráidí parsála

Crainn pharsála / léaráidí parsála  
Sa tríú céim i bpróiseáil teanga nádúrtha, déantar machnamh ar shéimeantaic na habairte. Má fhaigheann an chlibeáil agus anailís chomhréire amach go bhfuil an chiall chéanna ag an fhocal, ansin déanfaidh an chéim seo próiseáil ar na focail atá ann roimh an fhocal chlibeáilte agus ina dhiaidh le ciall a chur i bhfeidhm ar an abairt.

Mar shampla:

*Cut the bread on the board.*

*He is a member of the board.*

Duine a léann an dá abairt seo, beidh sé/sí ábalta an difríocht a aithint idir *chopping board* (clár gearrtha) agus *board of directors* (bord stiúrtha) sa dá chás, ach caithfidh feidhmchlár próiseála teanga nádúrtha a ‘fhoghlaim’ gur tagairt atá ann do chlár gearrtha má tá an focal ‘bread’ roimh nó i ndiaidh an fhocail, ach má tá an focal ‘member’ roimhe nó ina dhiaidh, is mó an dóchúlacht gur *board of directors* atá ann.

Is limistéar casta í próiseáil teanga nádúrtha agus tá a lán rudaí le feabhsú go fóill, go háirithe maidir leis an **phragmataic**. Bíonn comhthéacs i gceist leis an chuid is mó de na habairtí a bhfuil tuiscint ghinearálta ar shaol na ndaoine de dhíth air agus is féidir leis bheith deacair sin a mhúineadh do ríomhaire, go háirithe maidir le magadh nó tarcaisne, mar shampla.

### Cad é mar a oibríonn córais aitheanta gutha?

An téarma aithint gutha (VR), tagraíonn sé don teaghlaim de chórais chrua-earraí agus bhogearraí a bhfuil ar a gcumas ordú labhartha a dhíchódú. Is minic a úsáidtear aithint gutha le gairis a oibriú nó le horduithe a rith gan an gá le gairis fhorimeallacha a úsáid amhail eochairchláir, lucha nó cip rianaithe, mar shampla.

Is é an chéad chéim in aithint gutha ná an t-ionchur agus digitiú na bhfocal labhartha isteach i mbogearraí aitheanta gutha. Leis seo a chur i gcrích, tá ríomhaire / gaireas le cárta fuaimne de dhíth, chomh maith le micreafón nó fearas cinn. Roinnt feidhmchlár, bainfidh siad úsáid as gairis chrua-earraí ar leith le tacú le haithint gutha, agus cinn eile ar nós fóin chliste, beidh na cruu-earraí go léir atá de dhíth ionsuite sa ghaires. Ní mór bogearraí speisialaithe a bheith suiteáilte ar an ghaires le tacú le haithint gutha.

### An dóigh a n-aithníonn ríomhairí caint

Is féidir le ríomhairí ceann de roinnt cur chuige a ghlacadh maidir le haithint cainte. Ina measc siúd, tá: –

- Meaitseáil patrún ina n-aithnítear na focail a deir an t-úsáideoir ina n-iomláine. Is minic a úsáideann gnólachtaí a bhfuil lasc-chláir uathoibrithe acu córais den chineál seo. Cuirfear ceisteanna le freagraí teoranta agus simplí acu ar an úsáideoir (m.sh. *Yes/No*). Déanfaidh an ríomhaire anailís ar an ionchur ón úsáideoir agus déanfaidh sé iarracht é a mheaitseáil le liosta de phatrúin fuaime fhéideartha a léiríonn gach ceann de na freagraí atá ar fáil.
- Anailís phatrún agus ghné. Anseo, déantar an t-ionchur labhartha a thaifeadadh le micreafón agus ansin déantar é a dhigitiú ag úsáid ADC (tiontaire analógach go digiteach). Déantar na sonraí digiteacha seo a anailísiú ansin agus cuirtear i gcomparáid iad le foclóir stóráilte arbh fhéidir úsáid a bhaint as leis an méid a

dúirt an t-úsáideoir a shainaithint. Is féidir ionchur níos casta a anailísiú leis an mhodh seo d'ionchur agus níl an t-úsáideoir teoranta do na freagraí atá siad ábalta a dhéanamh.

- Anailís staitistiúil. Is féidir le córais níos casta cur chuige níos staitistiúla a ghlacadh maidir le hanailís a dhéanamh ar ionchur cainte. Is féidir leis na córais seo rialacha gramadaí a chur i bhfeidhm le cuidiú le focail a thuar le tacú le sainaithint cainte; go háirithe i gcásanna nuair nach raibh an focal labhartha go hiomlán soiléir.
- Tá líonraí néaracha saorga (*ANN*) á bhfiosrú go fóill mar dhóigh le tacú le haithint gutha. Tá eolaithe ag amharc ar an dóigh ar féidir iad a oiliúint trí úsáid a bhaint as samplaí le hionchur labhartha a aithint. Tá staidéir níos déanaí ag amharc ar líonraí néaracha saorga agus anailís staitistiúil a úsáid le chéile le cuidiú le cruinneas na bhfeidhmchlár aitheanta gutha a fheabhsú.

### Úsáid teanga nádúrtha agus córais aithint gutha a luacháil

	Buntáistí	Míbhuntáistí
<b>Teanga Nádúrtha</b>	<ul style="list-style-type: none"> <li>• Níl oiliúint ar bith de dhíth leis an fheidhmchlár a úsáid</li> <li>• Méadaítear inrochtaineacht an fheidhmchláir</li> <li>• Má úsáidtear le chéile le haithint cainte é is féidir leis lámha an úsáideora a shaoradh le tascanna eile a dhéanamh</li> </ul>	<ul style="list-style-type: none"> <li>• Ní aithneofar/ ní shainaithneofar gach ordú mar is ceart, an t-am ar fad</li> <li>• Tá feidhmchlár casta a chruthú agus mar sin de, bíonn siad costasach</li> </ul>
<b>Aithint gutha</b>	<ul style="list-style-type: none"> <li>• Níl oiliúint ar bith de dhíth leis an fheidhmchlár a úsáid</li> <li>• Litriú cruinn focal d'úsáideoirí a bhfuil fadhbanna litearthachta acu</li> <li>• Do dhaoine a bhfuil gluaiseacht theoranta acu, is modh níos éifeachtaí den ionchur é</li> </ul>	<ul style="list-style-type: none"> <li>• Tugtha don trasnaíocht</li> <li>• Seans go mbeadh ar an úsáideoir labhairt níos fadálaí agus níos airde ná mar is gnách</li> <li>• Seans nach n-aithneodh sé ach raon teoranta guthanna / is gá é a oiliúint go sainiúil le guth úsáideora a aithint</li> </ul>

## Ceisteanna

1. Mínigh an difríocht idir córais phróiseála teanga nádúrtha agus córais aitheanta gutha. [2]
2. Cuir síos ar na téarmaí seo a leanas atá ag baint le córais phróiseála teanga nádúrtha
  - a. Moirfeolaíocht [1]
  - b. Comhréir [1]
  - c. Séimeantaic [1]
3. Déan taighde ar úsáidí féideartha de phróiseáil teanga nádúrtha agus déan liosta de na húsáidí féideartha. [3]
4. Mínigh an dóigh a n-úsáideann córais aitheanta gutha ar nós Siri úsáid as córais aitheanta gutha agus próiseála teanga nádúrtha le hiarratais úsáideora a phróiseáil. [6]
5. Mínigh na téarmaí parsáil agus ranna cainte i gcomhthéacs próiseála teanga nádúrtha. [4]
6. Sainaithin na trí phríomheilimint chrua-earraí a bhfuil baint acu le córais aitheanta gutha. [3]
7. Sainaithin agus cuir síos ar 3 chineál cur chuige dhifriúla atá á ndéanamh faoi láthair le tacú le haithint gutha. [6]
8. Liostaigh 3 mhír chrua-earraí a úsáidtear le tacú le haithint gutha. [3]
9. Déan taighde ar úsáid phróiseála teanga nádúrtha agus aitheanta gutha. Déan liosta d'fheidhmchláir ina n-úsáidtear gach ceann. Taobh le gach feidhmchlár, cuir isteach liosta de na buntáistí agus na míbhuntáistí a bhaineann le húsáid phróiseála teanga nádúrtha agus aitheanta gutha don fheidhmchlár sin.

## Leabharliosta

BCS Academy Glossary Working Party, 2013, BCS Glossary of Computing and ICT, 13ú hEagrán, Swindon, BCS Learning and Development Ltd

## Naisc

- <http://blog.neospeech.com/what-is-natural-language-processing/>
- <https://www.lifewire.com/applications-of-natural-language-processing-technology-2495544>
- <https://www.1and1.co.uk/digitalguide/online-marketing/online-sales/how-does-natural-language-processing-work/>
- <https://www.computerhope.com/jargon/v/voicereco.htm>
- <http://www.explainthatstuff.com/voicerecognition.html>

